PR13 Efficiency Benchmarking of Network Rail using LICB

August 2013

Contents

1. Introduction

1.1 This paper documents the process followed in the 2013 Periodic Review (PR13) econometric benchmarking of Network Rail using the Lasting Infrastructure Costs Benchmarking (LICB) dataset[1]. It records ORR's work on improving the quality and reliability of our econometric estimates of Network Rail's efficiency gap, and the steps taken to address questions put forward by Network Rail and recommendations from reviews of the 2008 Periodic Review (PR08) work.

Principles of Frontier analysis:

1.2 This work uses an econometric technique known as (cost) frontier analysis, the starting point of which is the estimation of a cost function. The econometric estimation of a cost function involves looking at how variations in cost drivers (e.g. network size, train density, single track) are correlated with variations in costs. By taking into account differences in cost drivers (between infrastructure managers and over time) we get some indication about the likely effect on costs of a certain percentage increase in a cost driver (for instance, a 1% increase in passenger train density might increase total costs by 0.5%), known as the elasticity of costs with respect to that cost driver. A large positive (negative) elasticity suggests that costs increase (decrease) significantly with variations in that cost driver, while an elasticity close to zero indicates that costs are largely insensitive to variations in it.

1.3 Having taken into account the effect on costs of all the observable cost drivers (e.g. network size, train density, single track), frontier analysis seeks to assign some part of the *unexplained* variation in costs between firms to inefficiency. Three factors could account for this variation:

- cost drivers which vary between firms but are not included in the model due to lack of data (e.g. labour market conditions, regulatory structure, climate) – referred to in literature as "unobserved heterogeneity";
- noise (random factors that might drive costs up or down e.g. freak weather); and/or
- inefficiency.

1.4 Additionally, we can make various assumptions about time components of inefficiency. The models considered assume that inefficiency is either one of, or some combination of, the following:

- time-invariant – i.e. no change in inefficiency over time for any particular infrastructure manager (IM);
- time varying:

---

[1] A dataset produced from the submissions of the 14 European infrastructure managers that are members of the International Union of Railways.

    - independently between time periods (no systematic connection within IMs across time periods);
    - systematically between time periods (prior year's inefficiency related to future year's inefficiency) – we have considered models with linear and quadratic trends in inefficiency.

PR13 Frontier Analysis

1.5 Our PR13 analysis has considered a total of 21 models which include a variety of different approaches.[2] Of these 21 models, there are 10 basic approaches, and 11 variations built upon these. They vary either in the estimation of unobserved heterogeneity, the presumed time trend of inefficiency or in the estimation techniques used.

1.6 To investigate the effect of unobserved heterogeneity, we have introduced a "Mundlak transformation" into five of the models (these are the models labelled with an "M"). The Mundlak transformation introduces the group means of each of our regressors as additional regressors themselves.

1.7 Of the 21 models considered we view four as being sufficiently robust to form part of our final assessment. These models are:

- Corrected Ordinary Least Squares (COLS): This is a fairly basic model, commonly used by regulators, which treats firms inefficiency scores as independent over time and interprets the entire residual as inefficiency. To accommodate the possibility of noise we assume that 25% of the estimated efficiency gap for each IM is noise;
- Cuesta Linear (CUESTAL): This model, unlike the COLS model, allows for linear correlation in efficiency between periods and also allows for random shocks (i.e. noise);
- Cuesta Linear with a quadratic trend for Network Rail (CUESTAN): This is a development of the model used in PR08. It works on the same principle as the CUESTAL model except that it allows for quadratic variation in Network Rail's efficiency score – we take this assumption to be warranted on the basis of the significant upheaval in Network Rail's management over the period of the sample as a consequence of the changeover from Railtrack. We have also tested additional quadratic terms for other IMs, and found these in general to be insignificant. Details of these tests can be found in Annex B; and
- Cornwell, Schmidt and Sickles with Random Effects (CSSRE): This model is a sophisticated "random effects" model that allows for quadratic variation in efficiency scores over time as well as allowing for noise in the data.

Methodology

1.8 We have based our cost function on economic and engineering expertise, but there is inevitably an interaction with how theory fits and works with the data in practice. We therefore have followed an iterative approach: we have specified a cost function (informed by economic and engineering principles and expertise), then tested parameter estimates of this against a wide set of models, sifted out the less

---

[2] Table 5-4 in Annex A provides details of each of these models.

plausible models, and then used the remaining set of models to further inform whether our cost function specification is reasonable. We have aimed to be theory driven, but informed by the data. As an example, there may be a reasonable case from a theoretical perspective for the proportion of electrified track to be included in our models. However through testing this, the parameter estimates associated with this variable did not take the sign (that is, the direction of movement of costs in response to an increase in electrified track) we would expect from an engineering perspective. Given that we did not have a prior view as to what our final preferred model (or set of models) would be, we needed to test these parameter estimates both against a wide set of models initially, and then again against our final preferred set used, in order to develop the most robust model specification.

1.9 The final range of Network Rail's efficiency gap is based on the results of the above four models. These results are presented in Section 4.

1.10 Additionally we conducted a series of sensitivity tests to understand how stable the results were to variations in the data, sample, or model specification. These can be found in Annex B and Tables 4-3 and 4-4.

Structure of this document

1.11 The remainder of this note is divided into three main sections with two annexes. Section 2 details the steps taken in reviewing and updating the LICB data and Section 3 covers the evaluation of the cost specification. Section 4 reviews the findings of the 21 models considered and outlines the basis for moving to a final set of four models for the analysis. It also presents the range of Network Rail's efficiency gap estimates based on these models. Finally, Annex A includes further details of the models and results, and Annex B details some of the sensitivity tests undertaken.

2. Data Review

Data Quality Assurance

2.1 As well as updating the dataset to include 2010 data, our PR13 work has included a wholesale review of outliers and systematic data issues for particular countries. Three main methods have been used to identify outliers:

- The number of standard deviations from the mean (points greater than two standard deviations from the mean were immediately considered outliers);
- The percentage change year on year; and
- Visual inspection of plots of country histories (absolute and indexed to 1996).

2.2 Systematic issues were identified using visual inspection, looking at trends, and divergences of key variables. Volatility and/or implausibility over time were also considered. The subsequent ‘Removal of IMs’ section details these systematic issues, with the ‘Identification of other outliers’ section covering outliers. We only considered exclusion of countries where data showed persistent volatility or implausibility throughout the period of the sample and was deemed unsalvageable. Where there were single outlying observations for particular countries, either plausible explanations were found, or they were corrected for.

Review of IMs

2.3 The raw LICB dataset includes data from fourteen infrastructure managers across Europe: OBB (Austria), Infrabel (Belgium), BDK (Denmark), RHK (Finland), DB (Germany), Irish Rail (Ireland), RFI (Italy), CFL (Luxembourg), ProRail (the Netherlands), JBV (Norway), REFER (Portugal), BV (Sweden), SBB (Switzerland) and Network Rail. Although some countries have more than one infrastructure manager, given that each infrastructure managers is the main one of their country and that the dataset only ever includes one infrastructure manager from any particular country, throughout this report we will refer to infrastructure managers by their country of origin. Owing to either non-comparability (e.g. non-similar operating or infrastructure conditions), or data limitations the final dataset used for this analysis includes data from Austria, Belgium, Finland, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the UK (Network Rail).

Identification of other outliers

2.4 We have analysed unusual trends, atypical volatility, and outliers for countries, including the UK, and where appropriate applied corrections. This has built upon the set of data adjustments made for PR08.

2.5 Our analysis focussed on the following candidate dependent and explanatory variables:

- maintenance expenditure;
- renewals expenditure;
- route km;
- track km;
- passenger train kilometres;
- freight train kilometres;
- single track (also considered in context of proportion single track); and
- electrified track (also considered in context of proportion of track electrified).

2.6 Our method of identifying outliers used the full set of techniques outlined in paragraph 2.1, combined with our understanding of expected behaviour. For instance, renewals expenditure is inherently lumpy, so year-on-year percentage changes may be a relatively weak indicator of outliers. In contrast, route and track km ought to be very consistent, so year-on-year changes should be a far more effective tool to identify outliers.

2.7 Table 5-3 provides details on the outliers identified, explanations of the causes, and the actions taken to address them. Our preference has been only to apply adjustments where there has been a clear basis to do so. In addition to the outliers detailed in that table, we identified a few unanticipated trends across countries – a trend decrease in freight train kilometres since 2008, and a trend decline in maintenance expenditure across several countries. After discussions with the managers of the LICB dataset, the former can be accounted for through lowered freight volumes due to the recession, and the latter through a switch from a maintenance-driven to a renewals-driven asset policy in several countries.

Network Rail Renewals Expenditure Adjustment

2.8 In 2010 Network Rail revised its classification of renewals costs (for the purposes of submission to the LICB) in line with inconsistencies that it had identified between its approach and the approach taken by others. In particular, Network Rail had been classifying some expenditure on signalling, telecoms, civils, plant and machinery and other cost categories as renewals while other countries had been classifying these as enhancements (which are not included in our model, which considers only maintenance and renewals costs).

2.9 In discussion with other LICB members, the “Paris accord” was agreed, which set out guidelines for re-classifying renewals expenditure. Network Rail has now re-classified its renewals expenditure in line with these guidelines, resulting in adjustments back to 2003. Network Rail state that it is only practically possible and necessary to undertake adjustments from 2003. The adjustments are generally smaller for earlier years in the sample[3]. It is not possible to make robust adjustments internally, but we have conducted sensitivity tests to look at the impact on estimates of excluding pre-2003 years. These can be found in Annex B. It may be noted that all the adjustments act to lower reported renewals expenditure – but given the classification principles were agreed upon by all UIC members, and that costs are now being submitted on this basis, we are confident that the data is now being reported on a consistent basis.

2.10 For PR13 we have accepted Network Rail's adjustments wholesale – in line with the agreements made in the "Paris accord".

Steady State Adjustment

Motivation:

2.11 The steady state adjustment changes Network Rail's level of renewal expenditure for each year to what it would have been had the company been renewing at steady state levels. So, if Network Rail were renewing below steady state in a particular year, renewals expenditure is adjusted upwards to reflect what would have been expected if they were operating at steady state levels.

2.12 We take it that an adjustment is necessary since, following the shift from Railtrack to Network Rail, there was a significant increase in renewals expenditure as Network Rail sought to clear a renewals backlog arising from systematic underinvestment in renewals by Railtrack, the latter occurring predominantly in the years before the Hatfield crash of 2000.

2.13 Without an adjustment our econometric models may interpret this increased renewals expenditure as a fall in efficiency, because a lower volume of renewals results in a lower level of renewals expenditure. As such, we make an adjustment to set expenditure levels to match the level at which renewals would be expected on a steady state basis.

Methodology:

PR08 Approach:

2.14 In PR08, the steady state adjustment was as follows:

- Rail was assumed to have a 40 year life, requiring 2.5% of total plain line rail to be renewed each year under steady state conditions. This resulted in an assumed steady state renewal of 788 km plain line rail, using a measure of average total track length at 31,478 km.

---

[3] We also note that given our focus is on 2010 efficiency estimates that earlier years ought to have less bearing on final year estimates.

- For each year, the steady state volume was divided by actual renewal volume to give a “scaling factor”. Where actual renewal was below steady state this resulted in a scaling factor greater than 1. Where actual renewal was above steady state the scaling factor was less than 1.
- For each year, the scaling factor was then multiplied by the actual costs for signalling and track (rail + ballast + sleepers) renewals to derive the steady state adjusted costs.

PR13 Approach:

2.15 The percentage of rail renewed for a given year is estimated by taking the actual volume of rail renewed (in track km) for that year and dividing by the total length of track (in track km). For instance, in 2007 Network Rail’s track length was 31,082 km, and they renewed 1,039 km of rail. From this we estimate the percentage of track renewed to be 3.34%.

2.16 We have assumed that the steady state rate of rail renewal remains constant over the sample period – i.e. that changes in technology neither increase nor decrease asset longevity.

2.17 There are several steady state rates of renewals estimates available, and for PR13 we have used the rate that Network Rail identifies for CP4 in its September 2011 track access policy: 2.3%[4]. We have conducted sensitivity tests on variations to this rate, which can be found in Annex B, these suggest that the impact on Network Rail’s efficiency score of a change in the steady state rate of renewals is quite modest.

2.18 The steady state *volume* of renewal is how much rail an IM would renew if they were renewing at their steady state rate. It is calculated by multiplying the steady state rate of rail renewal by the length of track for that year. So, in 2007 a track length of 31,082 km, with a steady state rate of 2.3%, would result in a steady state volume of renewals of 715 km.

2.19 The “scaling factor” is the ratio of the steady state volume of renewal to the actual volume of renewal for that year. If rail renewals for a particular year fall below this steady state volume (i.e. the scaling factor is lower than 1) then the IM has under-renewed, while if it is exceeded (i.e. the scaling factor is greater than 1) they have over-renewed. For 2007 Network Rail’s “scaling factor” was 715 km (what we ‘expected’ them to renew) / 1,039 km (what they actually renewed) = 0.69.

2.20 The scaling factor is used to make the final steady state adjustment. To make this final adjustment we first break down total renewals expenditure into track and non-track renewals expenditure. We only apply the steady state adjustment to track renewals expenditure (so assume that other areas of renewals are at steady state levels).

2.21 Then the steady state adjusted level of renewals expenditure is found by multiplying the level of track renewals expenditure by the scaling factor (calculated from rail renewals volumes above) and adding this

---

[4] Others include the rate used for PR08, 2.5%, or the average percentage of rail renewed across all the LICB countries (excluding the UK), 2.65%. See Section 2.40 for more details.

to the remaining, unadjusted, non-track renewals expenditure. In multiplying by a constant scale factor we assume that unit renewal costs (i.e. costs per km) are constant with respect to volumes renewed[5]. This assumption means that when we scale up or down renewals costs, we can do so by a constant factor, i.e. if we think their renewals volumes ought to have been 10% higher, we can scale up renewals costs by 10%.

2.22 While, from an engineering basis, there is some support for the view that there are constant returns to scale in renewals activity, we would have preferred not to impose this assumption on the data. However, given that the data on renewals volumes is relatively sparsely populated in the LICB (72 observations out of a possible 150); an assumption of constant returns to scale in making the steady state adjustment is the best approach we have available.

2.23 Network Rail's track renewals expenditure in 2007-08 was £1,015m (2007-08 prices) and with scaling factor for that year of 0.69, their adjusted track renewals expenditure is £1,015m × 0.69 = £698m.

2.24 Non-track renewals expenditure in 2007-08 was £1,135m, and so the total steady state adjusted level of renewals expenditure in 2007-08 (in 2007-08 prices) is £698m + £1,135m = £1,833m.[6]

2.25 To summarise:

- $Steady\ State\ (SS)\ volume\ of\ track\ renewal\ (km) = steady\ state\ rate\ of\ rail\ renewal \times total\ length\ of\ track\ (km)$
- $Scaling\ factor = \frac{Steady\ state\ volume\ of\ track\ renewal\ (km)}{actual\ volume\ of\ track\ renewal\ (km)}$
- $SS\ adjusted\ track\ renewals\ cost = scaling\ factor \times actual\ track\ renewals\ cost$
- $SS\ adjusted\ renewals\ cost = SS\ adjusted\ track\ renewals\ cost + actual\ non - track\ renewals\ cost$

2.26 Further detail on the effect of the steady state adjustment on Network Rail's renewals expenditure can be found in Annex B[7].

Issues with the adjustment methodology:

PR08 Approach:

2.27 In 2011 the following questions were identified with respect to the PR08 methodology:

- There could be a clearer basis for the assumption of a 40 year life for rail.
- The adjustment was applied to track expenditure even though it did not take into account sleeper and ballast renewal volumes.

---

[5] That is, there are neither economies of scale nor diseconomies of scale in renewals costs.

[6] This compares to £2,253.54m unadjusted.

[7] In general, as we should expect, removing the steady state adjustment results in an increase in Network Rail's efficiency gap.

- It was also applied to signalling expenditure. There is no reason to assume signalling renewal volumes are correlated to rail renewal volumes.
- It assumes steady state volumes are the same for the whole period. This may be unlikely. Network Rail has identified tonnage as the main driver of track renewal volumes and this has increased considerably during the period. For example, in 2009 Network Rail's passenger tonnage was 115,167 gross tonne km, whereas in 1996 it was 80,510 gross tonne km. There has been a similar increase in freight gross tonne km.

Current Approach:

2.28 The current steady state methodology addresses the highlighted concerns as follows:

- *Assumed 2.5% rate of rail renewal*. We have two proposed new approaches to determine the steady state rate of rail renewal:
    - Use the average percentage of rail renewed across the sample. Since not all IMs submit renewal volumes data every year there are two candidates for the sample mean – an average of all the 59 observed values in the LICB[8] or an average across countries. The latter estimates an average for each country and then takes the average of those averages. Neither method is perfect – whilst the first reduces the weight placed on specific country outliers, the second reduces the weight placed on countries that submit more data. The second method is judged to be fairer as it does not bias an estimate to countries that happen to provide more returns. This method (excluding the United Kingdom), results in a steady state rate of rail renewal of 2.65%.
    - Identifying steady state rates on the basis of Network Rail's asset policy (i.e. using Network Rail's asset life projections as the basis of determining the steady state renewal rate). This is based on Network Rail's September 2011 track asset policy submission, submitted to ORR as part of the IIP. This states a 2.3% rate of renewal for rail – this is consistent with a 40 year life cycle for rail alongside the approximate 0.22% per annum rate of track reductions. As discussed above, this is the rate we have adopted for PR13.
- *Applying the steady state adjustment to track renewals expenditure as a whole:* Track renewals expenditure includes expenditure on ballast, sleeper and switches and crossings, as well as rail, and we do not have access to track renewals expenditure broken down into these sub-categories. However, our steady state renewals volumes have been estimated on the basis of rail renewals, i.e. excluding ballast, sleeper and switches and crossings.

  We have asked Network Rail to submit data on track renewals expenditure broken down by asset type and we have discussed the possibility of getting such data with the engineering team within ORR. Network Rail has said that they do not have track expenditure data broken down into rail, sleeper and ballast (RSB). Although the engineering team have been able to provide us with an alternative breakdown of track expenditure, we are unable either to assign these to RSB categories, or to convert the RSB steady state rates of renewal into the asset categories (e.g. steel relay, single rail) used in the expenditure data.

  On the basis that, given these data limitations, any further adjustments are likely to introduce more error than benefit, we will continue to apply the rail renewal steady state rate to track renewals expenditure as a whole.
- *Applying the steady state adjustment to signalling expenditure*: In PR08 the steady state adjustment was applied to track and signalling expenditure jointly. It is now applied only to track – there is little

---

[8] Excluding Network Rail, across the nine IMs forming the analysis we have 59 observations of rail renewal rates out of a possible 135 (that is, 9 IMs across 15 years).

engineering basis for assuming that track and signalling have the same steady state rate of renewal. Since no adjustment is applied to signalling renewals expenditure, we have effectively assumed that signalling renewals are at steady state.

- *Constant steady state rate:* As the main driver of track renewal volumes, increased tonnage over the period of the sample would likely increase the rate of steady state renewal. However, improvements in technology generally reduce the level of steady state renewal by increasing asset longevity. It is difficult to estimate what the net effect on any steady state rate is. Since Network Rail has assumed a constant rate of steady state renewal for CP4 of 2.3%, we take this to a reasonable figure to use for this analysis.

PPP and Inflation Adjustment

2.29 For PR08, we needed to adjust for different price levels between railways in our sample group. Our approach involved converting cost data using purchasing power parities (PPPs) exchange rates, which is a widely used method to convert between currencies. This section outlines the theoretical premises of PPPs, why they have been used, and the approaches used for PR08 and PR13.

Background

2.30 PPPs are indicators of price level differences across countries. They tell us how many currency units a given quantity of goods and services cost in different countries. PPPs can be used as currency conversion rates to convert expenditures expressed in national currencies into an artificial common currency (the purchasing power standard) eliminating the effect of price level differences across countries.

2.31 The main use of PPPs is to convert national accounts aggregates, like the gross domestic product (GDP) of different countries into comparable volume aggregates. We would view this as more appropriate to using nominal exchange rates, as these may be influenced by non-price level effects, such as interest rates or interest rate expectations. The use of PPPs should ensure the GDP of all countries is valued at a uniform price level and that the only differences remaining are the actual volumes of the country. The production of PPPs is a multilateral exercise involving the national statistics offices of participating countries, Eurostat and the OECD.

Eurostat[9], from which we have obtained our PPP data, provide the following account of the construction and use of measures of PPP:

> In their simplest form PPPs are price relatives that show the ratio of the prices in national currencies of the same good or service in different countries. For example, if good X costs €3.11 in France and £1.94 in the UK, the PPP for good X between France and the UK is 1.60. This means that for every pound spent on good X, €1.60 would need to be spent in order to obtain the same quality and quantity (in other words, the same volume) of good X. To compare the volumes of good X

---

[9] http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/prc_ppp_esms.htm

purchased in the two countries, either the expenditure on good X in France can be converted to pounds by dividing it by 1.60, or the expenditure on good X in the United Kingdom can be converted to Euros by multiplying it by 1.60.

PPPs can refer to a single product, a product group, or the economy as a whole. In moving up the hierarchy of aggregation, the PPPs refer to an increasingly complex assortment of goods and services. If the PPP for GDP between France and the UK is 1.39 euros to the pound, it can be inferred that for every pound of GDP in the United Kingdom, €1.39 would have to be spent in France to purchase the same volume of goods and services. Purchasing the same volume of goods and services does not mean that exactly identical baskets of goods and services will be purchased in both countries. The composition of the baskets will vary between countries and reflect differences in tastes and cultural backgrounds, but both baskets should, in principle, provide equivalent satisfaction or utility.

Use of PPPs in PR08 and the September 2010 econometrics update

2.32 For PR08, we used GDP PPPs to adjust cost data into a common currency. In the LICB dataset, cost data were supplied in both local currency and PPP Euros. So that we could be sure of the underlying assumptions concerning inflation and the PPP adjustment, we started with the local currency information and used PPP exchange rate data from the OECD to convert the data to a common currency and price level for each year. In this way, differences in national price levels which affect costs were controlled for. To control for inflation, the data was then deflated to a common year price level (with the price level of German Euros in 2006 as the base value). These real costs were then the final figures used in our analysis.

Use of PPPs in PR13

2.33 For PR13 we revised the approach to PPP adjustments to make the audit trail for these adjustments more straightforward. Our PR13 analysis has converted all costs to GBP and adjusted for inflation using the GDP deflator series published by HM Treasury.[10]

2.34 The procedure has been as follows:

- Convert Eurozone country cost data to Euros for the whole period of the sample. Because the Eurostat PPP data already includes a conversion of pre-Euro currencies into Euros (e.g. Austria's 1996 local currency to EU27 PPP is reported in Euros to EU27, despite the Euro not having been introduced at that stage), it is necessary to adjust nominal pre-Euro cost data into Euros for all Eurozone countries.
- Convert PPP:EU27 to PPP:GBP. Eurostat data is in PPP:EU27, in view of our decision to switch to GBP it was necessary to convert all the PPP data accordingly, this was done, for each period, t, and each country, i, on the following basis:

---

[10] Available at: https://www.gov.uk/government/publications/gdp-deflators-at-market-prices-and-money-gdp-march-2013.

$$i{:}\,GBP\ PPP_t = \frac{i{:}\,EU27\ PPP_t}{GBP\ PPP{:}\,EU27\ PPP_t}$$

- Convert country cost data to PPP adjusted GBP. Using the above rates specific country cost data could then be converted, for each period, to PPP adjusted GBP figures:

$$Nominal\ Cost\ in\ GBP_{i,t} = \frac{Nominal\ cost\ in\ local\ currency_{i,t}}{i{:}\,GBP\ PPP_t}$$

- Adjust cost data, denominated in GBP, for inflation. We adjusted for inflation using ONS GDP deflator to 2010 prices to find real costs for each country:

$$Real\ Cost\ in\ GBP_{i,t} = \frac{Nominal\ Cost\ in\ GBP_{i,t}}{UK\ GDP\ Deflator_t}$$

QA Procedure

2.35 Changes to this dataset were constructed by one member of the team at ORR and independently reviewed by two others.

3. Review of the Cost Function

3.1 This section outlines the set of alternative cost specifications considered in response to internal discussion, external recommendations (e.g. the Jon Stern review, Oxera review), and questions from Network Rail on the model used for PR08.[11] All models are tested against the sample including both track and route km as measures of network size.[12] As stated in the introduction, the development of the cost function and the preferred set of efficiency models have occurred together. We employed an iterative approach: first examining the general performance of our cost function against a wide set of models (but not rejecting anything only on the basis of one or two of these), then retesting our cost function against a narrower, and ultimately our preferred set of models. For conciseness results here are presented against the preferred set of models only. Parameter results against the full set of models are available in Annex A.

3.2 Capitalised names represent model variables – detail on all variables used in the model is available in Table 5-2 of Annex A.

3.3 Data used in the model were transformed by dividing through by the mean[13] and taking natural logs of the result[14], i.e.:

$$X_{i,t} = \ln\left(\frac{x_{i,t}}{\bar{x}}\right)$$

Where: $X_{i,t}$ is a model variable (e.g. TOTEX)

$x_{i,t}$ is the value for the variable for IM *i* in time period *t* (e.g. total expenditure for Germany in 2001)

$\bar{x}$ is the mean for the variable across all IMs and time periods (e.g. mean total expenditure)

PR08 Specification

3.4 For PR08 the preferred cost specification was:

---

[11] Jon Stern's review is available at: http://www.rail-reg.gov.uk/pr13/publications/consultants-reports.php, and the Oxera report at: http://www.rail-reg.gov.uk/upload/pdf/econometric_update_2010_oxera_paper.pdf

[12] Given that network size accounts for most of the variation in costs between IMs and over time, we give results for both measures (route km and track km) despite viewing track km as the preferred measure.

[13] In order to "standardise" the data – this makes variables denominated in different units comparable.

[14] This ensures that the parameter estimates generated by the regression are estimates of elasticities, which give the percentage change in cost as a result of a percentage change in that variable.

Dependent Variable	Explanatory Variables
TOTEX	ROUTE; PASSDR; FRDR; SING; ELEC; TIME; TIME2

3.5 That is, our dependent variable was total expenditure (maintenance + renewals), and our explanatory variables were route length (ROUTE), passenger train density (passenger train km per route km (PASSDR)), freight train density (freight train km per route km (FRDR)), the proportion of the route that was single track (SING), the proportion of the route that was electrified (ELEC), and a measure of time (1996=1, 1997=2 etc. (TIME)) and time squared (1996=1, 1997=4 etc. (TIME2))[15].

3.6 The preferred model in PR08 was a variant of the "Cuesta"[16] model, which assumed a linear trend in inefficiency variation for all infrastructure managers except Network Rail, which was given a quadratic trend (specifically to account for the perverse effect of the Hatfield incident on their efficiency time-profile). This model is referred to below as the CUESTAN model.[17]

3.7 This document investigates the four main aspects of the PR08 cost specification:

- network size: The use of route rather than track length as a measure of network size;
- train density: The use of separate measures of train density rather than a single, total measure of train density;
- track density: The use of the single track variable as the measure of track density; and
- the proportion of track electrified: The use of some measure of the degree of electrification of a network.

3.8 Since the movement between route and track specifications is the most significant change to the model in comparison to our PR08 work, results here are presented against route and track specifications. While other sensitivity tests were carried out, they are not presented here on the basis that they were consistent with these findings.

3.9 Note that throughout this report parameter estimates that are significant at the 1%, 5% and 10% significance levels are highlighted in green, yellow and red, respectively.

3.10 The identification of tests for the cost specification have been driven by underlying economic and engineering expertise, with the aim of establishing whether the parameter estimates derived from our econometric analysis accord with this expertise. While the focus of this note is on each test individually, each test has also been undertaken against each cost specification, to help develop our understanding on parameter estimates of multiple changes to the cost specification. Formal tests of whether changes in the

---

[15] Note the ELEC variable was subsequently dropped for the September 2010 update to this work.

[16] See Annex A for full details on all models used in this analysis.

[17] The CUESTAN model is discussed in Annex A.

cost specification result in models that are significantly different from each other (either jointly or singly) have generally not been possible. This is because in general the changes in specification have been non-nested, in that one model cannot be defined as a restricted version of another. In this situation direct tests of whether cost functions as a whole are significantly different from each other are not possible.

Network Size

3.11 Route kilometres ("length of lines" in LICB) is a measure of the total length of the network in a given country. Track kilometres measure the total length of track ("length of main track km"). Since some routes may be multiple track, the total length of track km must always be greater than or equal to route km. The PR08 specification used route kilometres.

3.12 Network Rail has questioned the use of route kilometres as a main explanatory variable, arguing that track kilometres may be more appropriate. This is intuitively plausible – moving from an entirely single track network to one that was entirely double track would keep the route length constant while doubling track length and yield a significant increase in costs. It should be borne in mind however that the proportion of single track variable also plays an important role in the cost function in this respect, and should pick up some aspects of the impact of network complexity on cost.

3.13 The engineering team at ORR share the view that track kilometres are the main driver of maintenance and renewal costs. Given this, along with Network Rail's evidence, we accept there are good engineering reasons for using track km as the main driver of these costs.

3.14 It is perhaps worth noting that most of the cost savings from having multiple track are operational; from the perspective of maintenance and renewals costs, if train density remains constant (which the model controls for), we expect there to be relatively little in the way of cost savings from laying or maintaining multiple track over the same length of single track.

3.15 The correlation coefficients between total expenditure and track and route lengths, across the whole LICB dataset, also support the hypothesis that track may be the main driver of costs[18], as Table 3-1 shows:

	Track	Route
Normal	0.90	0.83
Logged/Standardised	0.87	0.78

Table 3-1: Coefficients of correlation between route/track and total operating expenditure

3.16 Table 3-2 shows that estimates of coefficients on TRACK are generally significant and as stable as those on ROUTE across different model variants[19], that is, the TRACK coefficients estimated by different

---

[18] A full set of correlation coefficients is available in Annex A.

models are generally as similar to each other as the ROUTE coefficients estimated by different models are. A direct test of whether a route specification is significantly different to a track specification is not possible as the hypothesis would be non-nested.

TRACK/ROUTE Point Estimates		
	Track	**Route**
COLS	0.9800	1.0009
CUESTAL	0.9745	1.0686
CUESTAN	0.8344	0.8760
CSSRE	0.9800	1.0009

Table 3-2 Route/Track estimates (Green – significant at 1% level; Yellow – significant at 5% level; Red – significant at 10% level)

Decision:

3.17 Given the engineering rationale behind track length as the main diver of costs, as opposed to route length, the higher correlation between track length and total expenditure, and the similar stability across model variants of coefficients, a cost specification using track length is preferred.

Decision: Use TRACK instead of ROUTE.

Train Density

3.18 The PR08 model included two measures of density in the model specification – passenger train density (PASSDT) (passenger train km / route or track km) and freight train density (FRDT) freight train km / route or track km). Network Rail has suggested that we should consider using total density (TOTDT) instead ((passenger train km + freight train km) / route or track km).

Network Size and Density Coefficients

3.19 Table 3-3 compares coefficient estimates for track and route length coefficients between specifications with separate density measures (i.e. passenger train density and freight train density) and with a single total density measure.

TRACK/ROUTE Point Estimates				
	Track		**Route**	
	Separate	**Total**	**Separate**	**Total**
COLS	0.9800	0.9846	1.0009	1.0045
CUESTAL	0.9745	0.9748	1.0686	1.1023
CUESTAN	0.8344	0.8060	0.8760	0.8506
CSSRE	0.9800	0.9846	1.0009	1.0045

Table 3-3: Parameter estimates of TRACK or ROUTE (depending on specification) across separate (i.e. passenger and freight density included separately) and total density specifications.

---

[19] Parameter results in this section are only presented against our preferred models. Parameter estimates for the full set of models examined can be found in Annex A.

3.20 Table 3-4 compares coefficient estimates on passenger density and total density across ROUTE and TRACK specifications. Because, in general, estimates of coefficients on freight train density are very small and statistically insignificant, they have been excluded from Table 3-4.

Density coefficient estimates	**Track**		**Route**	
	Separate	**Total**	**Separate**	**Total**
COLS	0.5583	0.6251	0.5938	0.6618
CUESTAL	0.6575	0.8226	0.4621	0.4572
CUESTAN	0.4353	0.4613	0.4000	0.3779
CSSRE	0.5583	0.6251	0.5938	0.6618

Table 3-4: Estimates on passenger OR total density coefficients on separate (i.e. passenger and freight density included separately) and total density specifications

3.21 Table 3-3 and Table 3-4 indicate that estimates on network size and on density coefficients are relatively stable across the core models (having similarly low variances in parameter estimates) – i.e. neither the specification with separated density measures, nor the specification with a single total, density measure produces estimates that are significantly more stable across models than the other. We conclude that there is no reason on basis of parameter estimates to prefer either specification.

Total Density and “Wrong Skew”

3.22 A final consideration that weighs against the total density specification is that running it results in a “wrong skew”. Finding “wrong skew” can be indicative of either all firms being similarly efficient or that the model has been miss-specified. The separate density specification does not suffer from this.

Decision

3.23 Given that there does not seem to be a good reason to prefer the total density measure over the separated measures, and that we do not have degrees of freedom problems, imposing this restriction would appear to run the risk of losing information without any significant gains.

Decision: Include PASSDT and FRDT in model specification and not TOTDT.

Track Density

3.24 The single track variable used in the PR08 model is a measure of the proportion of the route that is single, as opposed to multiple, track (length of single track / length of route km). We looked at what effects varying this element of the model had on parameter estimates.

Effect of Route/Track choice on Single Track Coefficient

3.25 Intuitively, in a route specification, having a higher proportion of a network that is single track might proxy for the omitted track variable (by helping capture the effects of multiple track on costs), leading to a situation in which the higher the level of single track, the lower the cost of the network. In a track specification the role of this variable is more complex. A higher proportion of single track on a network (taken as single track km per track km), could plausibly have both a cost-increasing and cost-reducing effect relative to an otherwise similar network that had a lower proportion of single track. Firstly, if there are cost benefits in having multiple track (e.g. maintenance teams can work on multiple tracks without travelling), then having a higher proportion of single track would, all else considered, have a cost-increasing effect, so we would expect the elasticity on single track to have a positive sign. Secondly, and in contrast, if there are greater costs associated with having more multiple track (because, for instance, it is more costly to take possession of multiple track routes than single track routes) then, all else considered, having a higher proportion of single track would have a cost-reducing effect, so we would expect the elasticity on single track to have a negative sign. It is likely that both of these forces have some effect on the outcome, and *a priori* it is not clear which effect will be stronger.

SING Point Estimates		
	Track	**Route**
COLS	-0.3945	-0.4932
CUESTAL	-0.5104	-0.7934
CUESTAN	-0.5382	-0.6919
CSSRE	-0.3945	-0.4932

Table 3-5: Estimates on proportion single track coefficients in different specifications

3.26 While Table 3-5 indicates a reduction in the effect of single track in the track specification as against the route specification, tests against the null hypothesis that the estimates of SING in the route specification are equal the point estimate of SING in track specifications are not significant for the preferred models as Table 3-6 demonstrates.

Model	**P values**
COLS	0.04
CUESTAL	0.00
CUESTAN	0.01
CSSRE	0.04

Table 3-6: Probability that the SING coefficient estimate in route specification was observed from a population in which the mean is equal to the point estimate of the SING coefficient in the track specification.

Using a measure of Average Track instead of Single Track

3.27 As there is some complexity around exactly what the single track variable is capturing in the model, we investigated both the effects of replacing single track with a different variable intended to be a proxy for “network complexity” – average amount of track per route km (length of track km / length of route km). The benefit of the latter measure is that it gives an indication of the degree of multiple track as opposed to just route that is and is not single track.

SING/AVTRACK POINT ESTIMATES	Track		Route	
	Sing	Avtrack	Sing	Avtrack
COLS	-0.3945	1.0737	-0.4932	1.3873
CUESTAL	-0.5104	-0.0493	-0.7934	-0.0721
CUESTAN	-0.5382	0.2682	-0.6919	0.2293
CSSRE	-0.3945	1.0737	-0.4932	1.3873

Table 3-7 Estimates of SING and AVTRACK with route and track specifications

3.28 Table 3-7 shows that using a measure of average amount of track per route kilometre appears to be less reliable than using the single track variable. Estimates of the average track variable generally tend to be insignificant, and with a sign that varies across models, i.e. some models suggest that a higher average track per route increases costs, while others suggest that it decreases costs. There is also far lower variance in the coefficient estimates on SING than on AVTRACK.

3.29 A possible explanation for the limited effectiveness of the AVTRACK variable is that, because our specification takes natural logs of all the variables, including AVTRACK is the same as including TRACK – ROUTE.[20]

Dropping Single Track as an Explanatory Variable

3.30 We also examined the effects of dropping single track on route and track coefficient estimates and on density estimates.

TRACK/ROUTE Point Estimates	Track		Route	
	Sing	No Sing	Sing	No Sing
COLS	0.9800	1.0040	1.0009	1.0024
CUESTAL	0.9745	1.0180	1.0686	1.0174
CUESTAN	0.8344	0.9174	0.8760	0.8635
CSSRE	0.9800	1.0040	1.0009	1.0024

Table 3-8: Estimates on network size measures with and without single track variable

Density coefficient estimates	Track		Route	
	Sing	No Sing	Sing	No Sing
COLS	0.5583	0.8402	0.5938	0.9128
CUESTAL	0.6575	0.9634	0.4621	0.9614
CUESTAN	0.4353	0.6581	0.4000	0.8133
CSSRE	0.5583	0.8402	0.5938	0.9128

Table 3-9: Estimates on passenger train density measures with and without single track variable

3.31 While ROUTE and TRACK coefficients remain fairly stable, there are statistically significant differences between the estimates of passenger density when SING is included and excluded. Given the positive effect on the estimates of passenger density, it seems likely that it is partly capturing the "network complexity" characteristics that the single track variable is proxying for. This is a reasonable result – we

---

[20] since (AVTRACK=log(track km/route km)=log(track km) – log (route km) = TRACK - ROUTE)

would expect a high negative correlation between passenger train density and proportion of route that is single track as networks with more frequent passenger services would need more multiple track than those with less frequent trains, which is born out in the data.

Decision:

3.32 Given that there is a reasonable degree of correlation between the proportion of route that is single track and total expenditure (correlation between logged values is -0.498), that the SING variable is significant and the effect of dropping this variable on train density, there appear to be good reasons against omitting it from the model (despite the complexity of the effects the proportion of single track can have on costs). As for replacing it with AVTRACK, the poor performance of that variable suggests that SING is a much better proxy for network complexity.

Decision: Include SING in cost specification.

Electrification

3.33 While a variable was included for the proportion of the network electrified for PR08 (using data up to 1996), it was not found to be statistically significant in any of the models used for the September 2010 update to the PR08 analysis. In response to this update Network Rail suggested the proportion of track electrified (ELEC) might be included as an explanatory variable to capture differences in cost functions between countries. This section considers the results of including ELEC in the model specification.

Electrification Coefficient

3.34 We looked at how the coefficient on electrification behaved when different samples and different measures of network size were used.

3.35 Table 3-10 shows the results against the standard comparators of route vs. track.

ELEC Point Estimates	**Track**	**Route**
COLS	-0.6284	-0.6528
CUESTAL	-0.5646	-0.4831
CUESTAN	-0.5789	-0.6830
CSSRE	-0.6284	-0.6528

Table 3-10: Point estimates on proportion of track electrified

3.36 While the estimates on the electrification variable are significant against our standard dataset, we found the parameter estimates to vary significantly with the inclusion/exclusion of specific countries. This lack of parameter stability suggests that the proportion of track electrified may not be a reliable explanatory variable.

Effect on other model parameters

3.37 The inclusion of electrification tends to have very little effect on estimates of network size coefficients, whether track or route. Similarly, it has little effect on estimates of single track. Conversely, it does appear to increase estimates on passenger train density as well as making estimates on freight train density significant, but only in those cases where the estimate on electrification is, itself, significant.

Other features of electrification

3.38 The proportion of an IM's track that is electrified is very stable over time for most IMs, and as such it is possible that this variable may be acting as a proxy for some other unobserved, time-invariant firm characteristics. To substantiate whether the significance of ELEC was plausibly a consequence of its explanatory power or more to do with it happening to be a proxy for some unobserved variable we:

- Note that the correlation between ELEC and TOTEX is very low (0.016 compared to -0.498 for SING). This is some evidence against thinking that it has the apparently high impact that some models suggest.
- We discussed with members of the engineering team at ORR whether significant, large negative parameter estimates were plausible – they confirmed that this is very unlikely; electrified assets are generally more costly to maintain and renew.

3.39 Additionally it was also found that the inclusion of the proportion of track electrified in the model generates "wrong skew".

Decision:

3.40 Since the estimate on ELEC is found to be very sensitive to which countries are included/excluded from the sample, along with its low degree of correlation with costs, and that the large negative coefficient is implausible we have decided not to include the proportion of electrified track in the cost specification.

Final Decisions:

3.41 In light of the comments above, the recommended model specification for PR13 is as follows:

Dependent Variable	Explanatory Variables
TOTEX	TRACK; PASSDT; FRDT; SING; TIME; TIME2

3.42 As at PR08, we have included time variables within our PR13 specifications. Given the long timeframe for this data (in panel terms) we view this as necessary to accommodate for frontier shift effects. We have also examined further squared terms (and cross products) in addition to those used for TIME. Details of these are provided in Annex B.

QA Procedure

3.43 Model runs have been independently performed by two members of the ORR to reach the same parameter estimates. Estimates produced by ITS Leeds using the same models and their own code have also been compared and verified, and match those produced by the ORR.

4. Model Analysis

4.1 This section reviews the 21 models considered and documents the processes used to eliminate models for use in the final estimates of any efficiency gap. The approach was as follows:

- Review of plausibility of model assumptions;
- Comparison of parameter estimates – models with estimates that were significantly different from others, and/or models that produced intuitively implausible estimates were rejected;
- Models that produced either unusable or intuitively implausible efficiency estimates were rejected on the basis that such efficiency estimates could not be a reliable basis for calculating Network Rail's efficiency gap.

4.2 Having excluded models on the basis of the above factors, we then sought to eliminate any models that were very sensitive to the data, sample, or minor changes to model specifications. This was to ensure we were confident that our models were robust, and that we were not accepting models on the basis that *by chance* they produced reasonable estimates, rather than being truly credible models for inefficiency.

Plausibility of Assumptions:

4.3 In terms of assumptions the following models stand out:

- FEI, PL and PLM assume that inefficiency does not vary over time. Given that the dataset covers 15 years and we have good independent reasons for thinking that Network Rail's efficiency has changed over the period, this assumption is not particularly attractive. Where efficiency has improved over time, these models would overstate past efficiency and understate present efficiency, and the extent of this over/under statement would worsen as the length of the period increases.
- COLS and COLSFE assume that there is no noise, i.e. that an infrastructure manager's costs are not subject to shocks that could shift their cost levels and they explain such shocks as inefficiency.

Decision:

4.4 Given that most models seem to suggest trending in inefficiency is occurring over time, and that we have good reasons for viewing this as appropriate given the length of our dataset, we made the decision that the FEI, PL and PLM models should be excluded from the analysis.

4.5 While the assumption that there is no noise is difficult to justify from a theoretical standpoint, the COLS model in particular is widely used in in efficiency benchmarking and in regulation because of its simplicity and ease of interpretation. Also, in recognition of the recommendations in Jon Stern's report that we make fuller use of COLS models in addition to more sophisticated techniques, we decided to continue to use COLS and COLSFE, but only with an explicit adjustment to reflect the noise limitation.

Parameter Estimates:

4.6 The parameter estimates for each of the default explanatory variables (track kilometres, passenger density, freight density and proportion of network that is single track) across each of the 21 models are given in Table 4-1.

4.7 Most of the models produce fairly similar parameter estimates that also correspond well with the different estimates we found when assessing the model specification.

4.8 However, both variants of the Cornwell, Schmidt and Sickles fixed effects model (CSSFEQ and CSSFEL) fare very poorly in terms of the plausibility of their estimates; they generate very low and statistically insignificant estimates for track kilometres and passenger train density. This result is not especially surprising when the character of the model is considered – the very flexible specification of inefficiency (which, in this model, is treated as an explanatory variable rather than a component of the residual) means that most of the variation over time can, and likely will, be captured as inefficiency rather than attributed to the other explanatory variables.

4.9 Additionally with regards to the PL and PLM model; while its parameter estimates are consistent in size with those of other models, they are insignificant.

4.10 The COLSFE and FEI models result in parameter estimates that are insignificant, and have much lower track and passenger density coefficient estimates compared to other models. The implication of the values being reported suggest little impact on costs of increasing passenger train density, and large economies of scale to network size (so a 10% increase in the amount of track resulting in a 5% increase in M+R costs). This seems to be at odds to other models. Also analysis of the impact of changes to model specification (see the detailed parameter estimates in Annex A) suggests particular sensitivity for this model.

4.11 The Mundlak transformation is in place to allow us to control for the influence of unobserved heterogeneity which is correlated with the regressors. All Mundlak transformed models have either low track coefficients (CUESTAM), low passenger density coefficients (PLM, PERRESM, COMPM) or both (CSSREM, TRREM). The group mean parameter estimates from the Mundlak models also appear to be highly insignificant. This could be a concern for use of these models for regulatory purposes as the frontier does not seem to be estimated with precision. However, we suspect that there is an argument to interpret the marginal effects (elasticities) in this model as the sum of the coefficient on the variable of interest and on the mean of the variable of interest (though this point has not been set out in the literature and some verification would be required). These sum (roughly) to the coefficient estimates in models without the Mundlak transformation and the sum is likely to be jointly significant. However, in this case, the point is moot, given that the Mundlak terms in the models are jointly insignificant. Therefore we conclude that the Mundlak transformation is not required.

Model	Constant	TRACK	PASSDT	FRDT	SING
COLS	-0.0638	0.9800	0.5583	0.0405	-0.3945
PSFA	-0.1101	0.9799	0.5586	0.0404	-0.3945
COLSFE/FEI[21]	0.0000[22]	0.4860	0.0856	0.1999	-0.9082
PL	-0.4007	0.9990	0.4077	0.1709	-0.6421
PLMUN	-0.3031	0.4896	0.0880	0.1996	-0.9114
CUESTAQ	-0.4240	0.8588	0.6915	0.0523	-0.3663
CUESTAL	-0.3775	0.9745	0.6575	0.1436	-0.5104
CUESTAN	-0.4664	0.8344	0.4353	0.0686	-0.5382
CUESTAMUN	-0.5129	0.0400	0.0255	-0.0143	-0.3446
CSSFEQ	0.3985	-0.0702	0.3613	-0.1566	-0.3240
CSSFEL	0.8447	-0.2600	-0.0653	0.1073	-0.7895
CSSRE	-0.0638	0.9800	0.5583	0.0405	-0.3945
CSSREO	-0.0638	0.9800	0.5583	0.0405	-0.3945
CSSREM	-0.0920	0.4176	0.0405	0.2037	-0.8503
PERRES	-0.1255	0.9618	0.3446	0.1626	-0.6166
PERRESMUN	-0.0875	0.4826	0.0833	0.2001	-0.9053
TRERE	-0.1214	0.9504	0.3013	0.1681	-0.6128
TRREMUN	-0.0920	0.4176	0.0405	0.2037	-0.8503
COMP	-0.1255	0.9618	0.3446	0.1626	-0.6166
COMPMUN	-0.1672	-0.0823	0.1385	-0.0702	-0.7528

Table 4-1: Parameter estimates across all models

Decisions:

4.12 Reject CSSFEL and CSSFEQ models, these models produce highly atypical and insignificant parameter estimates on key explanatory variables. Reject all Mundlak model variants – Mundlak transformations are insignificant and unsupported by evidence. Reject the PL and FEI models on the basis that their parameter estimates are insignificant, and in a number of cases at odds with the results of other models used in the analysis.

Efficiency scores:

4.13 Figure 4-1 shows the time path of efficiency scores for Network Rail across all 21 models.

---

[21] Note that FEI and COLSFE models are estimated using the same regression.

[22] No constant included to avoid colinearity – this model is estimated by subtracting time periods so time-invariant factors are excluded.

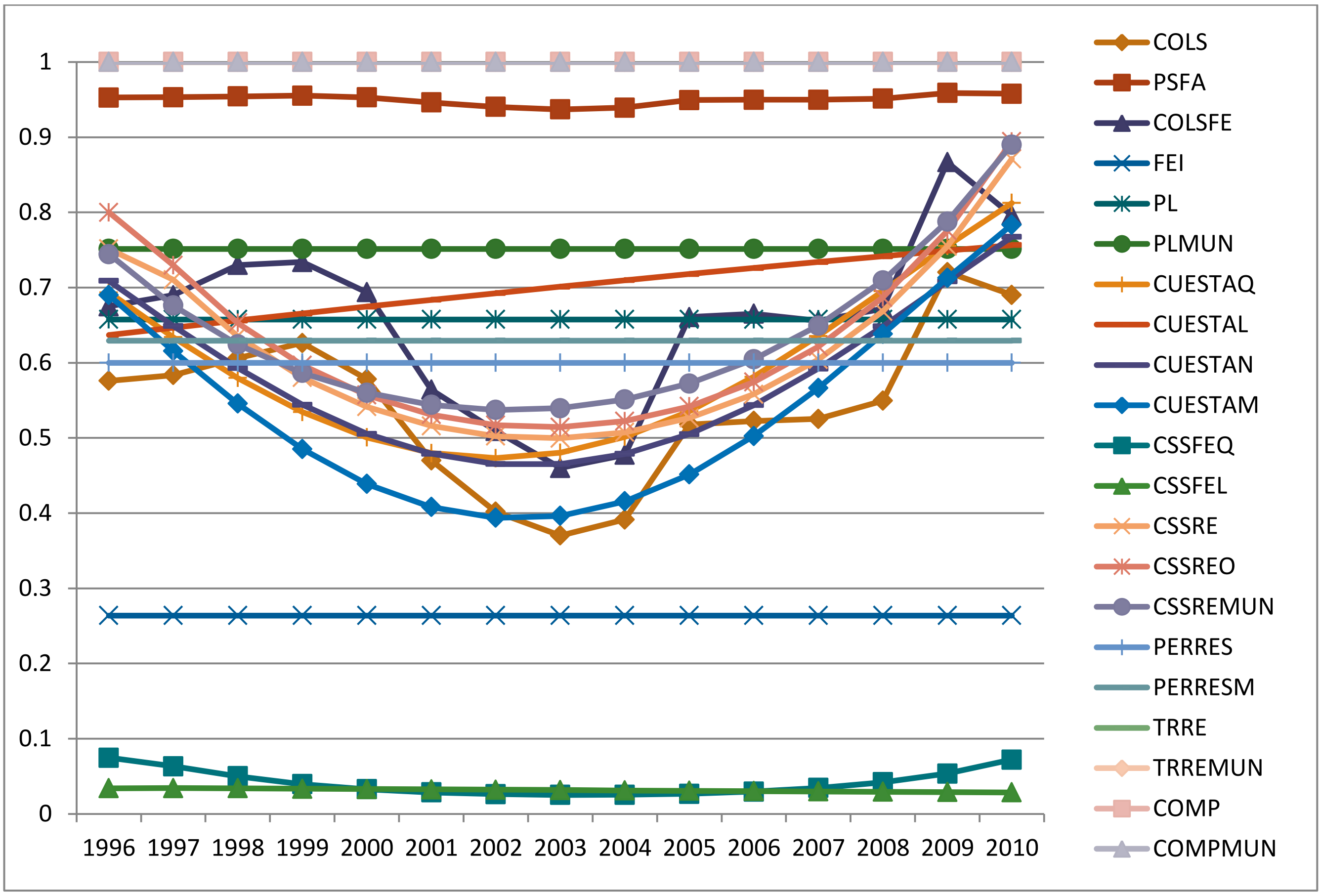

Figure 4-1: Network Rail efficiency scores over time across 21 models.

Models with no variation

4.14 The following models find no (or very little) variation in inefficiency across the whole sample (i.e. all countries are put on (or very close to) the frontier for the whole period): PSFA, TRRE, TRREM, COMP and COMPM. In the case of the first three models this is likely to be a product of the assumption of time independence - these models allocate all unexplained variation to noise and leave almost nothing to be explained as inefficiency. In the case of the COMP model “wrong skew” means that neither of the two components used to calculate inefficiency find any variation between IMs. Because they find no variation in efficiency scores across IMs (typically finding all IMs either very close to, or on, the frontier - which we do not consider credible), these models cannot be used to estimate Network Rail’s efficiency score and are therefore excluded.

Models with implausible efficiency scores

4.15 The CSSFEQ and CSSFEL models both produce very low estimates for most IM efficiency scores. It is an artefact of these models that the distribution of efficiency tends to be large – this is a consequence of their attributing a very large amount of between IM time-invariant and time-varying aspects of costs to inefficiency. Because of the implausibly large distribution of efficiency scores, we have also excluded these models from the analysis.

4.16 Additionally, the PERRES and PERRESMUN models both generate time invariant efficiency paths for all IMs. These models treat efficiency as being made up of two components, a time-invariant and an independent, time-varying component. Because the second stage of estimation used to produce the time varying component generates “wrong skew”, the model is unable to produce estimates for the time varying component, so only yields a constant efficiency score. Given the problems with time invariant efficiency profiles discussed earlier, we have also excluded these models.

COLS noise adjustment

4.17 As discussed in paragraph 4.3, the COLS model assumes that all noise is inefficiency, which is implausible. Ofwat, in their 2007-2008 relative efficiency assessments adjusted COLS residuals for water by 10% and for sewerage by 20%. In contrast Ofgem have benchmarked against the upper quartile for the most efficient firms and the upper third or average for less efficient firms.

4.18 In keeping with the view that there is some noise in the data that the COLS approach fails to account for, we have decided to apply what we would view to be a conservative 25% noise adjustment – our assumption is that, on average, 25% of an IM’s deviation from the frontier can be attributed to noise. In doing this we recognise that as the split between inefficiency and noise is unobserved, any adjustment is necessarily to some extent a matter of judgement. We have also undertaken sensitivities using 50% and 10% which are provided in Annex B. We used the following approach to calculate the noise adjusted efficiency scores (where $x_{i,t}$ is IM i’s COLS efficiency score in period t, and $x_{i,t}^{*}$ is its “noise-adjusted” efficiency score for that period):

$$x_{i,t}^{*} = x_{i,t} + 0.25(1 - x_{i,t})$$

4.19 We also considered adjusting the COLS approach by benchmarking to the upper quartile. This approach takes the efficiency score that sits 25% of the way down the ranked list of efficiency scores as the benchmark and then measures all efficiency gaps as distances from that point. We decided against this approach on the basis that it rather arbitrarily declares 25% of the observed efficiency scores to be on the frontier and is based more on a principle of excluding outliers than representing a systematic control for noise – no allowance is made for noise between the IM on the upper quartile and lower ranking IMs. In contrast, the approach we have used assumes some element of noise in every observation, and larger efficiency gaps are assumed to have a larger amount of noise, assumptions that seem more plausible than the alternative.

Decisions:

4.20 On the basis of implausible efficiency scores we have rejected: PSFA, TRRE, TRREM, COMP, COMPM, CSSFEQ, CSSFEL, PERRES, and PERRESM.

4.21 In addition we have retained the noise-adjusted efficiency score from the COLS model.

Robustness

4.22 On the basis of the above analysis the following six models are retained:

- COLS (ADJUSTED)
- CUESTAL
- CUESTAN
- CUESTAQ
- CSSRE
- CSSREO

4.23 As a final test we then look at the sensitivity of these models' efficiency scores to deviations in model specification and sample. For example we looked at the standard sensitivity tests of track against route specifications, and how estimates varied with the addition/removal of countries.

4.24 Our analysis found that the CUESTAQ model tended to show a larger range of variation in efficiency scores between changes in sample and model specification than other models. The model also proved to be highly sensitive to switching between route and track specifications (see Table 4-2).

4.25 The model was equally sensitive to the exclusion of years from the sample. Indeed, the model was even sensitive to the omission of single observations (e.g. Network Rail 1996) in a way that no other models were. It is worth noting that this effect on efficiency scores was found across infrastructure managers and that the shape of time-paths of inefficiency for several countries could be made to change dramatically with changes in model specification. For these reasons there seems to be a good basis for rejecting the CUESTAQ model.

4.26 The only other model that exhibited a level of sensitivity that was anywhere close to the CUESTAQ model was the CUESTAN model, which differs from the former in that only Network Rail has a quadratic time specification for inefficiency (all other IMs have linear specifications). However, since the range of variation is less than the CUESTAQ model (even for Network Rail's scores, where the time specification is the same), and since it contributes more in variety of modelling approaches than it brings in sensitivity to specification, we have elected to retain that model.

Decisions

4.27 Given the high degree of sensitivity of the efficiency scores generated by the CUESTAQ model to variations in model specification and marginal changes to the sample, we have excluded the CUESTAQ model from the analysis.

4.28 In addition, we have so far been considering both the CSSRE and CSSREO model. The only difference between these models is that the former is estimated by GLS while the latter is estimated by OLS, as a cross check; the two models thus produce almost the identical results and involve the same assumptions, so to include both would to place too much weight on a single model. As such we exclude the CSSREO model from our final range of models, on the basis that estimating the model using GLS (i.e. the CSSRE model) is the standard in the literature and is how the model is intended to be estimated.

	Standard Sample	
	Track	**Route**
COLS (Adjusted)	23.2%	24.0%
CUESTAL	24.4%	20.4%
CUESTAN	23.3%	22.8%
CSSRE	12.9%	12.7%
Average	**20.9%**	**16.6%**
Max	*24.4%*	*24.0%*
Min	*12.9%*	*12.7%*
Range Size	*11.5%*	*11.5%*

Table 4-2: 2010 efficiency gap estimates for Network Rail

Standard Sample	**All years**	**From 1997**	**From 1998**	**From 1999**	**From 2000**	**From 2001**
COLS	23.2%	22.0%	20.9%	19.6%	18.0%	17.3%
CUESTAL	24.4%	20.1%	18.9%	18.5%	16.0%	14.5%
CUESTAN	23.3%	18.7%	14.5%	14.1%	14.7%	18.2%
CSSRE	12.9%	7.3%	4.9%	5.7%	5.9%	11.1%

Table 4-3: Constraining sample to begin from specific year

Standard Sample	**Excl. IM 1**	**Excl. IM 2**	**Excl. IM 3**	**Excl. IM 4**	**Excl. IM 5**	**Excl. IM 6**	**Excl. IM 7**	**Excl. IM 8**	**Excl. IM 9**
COLS	23.0%	19.3%	23.6%	17.3%	22.7%	25.6%	24.1%	19.8%	21.5%
CUESTAL	20.9%	33.1%	21.6%	20.9%	43.9%	25.2%	30.4%	21.2%	14.7%
CUESTAN	10.5%	26.0%	25.1%	22.9%	31.0%	8.5%	30.3%	10.4%	9.5%
CSSRE	15.0%	1.0%	13.6%	0.0%	6.6%	5.4%	18.4%	8.8%	5.8%

Table 4-4: Removal of additional countries

Final Decisions:

4.29 On the basis of the above we retained the following four models to estimate Network Rail's efficiency score:

- COLS
- CUESTAL
- CUESTAN
- CSSRE

4.30 As stated previously we have retested our decisions around the specification of the cost function against these set of models, and found them to remain robust.

Network Rail's 2010 Efficiency Gap

4.31 On the basis of the four models selected in the previous section, Network Rail's time path for inefficiency over the period of the sample is as illustrated in Figure 4-2.

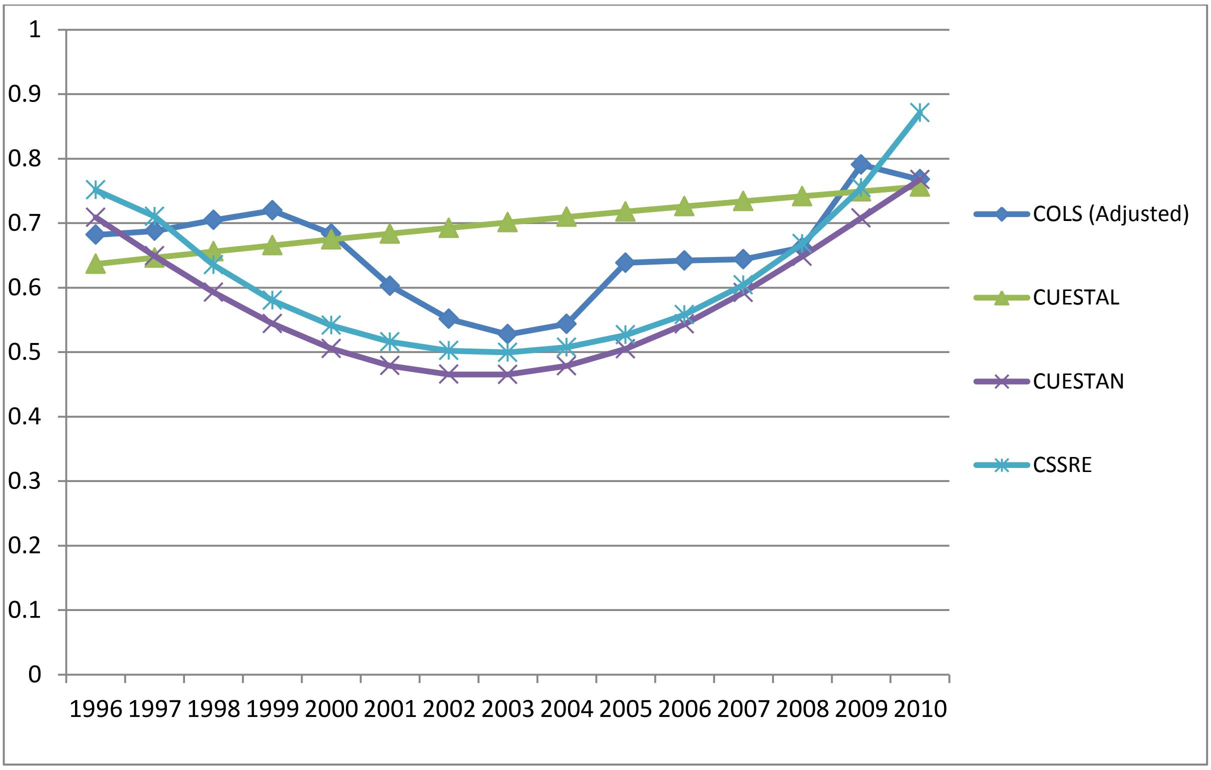

Figure 4-2: Network Rail's time-profile of inefficiency

4.32 The PR13 assessment of Network Rail's efficiency gap is thus:

Model	Efficiency Gap
COLS (Adjusted)	23.2%
CUESTAL	24.4%
CUESTAN	23.3%
CSSRE	12.9%
Mean	20.9%
Median	23.2% - 23.3%
Max	24.4%
Min	12.9%
Range Size	11.5%

QA Procedure

4.33 All models were run independently by two members of ORR and results were cross checked with estimates independently produced by ITS Leeds, whose results matched those produced by the ORR.

5. Annex A

Pre and Post Steady State Adjusted Data (£m)

Year	Unadjusted Renewals Expenditure	SS adjusted Renewals expenditure	Difference
1996	£711.00	£994.37	£283.37
1997	£830.00	£1,075.59	£245.59
1998	£840.00	£1,143.51	£303.51
1999	£1,016.00	£1,175.50	£159.50
2000	£1,469.00	£1,288.76	-£180.24
2001	£1,787.00	£1,581.15	-£205.85
2002	£2,267.00	£2,015.00	-£252.00
2003	£2,871.27	£2,259.84	-£611.43
2004	£2,338.00	£2,215.56	-£122.44
2005	£2,036.00	£1,618.05	-£417.95
2006	£2,056.94	£1,745.48	-£311.46
2007	£2,150.02	£1,833.39	-£316.63
2008	£2,253.54	£1,836.20	-£417.34
2009	£1,400.86	£1,318.90	-£81.96
2010	£1,330.00	£1,462.42	£132.42

Table 5-1: Network Rail (nominal) renewals expenditure with and without steady state adjustment

Model Variables

Variable	Description
TOTEX	Total Expenditure (GBP at 2010 Prices)
PASSDR	Passenger train km per route km
FRDR	Freight train km per route km
TOTDR	Total (passenger + freight) train km per route km
PASSDT	Passenger train km per track km
FRDT	Freight train km per track km
TOTDR	Total (passenger + freight) train km per track km
SING	Proportion of network that is single track (single track route km per route km)
AVTRACK	Average number of tracks per route km (track km/route km)
ELEC	Proportion of network electrified (electrified track km per track km)
TIME	Time period (1996 = 1, 1997 = 2, . . ., 2010 = 15)
TIME2	Time period squared (1996 = 1, 1997 = 4, . . ., 2010 = 225)

Table 5-2: Detail on model variables

Table of outliers and steps taken to address these

Country	Variable	Issue	Explanation	Action Taken
A	Route km	Significant drop 2009 to 2010	Closure/selling a significant part of their rural lines confirmed.	None.
B	Route km	Material fall 2003 to 2005	This was a result of a change to a more accurate assessment methodology.	Pre-2005 data adjusted by difference between 2003 and 2005 data.
C	Route km	Significant fall over period	The accuracy of this data has been confirmed from independent sources.	None.
	Proportion of track electrified	Significant drop 2000 to 2001	No clear explanation available but may be a consequence of changes to route.	None.
D	Maintenance	Significant fall 2006 to 2007	Likely a consequence of a change in asset policy with regards to maintenance and renewals ratios – this is borne out by changes to renewals expenditure.	None.
	Renewals	Significant increase from 2006 to 2007	It has been confirmed that this is both a consequence of a shift to a renewals focussed asset policy and government investment.	None.
E	Maintenance	Significant fall 2009 to 2010	The infrastructure manager has stated that the 2010 maintenance data has been submitted consistent with the LICB definitions, but that older data may not have been, and that it is not possible to adjust this.	Data to be retained but a sensitivity test to be run on back adjusting (by the proportionate difference in 2010 and 2009 maintenance values) historic maintenance expenditure data. Given that there may be other factors accounting for any difference between 2009 and 2010 we would not regard this sensitivity as being sufficiently robust to form a central case. Also given that the network tends to perform well in efficiency terms means that entirely dropping this data would likely lead to a general understatement of any Network Rail efficiency gap, and so we would not view dropping this data as reasonable.
	Renewals	Dip in 1999 and rise in 2000	We have no reason to doubt the accuracy of this data; this is likely just a consequence of lumpy renewals expenditure.	None.
	Single Track	Peak in 2006	This seems to be an error – the amount of single track rises 13% from 2005 to 2006 and falls by almost the same amount in 2007.	2006 peak replaced with average of 2005 and 2007.

	Track km	Peak in 2006 and 2007	This may be due to disused track being re-opened and subsequently re-closed to accommodate maintenance or renewals work.	None – insufficient evidence to suggest that this is a true outlier given that there are two corroborating years.
F	Renewals	Peak in 2009	Plausibly just a consequence of lumpy renewals expenditure. Year on year changes are not atypical of other IMs.	None.
G	Single track	Trough in 2009 and peak in 2010	We have not been able to obtain independent data on which observation is accurate.	None – without clearer evidence on which points are reliable we have left the data as it is.
H	Passenger train km	Peak in 2007	This appears to be an error in the data owing to the large rise and subsequent fall to same trend.	2007 peak replaced with average of 2006 and 2008.
I	Track km	Significant increase 1998 to 1999	It has been confirmed that pre-1999 track km data was incorrectly calculated as twice the length of route less single track plus the length of single track.	See below for detail.

Table 5-3: Identified outliers and action taken to address them.

5.1 The adjustment made to country I's data for the years prior to 1999, track km had been estimated on the basis of the following formula:

$$(Route_t - Length\ of\ route\ in\ single\ track_t) \times 2 + Length\ of\ route\ in\ single\ track_t$$

5.2 This process effectively assumed that the network only consisted of double and single track, which it has been confirmed was not the case. The 1999 data did not use this approach. In order to rectify the early year data, we assumed that the proportion of the network in greater than double track lines in 1999 was the same for previous years. We then adjusted the data according on the following basis (for values of t < 1999):

$$Adjusted\ Track_t = Track_t \times \frac{Track_{99}}{(Route_{99} - Length\ of\ route\ in\ single\ track_{99}) \times 2 + Length\ of\ route\ in\ single\ track_{99}}$$

Model Characteristics

		Decomposition of unexplained variation			Components of inefficiency				Estimation technique
						Time-varying			
							Trend		
		Noise	Unobserved Heterogeneity	Inefficiency	Time Invariant	Independent	Linear	Quadratic	
COLS	COLS			✓		✓			OLS
PSFA	PSFA	✓		✓		✓			MLE
Fixed Effects COLS	COLSFE		✓	✓		✓			OLS
Fixed Effects as Inefficiency	FEI	✓		✓	✓				OLS
Pitt and Lee	PL	✓		✓	✓				MLE
Pit and Lee + Mundlak	PLMUN	✓	✓	✓	✓				MLE
Cuesta Quadratic	CUESTAQ	✓		✓				✓	MLE
Cuesta Linear + Quadratic Network Rail	CUESTAN	✓		✓				✓	MLE
Cuesta Linear	CUESTAL	✓		✓			✓		MLE
Cuesta Quadratic + Mundlak	CUESTAMUN	✓	✓	✓	✓			✓	MLE
CSS Fixed Effects Quadratic	CSSFEQ	✓		✓	✓			✓	OLS
CSS Fixed Effects Linear	CSSFEL	✓		✓			✓		OLS
CSS Random Effects GLS	CSSRE	✓		✓				✓	GLS
CSS Random Effects OLS	CSSREO	✓		✓				✓	OLS
CSS Random Effects + Mundlak	CSSREMUN	✓	✓	✓				✓	OLS
Persistent/ Residual	PERRES	✓		✓	✓	✓			MLE
Persistent/ Residual + Mundlak	PERRESMun	✓	✓	✓	✓	✓			MLE
“True” Random Effects	TRERE	✓	✓	✓		✓			MLE
“True” Random Effects + Mundlak	TRREMUN	✓	✓	✓		✓			MLE
4 Components	COMP	✓	✓	✓	✓	✓			MLE
4 Components Mundlak	COMPMUN	✓	✓	✓	✓	✓			MLE

Table 5-4: Model Characteristics

Model Detail

5.3 This annex provides further information on the models used in this analysis.

COLS: Corrected Ordinary Least Squares (pooled)

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it}$$

5.4 This model represents the simplest approach to efficiency analysis; it assumes no correlation in inefficiency over time and because it interprets the entire residual as inefficiency, it involves no noise component.

PSFA: Pooled SFA

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_{it} \sim \left| N(0, \sigma_u^2) \right|, v_{it} \sim N(0, \sigma_v^2)$$

5.5 This model is a development on the COLS model, in so far as it has both a noise and inefficiency component. However, importantly, it also does not allow for correlation in inefficiency over time.

COLSFE: COLS with Fixed Effects

$$\ln C_{it} = \ \alpha_i + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it}$$

5.6 This model operates on the same principle as the COLS model but also allows for the effect of time-invariant firm characteristics (unobserved heterogeneity) that do not count towards inefficiency, thereby partly exploiting the panel nature of the data, but still not allowing for correlation in inefficiency over time.

FEI: Fixed Effect as Inefficiency (Schmidt and Sickles, 1984)

$$\ln C_{it} = \ f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_i + v_{it}$$

5.7 This model is estimated in the same way as COLSFE, only the fixed effect is interpreted as time-invariant technical inefficiency and the residual is interpreted as noise.

PL: Time invariant SFA (Pitt and Lee, 1981)

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_i + v_{it} \quad u_i \sim \left| N(0, \sigma_u^2) \right|, v_{it} \sim N(0, \sigma_v^2)$$

5.8 This model treats inefficiency as wholly time invariant; there is perfect correlation over time. There are two major implications for long panels. Firstly, when the panel is long it is difficult to maintain the assumption that firm’s performance does not change over time. Secondly, even if this assumption is correct, there will be a clear convolution between inefficiency and other time invariant unobserved heterogeneity in this model.

PLM: Time invariant model with Mundlak (Farsi et al, 2005a)

5.9 This model utilizes the result in Mundlak that the fixed effects linear model is equivalent to random effects and the use of the group means of the regressors as additional regressors themselves. Thus we have:

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}}, \overline{\mathbf{X}}_{\mathrm{i}};\ \boldsymbol{\beta}) + u_i + v_{it} \quad u_i \sim \left| N(0, \sigma_u^2) \right|, v_{it} \sim N(0, \sigma_v^2)$$

5.10 Importantly this formulation develops PSFA by controlling for influence of unobserved heterogeneity which is correlated with the regressors. Of course there may still be elements of unobserved heterogeneity that are not inefficiency but are still uncorrelated with the regressors and this may show up in inefficiency. Further inefficiency is still assumed to be invariant over time.

5.11 As pointed out in Farsi et al (2005a) while the decomposition holds in linear models it is only approximate in non-linear models. Nevertheless it is an easy extension of the Pitt and Lee (1981) model which should provide some improvements.

CUESTAL/CUESTAN/CUESTAQ – Deterministic time varying model (Cuesta, 2000)

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_i \sim \left| N(0, \sigma_u^2) \right|, v_{it} \sim N(0, \sigma_v^2)$$

We ran the following variants of this model (which vary in how the u_{it} component of the above model is treated):

$$(\boldsymbol{CUESTAL})\ \boldsymbol{Linear}: u_{it} = exp(\delta_{1i}t) \cdot u_i$$

$$(\boldsymbol{CUESTAN})\ \boldsymbol{NR\ Quadratic}: u_{it} = exp(\delta_{1i}t + \delta_{2i}t^2) \cdot u_i, \quad \forall i_{(i \neq NR)}\ \delta_{2i} = 0$$

$$(\boldsymbol{CUESTAQ})\ \boldsymbol{Quadratic}: u_{it} = exp(\delta_{1i}t + \delta_{2i}t^2) \cdot u_i$$

5.12 In these models we allow firm inefficiency to vary over time by scaling a time invariant one sided error by a deterministic function of time which has parameters unique to each firm. They can be traced back to Batesse and Coelli (1992) and Kumbhakar (1991) who specified this model without firm specific parameters. However, adopting a model with the same parameters across all firms is likely to be mis-specified, given the long length of the panel. There are two main formulations, a linear one (CUESTAL), and a quadratic formulation (CUESTAQ) that allows for a non-monotonic evolution of inefficiency for each firm. An additional formulation (CUESTAN) allows only for linear time paths for each IM except Network Rail, which is allowed a quadratic time-path – this is a similar to the model that was used in PR08. A sensitivity test looking at quadratic terms for other countries is provided in Annex B.

CUESTAM: Deterministic time varying model with Mundlak transformation*[23]

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}}, \overline{\mathbf{X}}_{\mathrm{i}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_i \sim \left| N(0, \sigma_u^2) \right|, v_{it} \sim N(0, \sigma_v^2), u_{it} = exp(\delta_{1i}t + \delta_{2i}t^2) \cdot u_i$$

5.13 As CUESTA models but with the addition of the Mundlak transformation in the frontier function.

---

[23] * denotes a model not yet explicitly in the literature.

CSSFE: Fixed Effects time varying CSS models (Cornwell et al, 1990)

$$(\boldsymbol{CSSFEL})\boldsymbol{Linear}\text{: } \ln C_{it} = \alpha_{1i} + \alpha_{2i}t + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + v_{it}$$

$$(\boldsymbol{CSSFEQ})\ \boldsymbol{Quadratic}\text{: } \ln C_{it} = \alpha_{1i} + \alpha_{2i}t + \alpha_{3i}t^2 + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + v_{it}$$

5.14 Unlike the previous models no strict distributional assumptions are made for this model. Inefficiency is recovered via a Schmidt and Sickles (1984) type transformation for the evaluation of the $\alpha_{1i} + \alpha_{2i}t\ (+\alpha_{3i}t^2)$ component of the model for each year. Thus in each year one firm is always efficient. Note that this model has many firm specific parameters (N more than the equivalent specification of the CUESTA models).

CSSRE: Random Effects time varying CSS models (Cornwell et al, 1990)

$$\ln C_{it} = \alpha_{1i} + \alpha_{2i}t + \alpha_{3i}t^2 + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + v_{it} = \mathbf{Z}'\boldsymbol{\alpha}_\mathbf{i} + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) \qquad E[\boldsymbol{\alpha}_\mathbf{i}] = \boldsymbol{\alpha}, v[\boldsymbol{\alpha}_{\boldsymbol{i}}] = \boldsymbol{\Omega}$$

5.15 This model is similar to CSSFEQ except that it treats the firm effects as random variables which can be correlated. It is estimated by GLS and so contrasts to the random coefficients model estimated by maximum likelihood (Greene, 2005). This model is less parameter intensive vis-à-vis the fixed effects CSS model, however it does assume no correlation between the effects and regressors (as is the assumption in all models except PL: Time invariant SFA). We estimate by both GLS (CSSRE) and OLS (CSSREO). Both methods provide consistent parameter estimates.

CSSREM: Random Effects time varying CSS models with Mundlak transformation*

$$\ln C_{it} = \alpha_{1i} + \alpha_{2i}t + \alpha_{3i}t^2 + f(\mathbf{X_{it}}, \overline{\mathbf{X}}_\mathrm{i};\ \boldsymbol{\beta}) + v_{it} = \mathbf{Z}'\boldsymbol{\alpha}_\mathbf{i} + f(\mathbf{X_{it}}, \overline{\mathbf{X}}_\mathrm{i};\ \boldsymbol{\beta}) + v_{it}$$

$$E[\boldsymbol{\alpha}_\mathbf{i}] = \boldsymbol{\alpha}, v[\boldsymbol{\alpha}_{\boldsymbol{i}}] = \boldsymbol{\Omega}$$

5.16 This model has not been used in the literature but is a simple extension of the logic in Farsi et al (2005a) namely that there is a missing time invariant component in random effect models; a component which is correlated with regressors.

PERRES: Persistent and residual inefficiency model (Kumbhakar et al, 2005)

$$\ln C_{it} = \alpha + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + \tau_i + u_{it} + v_{it}$$

$$u_{it} \sim \left| N(0, \sigma_u^2) \right|,\ v_{it} \sim N(0, \sigma_v^2), E[\tau_i] = \tau > 0, v[\tau_i] = \sigma_\tau^2$$

5.17 This model comprises two inefficiency components. First there is a persistent component of inefficiency, τ_i, which is time invariant. Then there is a residual inefficiency component, u_{it}, which is independent over time. Thus inefficiency can vary for each firm in a flexible manner, but retain an element of persistency over time. The model is estimated in a two stage process. First the model is estimated as a linear model by generalised least squares (GLS random effects). The residuals from this regression are averaged for each firm and this set of averages is used to compute τ_i. The second stage applies ML estimation to the demeaned residuals from the first stage.

PERRESM: Persistent and residual inefficiency model with Mundlak transformation*

$$\ln C_{it} = \alpha_i + f(\mathbf{X_{it}}, \bar{\mathbf{X}}_{\mathrm{i}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_{it} \sim \left|N(0,\sigma_u^2)\right|, v_{it}\sim N(0,\sigma_v^2),\ \ E[\alpha_i] = \alpha, v[\alpha_i] = \sigma_\alpha^2$$

5.18 This is an alternative method to simultaneously account for time invariant unobserved heterogeneity and persistent (time invariant) inefficiency. It is estimated in the same two stage method as for model PERRES (it simply involves including the group means of regressors alongside their levels).

TRRE: 'True' random effects model (Greene, 2005)

$$\ln C_{it} = \alpha_i + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_{it} \sim \left|N(0,\sigma_u^2)\right|, v_{it}\sim N(0,\sigma_v^2), \alpha_i\sim N(\alpha,\sigma_\alpha^2)$$

5.19 This model is the same as the PERRES, with two exceptions. Firstly the model is estimated via maximum likelihood and as such more precise distributional assumptions are required on the firm effects. Secondly, and more fundamentally, the firm effects are interpreted as unobserved heterogeneity rather than as inefficiency. Therefore PERRES and TRRE are different interpretations of the underlying model comprising a time independent inefficiency component and a time invariant component which is either interpreted as solely inefficiency (PERRES) or solely other unobserved heterogeneity (TRRE). In practice it is likely to represent both inefficiency and other time invariant unobserved effects.

TRREM: 'True' random effects model with Mundlak transformation (Farsi et al, 2005a and 2005b)

$$\ln C_{it} = \alpha_i + f(\mathbf{X_{it}}, \bar{\mathbf{X}}_{\mathrm{i}};\ \boldsymbol{\beta}) + u_{it} + v_{it} \quad u_{it} \sim \left|N(0,\sigma_u^2)\right|, v_{it}\sim N(0,\sigma_v^2), \alpha_i\sim N(\alpha,\sigma_\alpha^2)$$

5.20 This model is Model TRRE with the addition of the Mundlak transformation. Both the firm random effect, α_i , and the group means control for firm unobserved heterogeneity. Therefore persistent inefficiency is still not allowed for in this model (this is due to the interpretation of α_i as solely capturing time invariant unobserved heterogeneity and not time invariant inefficiency).

COMP: 4 components model: Persistent and residual inefficiency model with unobserved heterogeneity (Kumbhakar et al, 2012)

$$\ln C_{it} = \alpha_i + \tau_{\mathrm{i}} + f(\mathbf{X_{it}};\ \boldsymbol{\beta}) + u_{it} + v_{it}$$

$$u_{it} \sim \left|N(0,\sigma_u^2)\right|, v_{it}\sim N(0,\sigma_v^2), \alpha_i\sim N(\alpha,\sigma_\alpha^2),\ \tau_i\sim N(0,\sigma_\tau^2)$$

5.21 This model includes four error components. It is an extension of the PERRES model. It is estimated by a two stage process as in the PERRES model, but additionally ML estimation is applied to the group mean residuals from the panel model to decompose the random effect into α_i and τ_i .

COMPM: 4 components model: Persistent and residual inefficiency model with unobserved heterogeneity and Mundlak*

$$\ln C_{it} = \alpha_i + \tau_{\mathrm{i}} + f(\mathbf{X_{it}}, \bar{\mathbf{X}}_{\mathrm{i}};\ \boldsymbol{\beta}) + u_{it} + v_{it}$$

$$u_{it} \sim \left| N(0,\sigma_u^2) \right|, v_{it} \sim N(0,\sigma_v^2), \alpha_i \sim N(\alpha,\sigma_\alpha^2),\ \tau_i \sim N(0,\sigma_\tau^2)$$

5.22 This is the COMP model with the Mundlak transformation. It is estimated in exactly the same way as the COMP model.

Set of tables for Section 3 presented for full set of models

5.23 In the following tables green indicates significant at a 1% level, yellow at a 5% level and red at a 10% level.

5.24 Table 3-2 (route/track estimates) expanded:

TRACK/ROUTE Point Estimates	**Track**	**Route**
COLS	0.9800	1.0009
PSFA	0.9799	1.0009
COLSFE/FEI	0.4860	1.4301
PL	0.9990	1.0730
PLMUN	0.4896	1.2685
CUESTAQ	0.8588	0.9789
CUESTAL	0.9745	1.0686
CUESTAN	0.8344	0.8760
CUESTAMUN	0.0400	1.0643
CSSFEQ	-0.0702	2.4637
CSSFEL	-0.2600	1.2431
CSSRE	0.9800	1.0009
CSSREO	0.9800	1.0009
CSSREM	0.4176	-0.3074
PERRES	0.9618	1.0005
PERRESMUN	0.4826	1.0149
TRERE	0.9504	N/A[24]
TRREMUN	0.4176	0.8234
COMP	0.9618	1.0005
COMPMUN	-0.0823	1.0149

---

[24] Note that features of this model means that under certain specifications it cannot

5.25 Table 3-3 (parameter estimates of tack or route across separate and total density specifications) expanded:

TRACK/ROUTE Point Estimates	**Track**		**Route**	
	Separate	**Total**	**Separate**	**Total**
COLS	0.9800	0.9846	1.0009	1.0045
PSFA	0.9799	0.9846	1.0009	1.0045
COLSFE/FEI	0.4860	0.5087	1.4301	1.3564
PL	0.9990	0.9928	1.0730	1.0598
PLMUN	0.4896	0.4911	1.2685	1.1931
CUESTAQ	0.8588	0.7584	0.9789	0.7688
CUESTAL	0.9745	0.9748	1.0686	1.1023
CUESTAN	0.8344	0.8060	0.8760	0.8506
CUESTAMUN	0.0400	-0.0840	1.0643	1.0226
CSSFEQ	-0.0702	-0.2471	2.4637	2.3398
CSSFEL	-0.2600	-0.2819	1.2431	1.1612
CSSRE	0.9800	0.9846	1.0009	1.0045
CSSREO	0.9800	0.9846	1.0009	1.0045
CSSREM	0.4176	0.7889	-0.3074	-0.0043
PERRES	0.9618	0.9630	1.0005	1.0013
PERRESMUN	0.4826	0.5279	1.0149	0.9858
TRERE	0.9504	0.9073	N/A[24]	0.9331
TRREMUN	0.4176	0.6877	0.8234	0.0089
COMP	0.9618	0.9630	1.0005	1.0013
COMPMUN	-0.0823	0.5279	1.0149	0.9858

5.26 Table 3-4 (parameter estimates of passenger OR total density coefficients) expanded:

Density coefficient estimates				
	Track		**Route**	
	Separate	**Total**	**Separate**	**Total**
COLS	0.5583	0.6251	0.5938	0.6618
PSFA	0.5586	0.6251	0.5938	0.6618
COLSFE/FEI	0.0856	0.3004	0.0521	0.2792
PL	0.4077	0.5897	0.2947	0.4841
PLMUN	0.0880	0.2831	0.0572	0.2714
CUESTAQ	0.6915	0.6777	0.6645	0.6894
CUESTAL	0.6575	0.8226	0.4621	0.4572
CUESTAN	0.4353	0.4613	0.4000	0.3779
CUESTAMUN	0.0255	-0.1016	-0.0221	-0.0302
CSSFEQ	0.3613	-0.0477	0.1788	-0.1916
CSSFEL	-0.0653	-0.0114	-0.0384	-0.0751
CSSRE	0.5583	0.6251	0.5938	0.6618
CSSREO	0.5583	0.6251	0.5938	0.6618
CSSREM	0.0405	0.5814	0.0580	0.4879
PERRES	0.3446	0.5078	0.3141	0.4734
PERRESMUN	0.0833	0.3193	0.0588	0.3122
TRERE	0.3013	0.5719	N/A[24]	0.6075
TRREMUN	0.0405	0.3858	-0.0692	0.4247
COMP	0.3446	0.5078	0.3141	0.4734
COMPMUN	0.1385	0.3193	0.0588	0.3122

5.27 Table 3-5 (parameter estimates on the proportion of single track coefficients) expanded:

SING Point Estimates		
	Track	**Route**
COLS	-0.3945	-0.4932
PSFA	-0.3945	-0.4932
COLSFE/FEI	-0.9082	-1.1064
PL	-0.6421	-0.9057
PLMUN	-0.9114	-1.0792
CUESTAQ	-0.3663	-0.5074
CUESTAL	-0.5104	-0.7934
CUESTAN	-0.5382	-0.6919
CUESTAMUN	-0.3446	-0.4119
CSSFEQ	-0.3240	-0.4859
CSSFEL	-0.7895	-0.8694
CSSRE	-0.3945	-0.4932
CSSREO	-0.3945	-0.4932
CSSREM	-0.8503	-0.7602
PERRES	-0.6166	-0.8124
PERRESMUN	-0.9053	-1.0295
TRERE	-0.6128	N/A[24]
TRREMUN	-0.8503	-0.9540
COMP	-0.6166	-0.8124
COMPMUN	-0.7528	-1.0295

5.28 Table 3-7 (parameter estimates of sing and avtrack with route and track specifications) expanded:

SING/AVTRACK POINT ESTIMATES	**Track**		**Route**	
	Sing	**Avtrack**	**Sing**	**Avtrack**
COLS	-0.3945	1.0737	-0.4932	1.3873
PSFA	-0.3945	1.0865	-0.4932	1.3984
COLSFE/FEI	-0.9082	-0.8733	-1.1064	-0.1458
PL	-0.6421	0.5569	-0.9057	0.6310
PLMUN	-0.9114	-0.6991	-1.0792	-0.0836
CUESTAQ	-0.3663	0.4219	-0.5074	0.4657
CUESTAL	-0.5104	-0.0493	-0.7934	-0.0721
CUESTAN	-0.5382	0.2682	-0.6919	0.2293
CUESTAMUN	-0.3446	-1.4519	-0.4119	-0.2508
CSSFEQ	-0.3240	-2.5104	-0.4859	-0.5955
CSSFEL	-0.7895	-1.2545	-0.8694	-0.5136
CSSRE	-0.3945	1.0737	-0.4932	1.3873
CSSREO	-0.3945	1.0737	-0.4932	1.3873
CSSREM	-0.8503	0.6212	-0.7602	0.3628
PERRES	-0.6166	0.5461	-0.8124	0.8045
PERRESMUN	-0.9053	-0.4543	-1.0295	-0.0001
TRERE	-0.6128	N/A[24]	N/A[24]	N/A[24]
TRREMUN	-0.8503	0.2992	-0.9540	0.0011
COMP	-0.6166	0.5461	-0.8124	0.8045
COMPMUN	-0.7528	-0.4543	-1.0295	-0.0001

5.29 Table 3-8 (parameter estimates on network size measures with and without single track variable) expanded:

TRACK/ROUTE Point Estimates				
	Track		**Route**	
	Sing	**No Sing**	**Sing**	**No Sing**
COLS	0.9800	1.0040	1.0009	1.0024
PSFA	0.9799	1.0040	1.0009	1.0016
COLSFE/FEI	0.4860	0.3550	1.4301	0.9749
PL	0.9990	0.8985	1.0730	0.8412
PLMUN	0.4896	0.3584	1.2685	0.8364
CUESTAQ	0.8588	0.9963	0.9789	0.8893
CUESTAL	0.9745	1.0180	1.0686	1.0174
CUESTAN	0.8344	0.9174	0.8760	0.8635
CUESTAMUN	0.0400	0.1951	1.0643	1.3543
CSSFEQ	-0.0702	-0.0246	2.4637	2.3189
CSSFEL	-0.2600	-0.1493	1.2431	1.0180
CSSRE	0.9800	1.0040	1.0009	1.0024
CSSREO	0.9800	1.0040	1.0009	1.0024
CSSREM	0.4176	0.2998	-0.3074	-0.4128
PERRES	0.9618	1.0076	1.0005	0.9963
PERRESMUN	0.4826	0.3521	1.0149	0.6303
TRERE	0.9504	N/A[24]	N/A[24]	0.8811
TRREMUN	0.4176	0.3580	0.8234	N/A[24]
COMP	0.9618	1.0076	1.0005	0.9963
COMPMUN	-0.0823	0.3521	1.0149	0.6303

5.30 Table 3-9 (parameter estimates on passenger train density measures with and without single track variable) expanded:

Density coefficient estimates				
	Track		**Route**	
	Sing	**No Sing**	**Sing**	**No Sing**
COLS	0.5583	0.8402	0.5938	0.9128
PSFA	0.5586	0.8402	0.5938	0.9130
COLSFE/FEI	0.0856	-0.0168	0.0521	-0.0561
PL	0.4077	0.5142	0.2947	0.5121
PLMUN	0.0880	-0.0146	0.0572	-0.0484
CUESTAQ	0.6915	1.0031	0.6645	0.8334
CUESTAL	0.6575	0.9634	0.4621	0.9614
CUESTAN	0.4353	0.6581	0.4000	0.8133
CUESTAMUN	0.0255	0.2018	-0.0221	0.0533
CSSFEQ	0.3613	0.3493	0.1788	0.1744
CSSFEL	-0.0653	-0.0303	-0.0384	-0.0075
CSSRE	0.5583	0.8402	0.5938	0.9128
CSSREO	0.5583	0.8402	0.5938	0.9128
CSSREM	0.0405	-0.0522	0.0580	-0.0220
PERRES	0.3446	0.6286	0.3141	0.6864
PERRESMUN	0.0833	-0.0187	0.0588	-0.0434
TRERE	0.3013	N/A[24]	N/A[24]	0.5359
TRREMUN	0.0405	-0.0153	-0.0692	N/A[24]
COMP	0.3446	0.6286	0.3141	0.6864
COMPMUN	0.1385	-0.0187	0.0588	-0.0434

5.31 Table 3-10 (point estimates on the proportion of track electrified) expanded:

ELEC Point Estimates		
	Track	**Route**
COLS	-0.6284	-0.6528
PSFA	-0.6284	-0.6528
COLSFE/FEI	-0.2630	-0.4278
PL	-0.5555	-0.4631
PLMUN	-0.5373	-0.5318
CUESTAQ	-0.4558	-0.4234
CUESTAL	-0.5646	-0.4831
CUESTAN	-0.5789	-0.6830
CUESTAMUN	-0.8540	-0.3628
CSSFEQ	-0.4608	-0.2239
CSSFEL	-0.3683	-0.2124
CSSRE	-0.6284	-0.6528
CSSREO	-0.6284	-0.6528
CSSREM	-0.6383	-0.6291
PERRES	-0.4332	-0.4098
PERRESMUN	-0.4730	-0.5092
TRERE	-0.6201	-0.6429
TRREMUN	-0.6301	N/A[24]
COMP	-0.4332	-0.4098
COMPMUN	-0.4730	-0.5092

Correlation Matrix of regression variables (red indicates a value above ±0.9)

	main	*ren*	*totex*	*maincon*	*rencon*	*totexcon*	*route*	*passdr*	*frdr*	*sing*	*elec*	*singb*	*elecb*	*track*	*totdr*	*passdt*	*frdt*	*totdt*	*railtrac*	*singt*	*avtrack*	*time*	*time2*
main	1.00																						
ren	0.90	1.00																					
totex	0.97	0.98	1.00																				
maincon	0.99	0.90	0.96	1.00																			
rencon	0.90	0.99	0.97	0.90	1.00																		
totexcon	0.97	0.97	0.99	0.97	0.98	1.00																	
route	0.80	0.72	0.78	0.81	0.73	0.79	1.00																
passdr	0.36	0.40	0.39	0.32	0.37	0.36	-0.19	1.00															
frdr	0.23	0.34	0.30	0.21	0.33	0.28	0.08	0.31	1.00														
sing	-0.47	-0.50	-0.50	-0.45	-0.48	-0.48	0.06	-0.79	-0.23	1.00													
elec	0.04	0.00	0.01	0.02	-0.02	-0.01	-0.17	0.26	0.62	-0.17	1.00												
singb	-0.52	-0.55	-0.55	-0.50	-0.53	-0.53	0.02	-0.88	-0.31	0.98	-0.20	1.00											
elecb	0.02	0.00	0.01	0.00	-0.02	-0.02	-0.23	0.35	0.62	-0.25	0.99	-0.27	1.00										
track	0.89	0.81	0.87	0.89	0.82	0.88	0.98	-0.02	0.12	-0.13	-0.15	-0.18	-0.20	1.00									
totdr	0.37	0.42	0.41	0.32	0.39	0.37	-0.17	0.99	0.45	-0.77	0.33	-0.87	0.42	-0.01	1.00								
passdt	0.23	0.27	0.26	0.18	0.23	0.22	-0.29	0.98	0.32	-0.67	0.31	-0.77	0.40	-0.14	0.97	1.00							
frdt	0.01	0.11	0.07	-0.01	0.10	0.06	0.03	0.04	0.94	0.09	0.61	0.02	0.58	-0.01	0.18	0.10	1.00						
totdt	0.21	0.27	0.25	0.16	0.23	0.20	-0.29	0.94	0.50	-0.62	0.40	-0.72	0.49	-0.15	0.96	0.98	0.30	1.00					
railtrac	0.21	0.24	0.24	0.20	0.23	0.22	0.18	0.10	-0.12	-0.19	-0.28	-0.18	-0.25	0.22	0.07	0.03	-0.21	-0.02	1.00				
singt	-0.53	-0.55	-0.55	-0.51	-0.54	-0.54	0.00	-0.80	-0.23	0.99	-0.13	0.99	-0.21	-0.20	-0.79	-0.67	0.11	-0.61	-0.21	1.00			
avtrack	0.63	0.65	0.65	0.62	0.64	0.64	0.16	0.78	0.19	-0.90	0.03	-0.93	0.10	0.36	0.76	0.62	-0.17	0.56	0.27	-0.95	1.00		
time	0.01	0.13	0.07	-0.01	0.12	0.06	-0.03	0.12	-0.02	-0.08	0.09	-0.06	0.09	-0.01	0.11	0.12	-0.04	0.11	-0.20	-0.08	0.08	1.00	
time2	0.01	0.13	0.07	-0.01	0.12	0.05	-0.03	0.11	-0.02	-0.08	0.10	-0.07	0.10	-0.02	0.11	0.12	-0.05	0.11	-0.17	-0.08	0.07	0.97	1

6. Annex B

Data Changes:

Construction PPP

6.1 Both Network Rail and Oxera recommended that we consider alternative measures of PPP. In particular, Network Rail stressed that construction PPP might be a better measure of comparative price levels than the "GDP" PPP measures used in PR08. As a sensitivity test we have considered the effects of construction PPP on parameter estimates and efficiency scores.

6.2 The construction PPP data used, which is taken from Eurostat and is the best available, only goes back as far as 1999. Our sensitivity tests using this data adjusted for this in two ways; the first was to use the 1999 PPP figures for 1996-1998 (whilst not as robust has the advantage of retaining observations in the sample); the second was to exclude all 1996-1998 observations from the sample.

Results

6.3 While parameter estimates using construction PPP were broadly similar, the effect on efficiency estimates was significant, as Table 6-1 indicates.

	1996-2010		**1999-2010**	
	Construction PPP	**GDP PPP**	**Construction PPP**	**GDP PPP**
COLS (Adjusted)	36.3%	23.2%	32.5%	19.6%
CUESTAL	47.6%	24.4%	9.6%	18.5%
CUESTAN	45.4%	23.3%	25.1%	14.1%
CSSRE	35.7%	12.9%	23.5%	5.7%
Average	**41.2%**	**20.9%**	**22.7%**	**14.4%**
Max	*47.6%*	*24.4%*	*32.5%*	*19.6%*
Min	*35.7%*	*12.9%*	*9.6%*	*5.7%*
Range Size	*11.9%*	*11.5%*	*22.9%*	*13.9%*

Table 6-1: Network Rail 2010 efficiency gap when expenditure levels adjusted by alternative PPP measures

6.4 As well as worsening Network Rail's 2010 efficiency gap estimates, the size of the range of the gap also increases when using construction PPP. This effect was also observed across countries and over time. Notably, efficiency gap estimates generally worsened for all time periods for almost all countries across every model – a possible explanation of this may be an improvement in costs for country E as a result of switching to construction PPP; since construction PPP makes them look less costly and since they set the frontier in most models even in the standard specification, it is not surprising that this results in a general worsening in efficiency gap estimates.

6.5 Owing to the nature of the compilation of PPP data, its precision increases with the level of aggregation, and Eurostat state that they expect measures of PPP based on GDP to be more accurate than those based on construction.

6.6 Since the most defensible use of construction PPP would mean excluding three years of data, since we have good reasons to have doubts about the relative quality of measures of PPP based on construction and since, in any case, adjusting by GDP PPP is the more conservative regulatory approach (on the basis of narrower ranges of efficiency scores and lower efficiency gaps), we view the use of GDP PPP in this review as warranted.

No Steady State Adjustment

6.7 In Section 2 we discussed the steady state adjustment that was applied to Network Rail's track renewal expenditure, in line with the view that Railtrack under-renewed and Network Rail has had to clear this renewals backlog on top of steady state renewals volumes. In this section we look at the effect of removing the steady state adjustment on Network Rail's efficiency gap.

6.8 Removal of the steady state adjustment, in general, reduces Network Rail's reported total expenditure for earlier years, and increases it in the immediate post-Railtrack years, so we would generally expect to find that the efficiency gap increases when the steady state adjustment is removed.

Results

6.9 Table 6-2 indicates, as expected, that Network Rail's 2010 efficiency gap increases when the steady state adjustment is removed. The linear model produces the most significant increase in the 2010 efficiency gap, which is as one would expect – as the post-Railtrack years are made more costly the gradient should, and does, decrease as the model attempts to account for a more costly middle period. All models give a higher 1996 efficiency score using the un-adjusted data, with the CSSRE model actually putting Network Rail on the frontier. Notably the COLS model is more efficient in 2010 when the steady state adjustment is removed; this is relatively easily understood as Network Rail under-renewed in 2010 relative to trend, so that the steady state adjustment increases their reported costs for that year, and because these models do not assume any correlation in inefficiency between periods, they are not forced to put a time path of inefficiency that moves in a particular direction.

	No Steady State	Preferred
COLS (Adjusted)	18.4%	23.2%
CUESTAL	32.3%	24.4%
CUESTAN	24.5%	23.3%
CSSRE	15.1%	12.9%
Average	**22.6%**	**20.9%**
Max	*32.3%*	*24.4%*
Min	*15.1%*	*12.9%*
Range Size	*17.2%*	*11.5%*

Table 6-2: Network Rail 2010 efficiency gap without and with steady state adjustment

6.10 The results of removing the steady state adjustment are consistent with our expectations and the fact that several models find Railtrack to be on or near the frontier lend more support to our decision in this and the previous Periodic Review to implement a steady state adjustment.

Alternative Steady State Adjustments

Steady State Rate:	2.3 (Preferred)	2	2.5	2.65	3
COLS	23.2%	22.2%	23.9%	24.3%	25.3%
CUESTAL	24.4%	22.3%	25.7%	23.6%	26.0%
CUESTAN	23.3%	19.9%	23.8%	26.0%	28.4%
CSSRE	12.9%	10.8%	14.2%	15.2%	17.3%
Average	**20.9%**	**18.8 %**	**21.9%**	**22.3%**	**24.3%**

Table 6-3: Network Rail 2010 efficiency gap with alternative steady state adjustments

No Early Year Data

6.11 The table below shows how sensitive Network Rail's 2010 efficiency gap estimate is to the removal of the early years of the data from the sample. It is included for reference as a standard sensitivity check; we have no reason for supposing that early year data is less accurate than data in the later years and the results (across other IMs as well as Network Rail) do not support any decision to remove earlier years from the sample.

	Full Sample	From 1997	From 1998	From 1999	From 2000	From 2001	From 2002	From 2003
COLS	23.2%	22.0%	20.9%	19.6%	18.0%	17.3%	18.1%	19.0%
CUESTAL	24.4%	20.3%	18.9%	18.5%	16.0%	14.5%	16.0%	16.3%
CUESTAN	23.3%	18.7%	14.5%	14.1%	14.7%	18.2%	20.8%	16.1%
CSSRE	12.9%	7.3%	4.9%	5.7%	5.9%	11.1%	19.0%	18.4%

Table 6-4: Network Rail 2010 efficiency gap when early years removed

Country E's Maintenance Expenditure

6.12 As discussed in Annex A, there could be a discontinuity in the data for country E pre-2010. The IM states that the 2010 submission is correct, but the previous values may have been provided under incorrect definitions and that there is no way to correct this. To assess the impact of this possible discontinuity we looked at the effect of back-adjusting pre-2009 maintenance expenditure by the same proportion as the 2010 to 2009 fall (i.e. assuming that any incorrect reporting had the same effect in each year). Table 6-5 shows Network Rail's 2010 efficiency gap under this sensitivity. However, as we do not have a good evidence base that there were no changes in underlying activities between 2009 and 2010, or that any definitional error is proportionate to maintenance values (rather than absolute for example), we view this back adjustment as being better presented as a sensitivity test rather than forming part of our central case. The results do indicate though that our estimates may be conservative with regard to Network Rail's overall efficiency gap.

	Preferred	Pre-2010 Adjusted
COLS (Adjusted)	23.2%	33.6%
CUESTAL	24.4%	55.4%
CUESTAN	23.3%	44.1%
CSSRE	12.9%	20.1%
Average	**20.9%**	**38.3%**
Max	*24.4%*	*55.4%*
Min	*12.9%*	*20.1%*
Range Size	*11.5%*	*35.4%*

Table 6-5: Network Rail 2010 efficiency gap with adjustments to country E's maintenance expenditure

6.13 As we should expect, back-adjusting maintenance expenditure (by reducing all pre-2009 maintenance expenditure by 40%) results in a significant worsening of Network Rail's efficiency gap – since (by most models) country E already sets the frontier with their pre-adjusted cost levels, the significant reduction in their costs thereby increases Network Rail's distance from the most efficient IM.

Monte Carlo Analysis

6.14 In response to a recommendation by Jon Stern in his 2012 review of top down benchmarking, we have also looked at implementing a simple Monte Carlo analysis to help indicate the potential impact data uncertainty might have on our results. To do this we have applied a 5% uncertainty factor to each observation in our dataset (so up to +2.5% or -2.5% from the actual value) using a uniform distribution, with appropriate bounds to ensure variables remain within possible limits (so no negative values for proportion of single track for example). The aim of this analysis is to address a Network Rail concern that the data may be poorly reported by some countries, and so to understand how this may impact upon the efficiency estimates produced. In total 300 replications were taken for each model and the table below presents the central outputs from the models, along with intervals produced by the Monte Carlo simulation at the 20%, 50% and 90% likelihood level. The impacts of additional uncertainty on the efficiency score of a particular country is complex, but an explanation for the pattern seen below could be that additional data uncertainty is likely to reduce model fit, and so increase the size of the residual, all else being equal. This increased residual then needs to be allocated between noise and inefficiency. If some of this is allocated to inefficiency we would expect increased noise to increase the efficiency gap estimate, all else being equal.

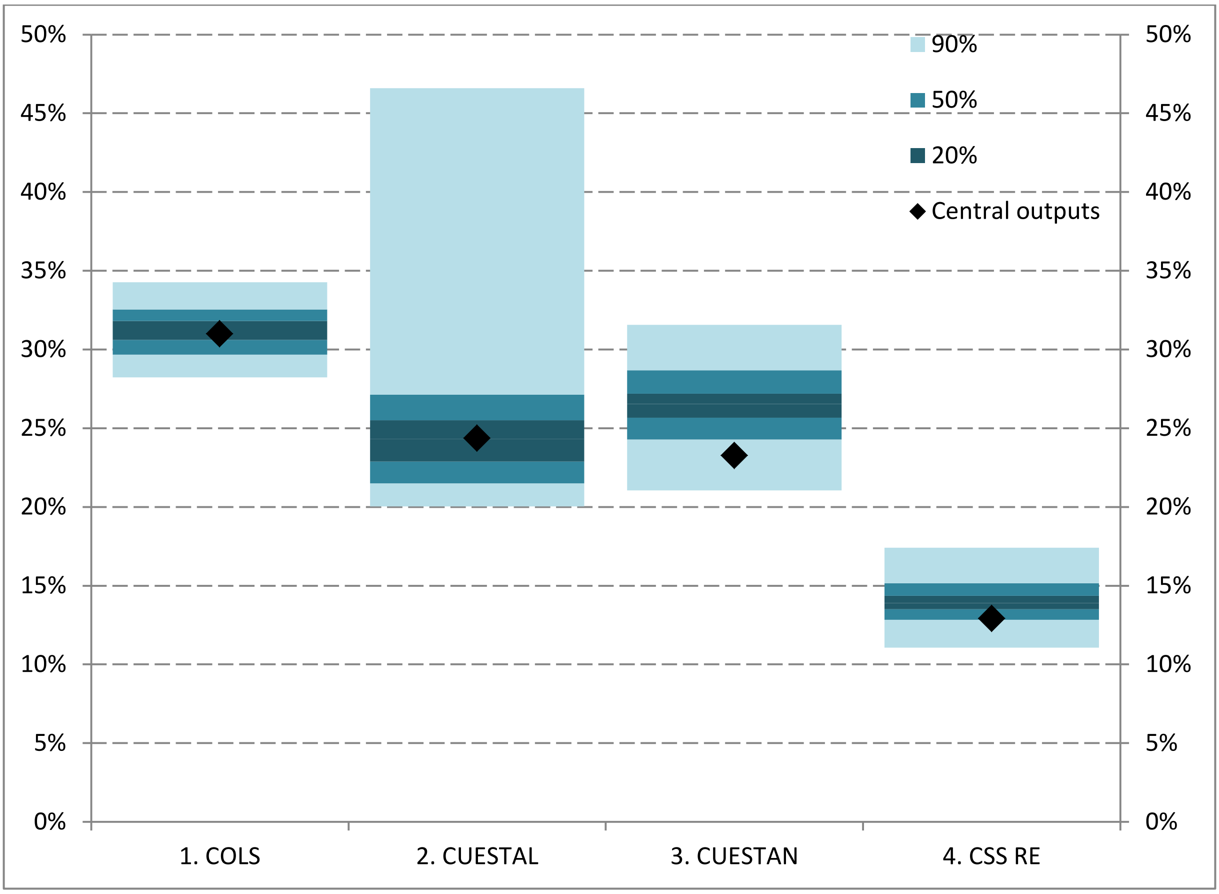

Figure 6-1: Network Rail 2010 efficiency gap Monte Carlo analysis - note that 25% adjustment to the COLS model is not included here.

The effect of systematic misreporting on efficiency scores

6.15 In addition to the Monte Carlo analysis above, we have also looked at possible impacts of any systematic misreporting by IMs. We looked at both the effects of “consistent” misreporting by the same amount on every variable, as well as misreporting only on total expenditure. For instance, when looking at total expenditure only, we looked at the effects of increasing/decreasing the expenditure of all IMs (except Network Rail) by 5% and 10% (effectively assuming that all IMs were under/over stating their costs by the same proportion). Table 6-6 shows the impact on Network Rail’s efficiency gap estimates.

		Change in all variables		Change in total expenditure only			
	Preferred	**+5%**	**-5%**	**+5%**	**+10%**	**- 5%**	**-10%**
COLS (Adjusted)	23.2%	23.3%	23.1%	22.1%	21.1%	24.4%	25.5%
CUESTAL	24.4%	25.2%	23.5%	21.2%	18.1%	24.8%	28.4%
CUESTAN	23.3%	22.7%	22.6%	20.0%	17.6%	24.9%	28.9%
CSSRERE	12.9%	13.0%	12.6%	10.4%	7.9%	15.5%	18.1%
Average	**20.9%**	**21.1%**	**20.4%**	**18.4%**	**16.2%**	**22.4%**	**25.2%**
Max	*24.4%*	*25.2%*	23.5%	*22.1%*	*21.1%*	*24.9%*	*28.9%*
Min	*12.9%*	*13.0%*	12.6%	*10.4%*	*7.9%*	*15.5%*	*18.1%*
Range Size	*11.5%*	*12.1%*	10.8%	*11.8%*	*13.2%*	*9.4%*	*10.8%*

Table 6-6: Network Rail 2010 efficiency gap with systematic misreporting by all other IMs

6.16 If all infrastructure managers other than Network Rail were to overstate or understate all of their submitted data by ±5% the effect on efficiency scores would be relatively minimal – this is as we would expect, so long as the cost function is near enough to constant returns to scale, increases in expenditure should be offset by proportionate increases in the explanatory variables. In contrast, if all IMs, with the exception of Network Rai,l were to understate or overstate total expenditure by 5% or 10% this would have some impact on efficiency scores, although it is important to note that this effect would still be limited - a 10% change in the total expenditure of all other IMs would change Network Rail's 2010 mean or median score by less than 5 percentage points.

6.17 Network Rail has also suggested that even with common definitions in place in the LICB dataset there could be remaining differences in how data is defined, collected, and reported between countries, which could reduce comparability. We looked at the effects of selectively varying the expenditure of each of these IMs to match some of Network Rail's suggestions. We considered increasing both renewals and total expenditure by 5% and 10% for all those IMs for which an increase was suggested, but we only considered a 5% reduction for those IMs for which a decrease in renewals was suggested (Network Rail has maintained that the decrease is only slight). It is worth noting two points: Firstly, since, on average, renewals expenditure accounts for about half of total expenditure across the LICB, if we assume that all of the misreporting of total expenditure is due to misreporting of renewals expenditure (which has generally been claimed to be the problematic area) then a 5% increase in total expenditure is equivalent, on average, to a 10% increase in renewals expenditure. Secondly, we adjust both country E's renewals and total expenditure, so that we are effectively increasing their costs twice (in line with Network Rail's view of misreporting). This would impact on Network Rail's efficiency gap as follows:

	Preferred	**+5%**	**+10%**
COLS (Adjusted)	23.2%	21.6%	21.7%
CUESTAL	24.4%	20.4%	20.4%
CUESTAN	23.3%	20.6%	9.1%
CSSRE	12.9%	9.9%	6.8%
Average	**20.9%**	**18.1%**	**14.5%**
Max	*24.4%*	*21.6%*	*21.7%*
Min	*12.9%*	*9.9%*	*6.8%*
Range Size	*11.5%*	*11.7%*	*14.9%*

Table 6-7: Network Rail 2010 efficiency gap with selective mis-reporting

6.18 As before, even with a 10% increase, the effects on Network Rail's efficiency gap are limited. Furthermore, what effects are observed may be a consequence of our adjusting country E's expenditure twice (at the 10% rate we effectively increase their total expenditure by approximately 15%) – as country E sets the frontier, an increase in their costs, all else considered, is likely to improve Network Rail's efficiency gap.

Model Specification Changes:

Dropping time as an explanatory variable

6.19 Time variables were included on the basis of the economic rationale to act as a proxy for technical change. Since the parameter estimates on the time and time squared explanatory variables are generally only significant at above the 10% level across most models, we have looked at the effects on efficiency gap of dropping time from the model specification entirely.

Results

6.20 Table 6-8 compares the results of dropping both the time and time squared variables on 2010 efficiency gap estimates for Network Rail. While the average is the same, some models are more affected than others. The CUESTAL and CUESTAN models show much more variation than the other models, and this pattern is replicated in the results for other countries and over time. The Cuesta models in general are the most sensitive to variation in model specification, so this may just be a consequence of their typical sensitivity. Alternatively, it may be down to some interaction between the time terms and the time trend imposed on inefficiency.

	No Time	Preferred
COLS (Adjusted)	23.2%	23.2%
CUESTAL	26.8%	24.4%
CUESTAN	20.1%	23.3%
CSSRE	35.2%	12.9%
Average	**26.3%**	**20.9%**
Max	*35.2%*	*24.4%*
Min	*20.1%*	*12.9%*
Range Size	*15.1%*	*11.5%*

Table 6-8: Network Rail 2010 efficiency gap without and with time as explanatory variable

Testing for a structural break between Network Rail and Railtrack

6.21 Oxera recommended that we test for the possibility of a structural break in cost functions between Railtrack and Network Rail. To do this we included a dummy variable set to one for the years 1996 to 1999 for Network Rail, and zero otherwise. Given the steady state adjustment to some extent may compensate for any structural break in the data we ran the models with the dummy variable included on the data with and without a steady state adjustment. The parameter estimates on this variable are given in Table 6-9.

Model	**Preferred**	**No Steady State Adjustment**
COLS	1.0416	1.0167
CUESTAL	1.1126	1.0672
CUESTAN	1.1188	1.1099
CSSRE	1.0416	1.0167

Table 6-9: Parameter estimates on Railtrack dummy variable
(Green – significant at 1% level; Yellow – significant at 5% level; Red – significant at 10% level)

6.22 There does not seem to be good evidence for accepting a structural break between Network Rail and Railtrack. Only one of the models (CUESTAL) supports such a position while the remainder generally find the coefficient on the Railtrack dummy to be statistically insignificant. Given that we have already controlled for an acknowledged structural difference between periods in the sample (via the steady state adjustment) and given that the majority of models do not find statistically significant evidence of a structural break, there does not seem to be a strong basis for carrying a structural break into the cost function itself.

6.23 It is worth also noting the effect of including the Railtrack dummy on the efficiency gap estimates, as Table 6-11 indicates. Note that the most dramatic effect is on CUESTAN in the sample with the steady state adjustment – which, as previously discussed, is generally more sensitive to changes in model specification.

	Standard		No Steady State Adjustment	
	With RT Dummy	**Preferred**	**With RT Dummy**	**Preferred**
COLS (Adjusted)	23.7%	23.2%	17.1%	18.4%
CUESTAL	22.3%	24.4%	32.1%	32.3%
CUESTAN	4.1%	23.3%	24.4%	24.5%
CSSRE	13.7%	12.9%	15.3%	15.1%
Average	**15.9%**	**20.9%**	**22.2%**	**22.6%**
Max	*23.7%*	*24.4%*	*32.1%*	*32.3%*
Min	*4.1%*	*12.9%*	*15.3%*	*15.1%*
Range Size	*19.6%*	*11.5%*	*16.8%*	*17.2%*

Table 6-10: Network Rail 2010 efficiency gap with and without Railtrack dummy/steady state adjustment

Squared Terms and Cross Products

6.24 A further sensitivity that has been undertaken has been to test squared terms for each variable and cross products for significance. This involved starting with all these terms and then iterating down to remove insignificant or implausible elements. Through this process we were ultimately only left with an interaction term between freight and passenger train densities. We do not view there being a strong reason as to why this additional variable is required, and note that it may be substituting for freight density in some models (its inclusion drives freight density negative in the COLS model for example). As such we do not view there to be a strong case to include this variable, and note it could to some extent also proxy for IM specific effects. Its impact on efficiency gap for most of our preferred models is small, with the exception of the CUESTAN model, which no longer appears credible.

	With interaction term	Preferred
COLS (Adjusted)	22.8%	23.2%
CUESTAL	23.4%	24.4%
CUESTAN	6.7%	23.3%
CSSRE	12.4%	12.9%
Average	**16.3%**	**20.9%**

Table 6-11: Network Rail 2010 efficiency gap with interaction terms

COLS noise adjustment

6.25 Given that an assumption that 25% of the COLS model residuals can be attributed to noise is to some extent necessarily subjective, we consider the impact of two alternative assumptions on our overall inefficiency estimates, one at 10% and one at 50%. The overall impact of these alternatives on the mean across models is small, at 22% or 19% rather than 21%. Looking at the inner range from these models (which is as close to a ‘median’ estimate as is possible), sees the 23% in the preferred model move to a range of 23%-24% in the lower adjustment case, or 16%-23% with the higher adjustment.

	COLS - 10%	COLS -25% (Preferred)	COLS - 50%
COLS	27.9%	23.2%	15.5%
CUESTAL	24.4%	24.4%	24.4%
CUESTAN	23.3%	23.3%	23.3%
CSSRE	12.9%	12.9%	12.9%
Average	**22.1%**	**20.9%**	**19.0%**
Max	*27.9%*	*24.4%*	*24.4%*
Min	*12.9%*	*12.9%*	*12.9%*
Range Size	*15.0%*	*11.5%*	*11.5%*

Table 6-12: Network Rail 2010 efficiency gap with alternative COLS noise adjustments

Quadratic terms for other countries

6.26 Quadratic terms for all countries have already been considered in the CUESTAQ model, which was rejected on the basis that it was extremely sensitive to small variations in the dataset (such as the omission of single observations, or certain years, as set out in paragraphs 4.22 to 4.27). As an additional sensitivity we have also considered two ‘partial’ quadratic models, where quadratic terms have only been retained for those countries for which in the CUESTAQ specification they were significant. These countries (and the probability that the parameter estimate on the squared term is zero) are as follows: country G (<10%), country F (<5%), country D (<5%), and Network Rail (<1%).

6.27 Given that the results of country G are less significant than the others, we present two sensitivity tests below, one including them, and one without. We also present the CUESTAN and CUESTAQ models for comparison purposes. It is worth noting that when both additional tests are undertaken only Network Rail and country D have significant parameter estimates associated with the squared term. As such we also present a further test with only squared terms for them included.

	Inefficiency estimates
CUESTA – Countries G,F, D,NR	21.5%
CUESTA – Countries F,D,NR	27.0%
CUESTA – Countries D,NR	27.0%
CUESTAN	23.3%
CUESTAQ	18.8%

Table 6-13: Network Rail 2010 efficiency gap with variations on Cuesta model

Model	Constant	TRACK	PASSDT	FRDT	SING
CUESTA - Countries G,F, D,NR	-0.4185	0.8208	0.6063	0.0829	-0.4529
CUESTA – Countries F,D,NR	-0.4378	0.7988	0.5564	0.1265	-0.4637
CUESTA – Countries D,NR	-0.4378	0.7988	0.5564	0.1265	-0.4637
CUESTAQ	-0.4240	0.8588	0.6915	0.0523	-0.3663
CUESTAN	-0.4664	0.8344	0.4353	0.0686	-0.5382

Table 6-14: Parameter estimates across variations in Cuesta model

6.28 The specification including a quadratic term for country D and Network Rail does produce plausible parameter estimates, and one where the term on freight density coefficient is significant. It does also produce estimates for Network Rail’s efficiency gap which are higher than that produced for the other models considered, and also exhibits greater sensitivity to changes to data. For example if 1996 is dropped from the dataset, the model estimates a change in efficiency score of around 10 percentage points. As such we have not carried this model forward into our final preferred set of models, but do note that this alternative may indicate that our results with our preferred set of models are slightly conservative.