Recommendations on how to model efficiency for future price reviews
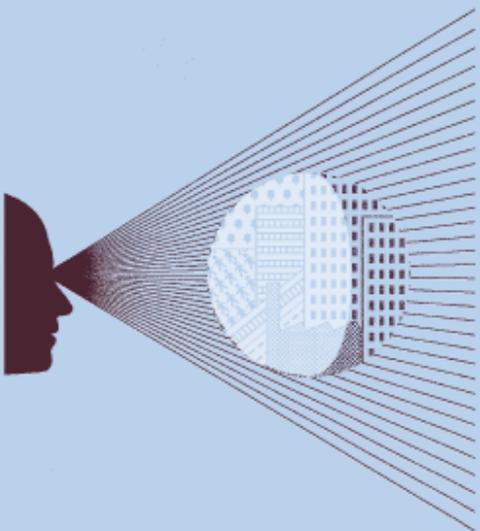
Prepared for
Office of Rail Regulation

November 2009

Oxera

Contents

1 Introduction

1.1 Scope

The Office of Rail Regulation (ORR) has asked Oxera to review the econometric models used to undertake international benchmarking for the 2008 Periodic Review (PR08) of Network Rail's access charges, with a view to providing recommendations on methodological issues and improvements for future price control periods.

The scope of this review is as follows:

– the choice of econometric models used in PR08 (focusing on the preferred model), including the functional form of the models and their statistical robustness;

– an assessment of the model variables (including omitted variables);

– the treatment of uncertainty.

Outside of the scope of this review are the following:

– providing comments on the ORR's use of the results of the econometric analysis in developing its overall judgements on the scope for efficiency improvement in Control Period 4 (CP4);

– providing comments on any other efficiency analysis, including the bottom-up analysis, undertaken by the ORR and its consultants in PR08;

– undertaking a detailed comparison or check of the sources of the data, their definitions and accuracy in terms of an assessment of on data quality and model variables. (Instead, the review focuses on issues relevant to the econometric modelling based on the material currently available);

– providing an exhaustive and detailed set of empirical recommendations pertinent to the next review. As the dataset used for the review develops further—indeed, as theoretical and software advances are made—the appropriate approach(es) will alter. As such, Oxera's recommendations are generally somewhat higher-level and more generic than what might be carried out during future price reviews, and should be developed further as the review progresses.

1.2 Objective

The objective of this report is to provide a basis from which to develop efficiency modelling used to make efficiency assumptions in future price control periods. As such, the report does not focus on detailed or specific modelling issues undertaken to estimate Network Rail's efficiency gap for PR08, but rather takes the PR08 analysis, together with the issues raised by Network Rail's consultants in their commentary on the modelling undertaken by the Institute for Transport Studies (ITS),[1] as a starting point to provide broad recommendations

---

[1] Smith, A., Wheat, P. (ITS) and Nixon, H, (ORR) (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs', June. Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October; LECG (2008), 'Assessment of the ITS/ORR Econometric Models', September; Horton 4 Consulting Ltd (2008), 'International Efficiency Benchmarking in the Network Rail 2008 Periodic Review', July. Smith, A. and Wheat, P. (2008), 'A Response to the LECG and Horton 4 Consulting Reports on the ITS/ORR International Benchmarking Study (Based on the LICB dataset)', October.

for undertaking efficiency comparisons of Network Rail going forward. Indeed, in the future modelling exercises, a large variety of issues would depend on the empirical outcome, and can only be reviewed during the implementation stage.

1.3 Evidence base

The evidence base used as part of this report is a desktop review of published reports[2] and data provided by ITS, and workshops and semi-structured interviews involving the ORR, ITS and Network Rail held in October 2009.

1.4 Summary of recommendations

Overall, Oxera considers that international benchmarking is a valid approach to assessing the cost efficiency of Network Rail and for the ORR's use in establishing efficiency assumptions as part of future price reviews.

Oxera's key recommendations are summarised below, indicating the sections in the report that set out the basis for the recommendations. As stated above, the following represents Oxera's recommendations for the ORR's efficiency modelling for future price control reviews. It builds on the work undertaken by the ORR, Network Rail and their respective consultants at PR08. As such, these recommendations should not be seen as Oxera commenting on the analysis undertaken in PR08, which has already tackled some of the issues discussed in this report but is now in need of updating.

Summary of recommendations

Context and process (section 2): recommendations
2.1 The context of the comparative analysis should be taken into account. That is, the added uncertainty and complexity involved in undertaking international comparisons should be allowed for. The modelling needs to test for the possibility of differences in technology; modelling techniques that are able to accommodate uncertainties in data should be used; and the results need to be tested for sensitivity and used with appropriate caution when forming part of a regulatory price review.
2.2 A well-structured process to undertake the efficiency analysis should be developed and followed. In particular, a wide range of modelling approaches and model specifications needs to be investigated before selecting a preferred model or synthesising the results.
2.3 The efficiency analysis should be undertaken on an annual basis and shared with Network Rail. This would assist in providing transparency and early indications of performance, and in enabling progress to be monitored. The ORR could go beyond sharing the analysis and work closely with Network Rail on the modelling.[1]
2.4 The ORR's approach should be continually developed and empirically tested. As the dataset used for the review develops, and as theoretical and software advances are made, the appropriate approach(es) will change.

Notes: [1] Indeed, it is Oxera's understanding that it is the intention to adopt this approach for the next review.

---

[2] Smith, A., Wheat, P. (ITS) and Nixon, H, (ORR) (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs', June. Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October; LECG (2008), 'Assessment of the ITS/ORR Econometric Models', September; Horton 4 Consulting Ltd (2008), 'International Efficiency Benchmarking in the Network Rail 2008 Periodic Review', July. Smith, A. and Wheat, P. (2008), 'A Response to the LECG and Horton 4 Consulting Reports on the ITS/ORR International Benchmarking Study (Based on the LICB dataset)', October.

Accounting for differences between companies and countries (section 3): recommendations

3.1	More work is required to extend the dataset and further improve consistency. Subsequent adjustments to the raw data should be clearly documented and audited, shared for comments and feedback, with the final dataset made available to all parties on an annual basis prior to the modelling being undertaken. Outlier tests should be undertaken to check for any anomalies that may warrant further investigation.
3.2	The efficiency analysis should continue to use panel data because this improves the accuracy of the analysis, as well as enabling unobserved company- or country-specific factors to be accounted for.
3.3	Renewals and maintenance costs are likely to be driven by, among other factors, required levels of activity and output (performance, safety and availability) requirements. A greater understanding of the impact of these factors is important in order to measure inefficiency accurately.
3.4	Differences in input prices need to be captured appropriately in the modelling. The robustness of the results to alternative purchasing power parity (PPP) measures and real exchange rates should be examined and the limitations and uncertainties involved in using PPPs recognised.[1] In addition, further effort should be devoted to accounting for differences in industry input prices.
3.5	Differences between countries and the possibility of differences in the technology used should be captured in the modelling. This includes testing for whether the technology used by different companies is the same, and potentially allowing for different technology frontiers for the companies.
	Some differences between countries and companies (known in other regulated industries as 'special factors') cannot be fully captured in the modelling. These warrant quantification by Network Rail and subsequent review, and, if deemed appropriate, should be accounted for by the ORR when developing its efficiency assumptions.
3.7	Structural breaks should be examined and, where relevant, captured in the modelling.

Notes: [1] It is Oxera's understanding that the ORR is working towards a different treatment of international monetary data via real exchange rates.

Modelling approach (section 4): recommendations

4.1	Several modelling approaches should be used to cross-check the robustness of the results.
4.2	The use of semi-parametric approaches should be investigated and the results contrasted with other approaches.
4.3	An approach should be developed to synthesise the results from these alternative approaches.

Specific modelling issues in PR08 analysis (section 5): recommendations

5.1	Substitution between cost categories should be captured where possible.
5.2	The appropriate functional form of the cost model requires detailed examination and testing.
5.3	For stochastic frontier analysis (SFA) models, alternative assumptions for the distribution of noise and inefficiency should be examined.
5.4	Explaining the efficiency term through explanatory factors assists in understanding inefficiency and improves the robustness of the modelling, and should be examined in further detail.
5.5	Modelling the time profile of inefficiency assists in understanding inefficiency and improves the robustness of the modelling, and should be examined in further detail.
5.6	A range of statistical and economic tests needs to be carried out to develop a robust cost model.

Accounting for the uncertainty of the results (section 6): recommendations

6.1	The degree of uncertainty in the modelling should be quantified.
6.2	Given the (quantified and unquantified) uncertainty, the ORR should use the modelling results with an appropriate degree of caution when making efficiency assumptions.

1.5 Structure of the report

The report is structured as follows.

–   Section 2 sets out the context for the international efficiency analysis undertaken at PR08, and provides an overview of the process of undertaking a benchmarking exercise in the context of developing efficiency assumptions.

–   Section 3 discusses the importance of the impact of using comparisons between countries and how to account for differences between them.

–   Section 4 looks at different modelling approaches.

–   Section 5 considers some of the specific modelling issues surrounding the PR08 analysis.

–   Section 6 explores the uncertainty of the efficiency results.

2   Context and process

> **Recommendation 2.1**: The context of the comparative analysis should be taken into account. That is, the added uncertainty and complexity involved in undertaking international comparisons should be allowed for. The modelling needs to test for the possibility of differences in technology; modelling techniques that are able to accommodate uncertainties in data should be used; and the results need to be tested for sensitivity and used with appropriate caution when forming part of a regulatory price review.
>
> **Recommendation 2.2**: A well-structured process to undertake the efficiency analysis should be developed and followed. In particular, a wide range of modelling approaches and model specifications needs to be investigated before selecting a preferred model or synthesising the results. (A suggested process is set out in Figure 2.1 below. This should be further developed as the ORR continues to develop its approach for the next review).
>
> **Recommendation 2.3**: The efficiency analysis should be undertaken on an annual basis and be shared with Network Rail. This would assist in providing transparency and early indications of performance and in enabling progress to be monitored. The ORR could go beyond sharing the analysis and work closely with Network Rail on the modelling.[3]
>
> **Recommendation 2.4**: The ORR's approach should be continually developed and empirically tested. As the dataset used for the review develops and as theoretical and software advances are made, the appropriate approach(es) will change.

Unlike many other regulated companies in Great Britain, Network Rail does not have any comparators within GB. In order to assess Network Rail's efficiency, along with other evidence, an international benchmarking study was used for PR08 to assess the performance of Network Rail's maintenance and renewal activities relative to other European operators. This benchmarking was based on econometric analysis in order to take into account some of the differences between the operators. The econometric analysis made use of data collected from rail operators in 13 European countries over the period 1996 to 2006.[4] Such international econometric benchmarking represents one of the few such studies by a UK regulator (the other notable example being Ofcom's assessment of BT).[5] Alongside other evidence, the ORR used the results from the econometric study in developing its efficiency assumption for Network Rail for PR08.[6]

Any comparative efficiency modelling needs to take into account the context and follow an appropriate process and methodology. Given the context discussed above, the key issue that will affect the modelling approach in a review of Network Rail's efficiency is that the comparator set is international (the impact of this is discussed in detail in section 3).

Given this context, Figure 2.1 outlines the workflow and decisions required in moving from the initial conceptualising of efficiency measurement to an efficiency assumption for use in a

---

[3] Indeed, it is Oxera's understanding that it is the intention to adopt this approach for the next review.

[4] The countries covered by the Lasting Infrastructure Cost Benchmarking (LICB) dataset are the UK, Netherlands, Norway, Portugal, Finland, Sweden, Ireland, Belgium, Germany, Austria, Italy, Denmark and Switzerland. More details of the dataset are provided in section 3.

[5] NERA (2008), 'The Comparative Efficiency of BT Openreach: A Report for Ofcom', March.

[6] ORR (2008), 'Periodic Review 2008: Determination of Network Rail's Outputs and Funding for 2009-14', October.

regulatory setting. Most of these steps are further discussed in the corresponding sections, indicated in the figure.

Figure 2.1   Comparative efficiency modelling process

Source: Oxera

While the focus of this report is to provide recommendations on elements of the modelling to assess Network Rail's efficiency going forward, it is also important to examine some aspects relevant to the overall process.

An important aspect of the process is the frequency of the analysis. Namely, should the analysis be undertaken only at each price review? Oxera considers that Ofwat's approach of publishing annual efficiency reports assists in providing transparency, early indications of estimated performance relative to peers and thus the likely position at the forthcoming reviews, and enables the regulator and the companies to monitor their progress.[7] Such annual information is of considerable value to the companies, and detailed comparisons can be useful for operational purposes in terms of identifying potential for improvement. Indeed,

---

[7] See, for example, Ofwat (2008), 'Relative Efficiency Assessment 2007-08: Supporting Information', January.

while the ORR might want to rely on higher-level comparisons, Network Rail might want to develop more detailed comparisons in order to gain additional operational insights. Undertaking the efficiency analysis on an annual basis and publishing the results—or, at least as a minimum, sharing the results with Network Rail—would result in similar benefits accruing to the GB rail industry. Indeed, it is Oxera's understanding that a joint approach is being adopted going forward.

Efficiency is a dynamic concept. Companies catch up to the efficiency frontier and convergence occurs. Equally, companies can leapfrog the frontier and divergence occurs. In addition, over time, it is possible to extend the dataset being used with additional variables and data over additional years. Similarly, continual theoretical developments and software developments imply that what may have been best practice at a particular point in time can quickly be superseded. As such, Oxera would recommend that the ORR continually reviews and develops its approach.

3 Accounting for differences between countries

Recommendation 3.1: More work is required to extend the dataset and further improve consistency; subsequent adjustments to the raw data should be clearly documented and audited, shared for comments and feedback, with the final dataset made available to all parties on an annual basis prior to the modelling being undertaken. Outlier tests should be undertaken to check for any anomalies that may warrant further investigation.

Recommendation 3.2: The efficiency analysis should continue to use panel data because this improves the accuracy of the analysis, as well as enabling unobserved company- or country-specific factors to be accounted for.

Recommendation 3.3: Renewals and maintenance costs are likely to be driven by, among other factors, required levels of activity and output (performance, safety and availability) requirements. A greater understanding of the impact of these factors is important in order to measure inefficiency accurately.

Recommendation 3.4: Differences in input prices need to be captured appropriately in the modelling. The robustness of the results to alternative PPP measures and real exchange rates should be examined and the limitations and uncertainties involved in using PPPs recognised. [8] In addition, differences in industry input prices should be accounted for.

Recommendation 3.5: Differences between countries and the possibility of differences in the technology used should be captured in the modelling. This includes testing for whether the technology used by different companies is the same, and potentially allowing for different technology frontiers for the companies.

Recommendation 3.6: Some differences between countries and companies (known in other regulated industries as 'special factors') cannot be captured in the modelling. These warrant quantification by Network Rail and subsequent review, and, if deemed appropriate, should be accounted for by the ORR when developing its efficiency assumptions.

Recommendation 3.7: Structural breaks should be examined and, where relevant, captured in the modelling.

3.1 Data issues

Efficiency analysis relies on consistent data to ensure like-for-like comparisons. Data errors and inconsistencies introduce uncertainties in the modelling and hence into efficiency estimates.

Regulators often use detailed regulatory accounting guidelines to collect consistent data from regulated companies. In the UK a number of regulated industries engage in extensive annual cost reporting, and similar annual reporting is also commonplace in other countries.[9] In

---

[8] It is Oxera's understanding that the ORR is working towards a different treatment of international monetary data via real exchange rates.

[9] For example, Ofwat requires all water companies in England and Wales to submit an annual return in June (known as the June Return) covering their activities in the previous financial year. See Ofwat website: http://www.ofwat.gov.uk/regulating/junereturn/prs_inf_junereturn.

industries with a single operator, such as Network Rail, regulators may need to resort to international benchmarking, where data uncertainty is often more pronounced owing to a lack of unified regulatory accounting standards. A regulatory example is provided below.

Case study: Ofwat's international comparisons

Ofwat undertakes international benchmarking to provide additional information on the performance of water companies in England and Wales. It is a high-level exercise, in contrast to the detailed modelling of relative efficiency used by Ofwat to set performance targets.[10] Instead of developing econometric models, Ofwat makes use of international data by making simple comparisons to support regulatory decisions and identify differences that allow the regulator to challenge the performance levels of the companies in England and Wales. (See section 3.2 for further discussion of Ofwat's international comparisons.)[11]

In the context of international benchmarking, particular attention needs to be paid to potential data inconsistency. The analysis for PR08 was based on data collected by the UIC (International Union of Railways) for its Lasting Infrastructure Cost Benchmarking (LICB) exercise.[12] Covering 13 rail infrastructure operators over the period from 1996 to 2006, the LICB dataset provides valuable information for the purpose of benchmarking Network Rail's performance against a set of international comparators. However, while considered robust enough for undertaking benchmarking analysis, the dataset is not without limitations. [13] Several aspects of the dataset are discussed below.

3.1.1 Consistency of LICB dataset

While the UIC enhanced data consistency by standardising the definitions of the variables in the dataset, as noted by Network Rail's consultants there are some anomalies in the data, and ITS had to correct for a number of missing data and implausible fluctuations in passenger and freight tonnage before undertaking the modelling; although other variables were assessed as being well behaved.

Data inconsistency could affect the precision of the efficiency estimates. Therefore, the analysis should be undertaken with due consideration for potential consistency issues, and the findings need to be interpreted and used appropriately. The SFA approach used by ITS for PR08 mitigates this effect, but only to some extent, as it explicitly allows for random noise in data (see section 6). Furthermore, the use of a multi-method approach can also assist in improving the confidence that can be given to the results (see section 4).

Further work is required to extend the dataset and further improve consistency. However, any adjustments to the raw data should be clearly documented. It has not been part of Oxera's remit to discuss detailed data issues with UIC, so Oxera is unaware of the specifics of the data collation exercise undertaken. Nevertheless, Ofwat's June Returns system provides some useful regulatory precedent. Following the water industry practice in England and Wales, Oxera would suggest that any adjustments to the raw data should be clearly documented and audited, shared for comments and feedback, with the final dataset made available to all parties on an annual basis prior to the modelling being undertaken. The latter is particularly important as it improves transparency, provides early indications of the likely

In 2005, Ofgem developed a Regulatory Reporting Pack (RRP) for annual submission of cost and related information, which is used to monitor actual costs incurred by the distribution network operators (DNOs) and to inform the price controls for electricity distribution. See Ofgem (2005), 'Electricity Distribution—Annual Cost Reporting', February.

[10] Ofwat (2008), 'Relative Efficiency Assessment 2007-08; Supporting Information', January.

[11] Ofwat website: http://www.ofwat.gov.uk/legacy/aptrix/ofwat/publish.nsf/Content/rpt_int_08intro.html

[12] UIC (2008), 'LICB—Lasting Infrastructure Costs Benchmarking', January.

[13] Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October; LECG (2008), 'Assessment of the ITS/ORR Econometric Models', September; Horton 4 Consulting Ltd (2008), 'International Efficiency Benchmarking in the Network Rail 2008 Periodic Review', July.

estimated inefficiency of the regulated company, and enables informed discussion and debate to occur during the consultation process. Oxera would also recommend that this be repeated annually, and not just as part of the price review process.

Furthermore, outlier tests should be undertaken at the modelling stage to identify potential data anomalies (see section 5.4).

3.1.2 Dimension of the dataset

The dataset used in PR08 contains information for 13 rail infrastructure operators over a period of 11 years. ITS made use of the entire time span for which data is available. Examples of the use of panel datasets in a regulatory setting are provided below.

> **Case study: The use of panel data in the water and electricity industries**
>
> Panel datasets have been examined in other UK regulated sectors. For example, in the UK electricity distribution network sector, Ofgem made use of data over a four-year period to assess the relative efficiency of the 14 distribution network operators (DNOs) as the basis for setting prices for the fifth distribution price control review (DPCR5).[14] In the water sector, while Ofwat continues to rely on cross-sectional data to model OPEX efficiency, in the merger inquiry of Mid Kent Water and South East Water the Competition Commission (CC) investigated the possibility of panel data modelling, and recognised the potential 'to contribute more information to Ofwat's models'.[15]

Compared with cross-section data for a single year, a dataset that spans a number of years provides more data points, which helps to generate more accurate estimates.[16] In addition, the panel structure of the dataset could generate additional insights not available in cross section data and can be used to control for additional issues. For example, fixed and random effects approaches can be used to control for country-specific factors. As such, Oxera considers that panel data should continue to be used going forward.

3.1.3 Coverage of dataset

The LICB dataset contains information on over 30 financial and structural variables. In general, four types of variables from this dataset were of interest to the cost benchmarking exercise undertaken for PR08:

– costs;
– network size;
– network usage; and
– network characteristics.

Based on this, ITS specified a Cobb–Douglas cost function,[17] using network size, usage and characteristics as determinants of total maintenance and renewal costs.

Special factors

However, a number of additional factors may influence maintenance and renewal costs and may warrant further investigation. While it is outside the scope of this review to develop this area, a list of potential special factors that may warrant further investigation includes, but is not be limited to:[18]

---

[14] Ofgem (2009), 'Electricity Distribution Price Control Review: Initial Proposals', August.

[15] Competition Commission (2007), 'South East Water Limited and Mid Kent Water Limited: A Report on the Completed Merger of South East Water Limited and Mid Kent Water Limited', May.

[16] Kumbhakar, S. and Horncastle, A. (2009), 'Improving the Econometric Precision of Regulatory Models', *Journal of Regulatory Economics*, revised and resubmitted.

[17] A Cobb–Douglas function is of the form, $Y=A.K^{\alpha}.L^{\beta}$, or in log form, $\ln(Y) = a + b.\ln(K) + c.\ln(L)$.

[18] It is Oxera's understanding that the ORR is developing its understanding of some of these factors as part of its regional analysis.

– customer objectives;
– access and incentive regimes;
– safety gauge constraints;
– the mix of traffic (freight, intercity passenger, regional passenger);[19]
– topology, relating to expenditure on tunnels, bridges, embankments, etc;
– number of stations, freight facilities and depots, number of level crossings;
– the amount of usage (by time of day, day of week); and
– the weight of trains (passenger, freight) and the length of trains (number of vehicles and wagons).

Such differences between countries and companies are known as 'special factors' in other regulated industries and often cannot be captured in the econometric modelling. This may be because the data is unavailable, the factor relates to only one (or a few) of the companies in the dataset, or the size of the dataset limits the number of factors that can be captured. An alternative approach is to use these cost items as special factors to adjust efficiency modelling results. Such an approach has been adopted by Ofwat and Ofgem as part of their most recent price reviews (PR09 and DPCR5 respectively).

Case study: Ofwat's approach to special factors

Ofwat requested water companies to submit special factor claims in May 2008, and assessed the submissions as part of the efficiency analysis for 2008–09.[20] It then communicated the reasons for its assessment to the companies, and invited re-submission of special factors to be considered for the Draft Determinations. In this regime, the onus is on the companies to submit valid special factor claims. Ofwat requests evidence to be submitted on the following:

– type of special factor: accounting issue, modelling issue or operating circumstance;
– offsetting positive special factors;
– net monetary impact of the costs, which should be greater than 1% of OPEX; and
– evidence that companies have mitigated the impact of additional costs.

After approving the special factors, Ofwat uses the results to adjust the efficiency gap estimated from its corrected ordinary least squares (COLS) regression models. In Ofwat's view, post-modelling adjustments have the advantage that uncertainties in one company's special factor would not have an impact on other companies' efficiency scores.[21] Alternatively, pre-modelling adjustment could be adopted.

The approach to special factors warrants further consideration as ORR's efficiency modelling approach develops.

Capturing the volume of work

Network size, usage and other characteristics are not perfectly related to the actual volume of work carried out by companies. The actual level of maintenance and renewal activities also varies with condition of network and investment life cycle, as well as the availability of funding. For example, an operator that sub-optimally defers track renewal may appear to be efficient. To avoid this potential bias, the dataset could be expanded to capture information on the condition of the network and actual maintenance or renewal activity undertaken by companies (see section 3.4) or this could be captured in a special factor (see above). However, the actual level of activity may also be inefficient and, as such, the efficient level of

---

[19] For example, there is a large commuter population around London; similar patterns are not observed in any other country, perhaps with the exception of France (Paris).
[20] Ofwat (2008), 'Relative Efficiency Assessment 2007-08; Supporting Information', January.
[21] Ibid.

activity would need to be determined or the condition of the network used as the cost driver.[22] In PR08, the length of track renewed was used to make a 'steady-state adjustment' for Network Rail only, as similar data was not available for other European operators.[23] As such, this approach is similar to the use of special factors (see above) and relies on the assumption that comparator operators are operating in their respective steady states. Whether and how this affects the results is an empirical issue.

Alternatively, the dataset could be expanded to include quality measures (such as network age/condition, safety standards) that act as proxies for the level of activities taken by companies. An advantage of this approach is that, unlike volume of work measures such as the length of track renewed, quality measures are more likely to be exogenously determined and closely related to the service standards received by train operators.

Input prices

A potential source of observed differences in costs (as opposed to efficiency) is differences in input prices. These are not available in the UIC dataset. The PR08 analysis sought to account for input price differentials across operators by using (PPP) exchange rates to standardise cost figures across countries. PPP adjustments are further discussed in the following section.

3.2   Summary

In the short term, due to practical considerations, analysis often needs to take datasets as given, and to focus on producing robust results within the data constraint. In the long term, it would be important to expand the existing dataset to include additional country-specific variables and to improve the consistency of the dataset.

3.3   PPPs and exchange rates

As highlighted above, the analysis for PR08 used PPPs to standardise cost figures across the 13 countries. PPP adjustment is a method used to capture potential differences in price levels between countries by converting different currencies into the same unit. PPPs have the advantage over spot exchange rates in capturing long-term differences in prices, whereas exchange rates are subject to volatility, reflecting demand and supply conditions for a given currency.

> **Case study: Ofwat's use of PPPs**
>
> PPPs have previously been used by Ofwat for its international comparison of water companies, which it has undertaken for a number of years. In this instance, Ofwat has used PPPs to adjust customer bills and unit cost for high-level comparisons between companies in Scotland, Northern Ireland, Canada, Portugal, the Netherlands, Australia and the USA. Since Ofwat can undertake comparative analysis with companies in England and Wales, the insights from its international benchmarking are not used as its primary tool for developing efficiency assumptions.

Instead of developing formal econometric models, Ofwat uses simple comparisons to provide a wider context of companies' performance, support regulatory decisions and identify differences that allow the regulator to challenge the performance levels of the companies in

---

[22] For example, Ofwat uses its Common Framework, a forward-looking risk-based assessment, to identify the efficient activity level. See Lumbers J. (Tynemarch System Engineering) and Kirby, R. (Kirby Associates Limited) (2003), 'The Common Framework for Capital Maintenance Planning in the UK Water Industry—From Concept to Current Reality'.

[23] Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October, pp. 23–24.

England and Wales. Along with bills and unit costs, the comparisons also cover the operating environment and population served by the companies in various countries.[24,]

While PPP adjustments make the cost data comparable, because the benchmarking involves comparing costs in the rail industry, it is desirable to have a PPP that is specific to this sector. Moreover, under ideal conditions, one needs to have PPPs specific to each input in order to make them comparable. The OECD publishes PPP information.[25] While this information used to be at a disaggregated level, the OECD has since stopped publishing such detailed PPP information. An example of the use of PPPs in a regulatory setting is provided below.

Case study: Ofcom's use of PPPs

Before 1993, OECD publications included PPPs for specific asset types, such as civil engineering works, transport equipment, etc. Ofcom used PPP by asset type to calculate a sector-specific PPP adjustment for BT when undertaking its comparative efficiency analysis with the US local exchange carriers (LECs).[26] However, this information is no longer available as the OECD stopped updating PPPs by asset type, pointing out that disaggregated data is less reliable because it is calculated based on a smaller sample size. In its latest benchmarking study on BT, the aggregate PPP measure (PPP for GDP) was used.[27]

Regardless of what approach is taken, the added uncertainty of the need for making such adjustments needs to be considered. For example, economy-wide PPPs can be calculated by a range of methodologies and, as illustrated by van Ark (1996), the resulting PPP measures can differ substantially.[28] Given the large uncertainty in PPP estimates, Oxera recommends that the benchmarking exercise considers the following alternative adjustments:[29]

– real exchange rates: given the caveat on market exchange rates, a long-term moving average real exchange rate could still be used as a sensitivity scenario, as it mitigates the impact of short-term volatility;

– PPPs calculated by alternative methodologies.

Finally, it is important to recognise that costs depend on input prices. At PR08, input prices were not available. Missing data on input prices creates omitted variable bias (see section 3.1). As such, input price data should be collated going forward in order to include this information in the modelling.

3.4   Technology

In efficiency measurement, the term 'technology' should be interpreted in a wider sense to encompass the engineering process that companies use, or could use, as well as the physical and policy environments in which the companies operate. Such factors may affect the appropriate cost function or cost frontier for a company.

---

[24] For detail, refer to www.ofwat.gov.uk/legacy/aptrix/ofwat/publish.nsf/Content/rpt_int_08intro.html.

[25] Details are available from OECD website: www.oecd.org/std/ppp.

[26] NERA (2005), 'The Comparative Efficiency of BT in 2003: A Report for Ofcom', March.

[27] NERA (2008), 'The Comparative Efficiency of BT Openreach: A Report for Ofcom', March.

[28] van Ark, B. (1996), 'Issues in Measurement and International Comparison of Productivity—An Overview', in 'Industry Productivity: International Comparison and Measurement Issues', OECD Proceedings, Paris.

[29] It is Oxera's understanding that the ORR is working towards a different treatment of international monetary data via real exchange rates.

In the discussion that follows, the choice of technology is assumed to be exogenous (ie, outside the control of Network Rail's management).[30] However, whatever the reason(s) for differences in technology, empirically one should at least recognise that the possibility of multiple technologies exists. The issue is whether, ex post, Network Rail's technology is different or, for that matter, whether there is more than one technology.[31] It may be desirable, but not necessary, to identify the possible source(s) of existence of multiple technologies.

To benchmark infrastructure operators' costs, a number of assumptions need to be made with regard to the technology that characterises the production process. One assumption could be that operators share a common technology frontier. However, this assumption may not hold in the present case. For example, in its international comparisons, Ofwat recognises the unique scale and structure of water companies in England and Wales, and selects as comparators 'enterprises that have a distinct corporate identity and independence'.[32] This statement highlights the importance of controlling for operating environment and technology in international benchmarking exercises. Similarly, Network Rail may operate under a different technological frontier to other European operators, for a variety of reasons, some of which are outlined as follows.

– Network Rail has broadly two[33] different types of electrification—25kv overhead lines, and powered third rails—on some parts of its network that it has to deal with.

– Health and safety requirements may differ from country to country, as may operating practices, such as the permitted length of possessions, which can affect the rate of introduction of new technology.

– In terms of policy environment, all 13 infrastructure managers in the sample differ in terms of their relationship with their respective governments, and in their degree of vertical integration. As such, their overall objectives and expenditure programmes may differ from those of Network Rail.

These differences could either reduce or increase Network Rail's efficiency estimate. As a result, the overall impact of technological difference and how these are dealt with would be an empirical issue. However, at this stage, Oxera's considers that the possibility that the infrastructure companies might operate under different technological frontiers cannot be discarded. It is important to consider and test whether technology is indeed the same across all companies and time periods.

The following discusses how differences in technology could be addressed using top-down and/or bottom-up approaches.

Top-down approach
As a result of differences in technology, the estimated parameters in the cost function may differ across companies and over time. This possibility should be tested for before modelling inefficiency. This is critical because efficiency scores are estimated based on the cost function, and efficiency estimates alone would not distinguish between differences in technology and differences in technical inefficiency. If all the companies do not have the exact same technology, the inefficiency estimates will be statistically biased. Using a top-down approach, the parameters in the cost function can be allowed to vary across companies and over time. Standard statistical tests can be employed to test which parameters remain constant across the entire sample. Stable parameters indicate the

---

[30] This issue needs further consideration and examination.

[31] See, for example, Kumbhakar, S .and Orea, L. (2004), 'Efficiency Measurement Using a Latent Class Stochastic Frontier Model', Empirical Economics, 2004, 29, pp. 169-183.

[32] For detail, refer to www.ofwat.gov.uk/legacy/aptrix/ofwat/publish.nsf/Content/rpt_int_08intro.html.

[33] There are further alternatives on the GB network, including dual voltage (AC/DC) routes, but these are generally small relative to the overall size of the network.

presence of a common cost frontier for all observations. If differences are found, separate frontiers could be specified, for example by including dummy variables for countries in the cost function.

If technology is country-specific, the panel structure of the dataset can be exploited to control for differences in technology. This is different from country-specific effects in inefficiency which can be modelled in a variety of ways (see section 5.3.3).

Bottom-up approach

While top-down models can be used to estimate and test different cost frontiers, this is not the only approach that warrants investigation. Evidence from bottom-up/engineering analysis can be used to examine and quantify country- or company-specific effects.[34] Such analysis may identify reasons for differences in efficiency not accounted for in the econometric model, which may or may not be exogenous. Clearly, as with Ofwat's approach to special factors (see section 3.1.3), there would be a need to demonstrate that these effects are exogenous if they are to be regarded as special factors. Such findings could then be used to adjust Network Rail's (and other companies') cost base (before or after the modelling). This approach would be similar to the adjustments undertaken by other UK regulators.[35] As the adjustment could be undertaken by the ORR outside the modelling process, this approach does not require significant changes to the top-down modelling approach.

3.5   Hatfield and steady state adjustment

The Hatfield derailment in 2000 represented a turning point in the level of activity undertaken by Railtrack/Network Rail and, as a result, Network Rail may not be operating in a long-term steady state in terms of its volume of activity.[36] Before 2000, a backlog of maintenance and renewal is considered to have been built up by Railtrack. Immediately after Hatfield, the level of activity increased significantly, and priorities shifted to improving safety.[37] Similar events could have happened in other countries as well, making other companies operate out of their respective steady states.

The PR08 modelling sought to account for the impact of Hatfield through the following:

– modelling Railtrack and Network Rail as two separate companies;
– making a steady-state adjustment to Network Rail's renewal costs; and
– allowing for a non-linear time profile of inefficiency.

Flexibility in the modelling framework is required in order to accommodate the potential impact of specific events such as Hatfield, while avoiding pre-judgement on the impact of such events. In addition to the approaches adopted in PR08, Oxera recommends that, before modelling inefficiency, the cost function should be tested for possible structural breaks following particular events, such as Hatfield.

[34] Bottom-up best-practice benchmarking can also be used (and, indeed, was used at PR08) to estimate Network Rail's efficiency independently of the top-down approach. Here, Oxera is referring to bottom-up analysis to be used to quantify the reasons for any differences in efficiency esitmated from the top-down modelling.

[35] For example, in PR09, Ofwat allowed new claims for special factors. See Ofwat (2008), 'Setting Price Limits for 2010-15: Framework and Approach', January, p. 41. Ofgem allows DNOs to report atypical costs as part of the annual RRP submissions. Most atypical costs, apart from those related to severe weather conditions, are excluded from efficiency modelling. See Ofgem (2009), 'Electricity Distribution Price Control Review Initial Proposals—Allowed Revenue – Cost Assessment', August.

[36] See, for example, LECG (2008), 'The Appropriate Assumption for Network Rail; Assessing Network Rail's Scope for Efficiency Gains: A Report for Network Rail', April, p. 45.

[37] See, for example, Oxera/LEK (2005), 'Assessing Network Rail's Scope for Efficiency Gains over CP4 and Beyond: A Preliminary Study', December.

3.5.1 Indentifying and dealing with structural breaks

Testing cost functions for structural breaks could be done in a number of ways, including adding dummy variables for such events.[38] Both UIC and the ORR have raised concerns about such an approach, on the grounds that it leads to a different treatment of railways in Britain, and events could have happened in other countries in the sample.[39] However, events that may have affected companies' efficiency could be reviewed at the European level. Ideally, the modelling needs to make adjustments for all comparators where appropriate. However, this may not be feasible (for example, in PR08, a steady state adjustment could not be done for countries other than the UK due to data limitations) or it could lead to over-parameterisation of models (ie, requiring too many parameters to be estimated, thus reducing the discriminatory power of the model). As well as testing for structural breaks, alternative modelling approaches, such as data envelopment analysis (DEA), could be used to demonstrate the impact of shock events without including an excessive number of variables. This is further discussed in section 4.2.

3.5.2 Dealing with 'lumpy' expenditure

A more general point, however, is the 'lumpy' nature of renewal costs and the critical issue of determining an efficient activity level that has often beset the modelling of maintenance or renewals expenditure. Renewals and maintenance costs are likely to be driven by required levels of activity. A possible approach could be to model average costs over a relatively long period of time, in order to reduce the influence of any particular year. Such approach has regulatory precedence.

> **Case study: Ofwat's approach to econometric modelling of capital maintenance**
>
> Ofwat and the water industry, in developing their capital maintenance expenditure models, considered this issue in some depth.[40] The models used explanatory variables from a base year, to model expenditure averaged over six or more years, in an attempt to smooth out the annual variation in capital maintenance spend. Early models included the condition of the asset base as a key cost driver. However, such a key cost driver was subsequently dropped from the modelling due to data consistency issues. The resultant modelling of capital maintenance expenditure has been criticised ever since, with companies arguing that the models simply picked up differences in levels of activity that could be explained by need. These econometric models have since been dropped by Ofwat.[41]

While the LICB dataset does not span an entire investment cycle, which might be decades given the long asset lives in the rail industry, this is the case in many infrastructure industries, including water. Exploratory statistical analysis (perhaps of other datasets) could be undertaken to assess the extent to which measures exhibit cyclical behaviour and whether cycles can be regarded as comparable. If certain companies appear to operate in part of the cycle, in particular where expenditure is suspected to be lower due to cyclical events, consideration should be given as to whether such companies may push out the frontier and making efficiency scores infeasible for those below the frontier.

As such, a greater understanding of the impact of rail network condition and degradation, output requirements, and political factors is important in order to measure inefficiency accurately.

---

[38] The core objective of the analysis would be to estimate Network Rail's efficiency at the end point of the period, so the critical issue is to ensure that this is estimated as robustly as possible.

[39] Based on discussion with ITS and the ORR.

[40] For further detail, refer to Ofwat (2007), 'Relative Efficiency Assessments 2006-07: Supporting Information'.

[41] Ofwat discontinued the use of these models when it introduced menu regulation for CAPEX. See, for example, Ofwat (2007), 'New Approaches to Expenditure and Incentives: A Discussion Paper', May, p. 13.

4　Approaches to modelling

> **Recommendation 4.1**: Several modelling approaches should be used to cross-check the robustness of the results.
>
> **Recommendation 4.2**: The use of semi-parametric approaches should be investigated and results contrasted with other approaches.
>
> **Recommendation 4.3**: An approach should be developed to synthesise the results from these alternative approaches.

4.1.1　Using alternative approaches

The academic literature has developed a wide range of modelling approaches to assessing efficiency, all of which have advantages and disadvantages. In developing practical approaches to measuring efficiency for use in a regulatory setting, these advantages and disadvantages need to be carefully evaluated. In general, using a number of different benchmarking techniques and being aware of the approaches' strengths and weaknesses can help improve the robustness of the efficiency exercise.

As noted in PR08, 'cross checking the results of benchmarking analysis of this sort in economic regulation by using alternative techniques is now considered regulatory best practice'.[42]

Before settling with a single preferred approach, a thorough comparison of various techniques should be undertaken, considering both their relative strengths and weaknesses. As a minimum, alternative approaches should be used to provide a cross-check for consistency. Preferably, these alternatives should be used jointly, and where results differ, the focus should be on understanding the underlying reason for the observed differences.

In practice, modelling techniques such as COLS, DEA, and SFA, as well as bottom-up benchmarking, have all been used in various regulatory contexts, either separately or jointly.

Up until now in this report, all the discussion has focused on the SFA modelling undertaken for PR08, although many of the issues discussed relate to econometric modelling more generally (including COLS). In section 4.2 the difference between SFA and COLS is discussed, while a more detailed discussion of DEA—an approach that was dropped in the final analysis in PR08—is provided in section 4.3. (Oxera does not examine bottom-up benchmarking approaches in this section as this is outside the scope of this review—it is only touched upon in section 3.1.3 on special factors. However, such an approach can also provide additional valuable information on Network Rail's efficiency).

4.2　Corrected ordinary least squares and stochastic frontier analysis

Both COLS and SFA (see Figure 4.1 below) are examples of 'parametric' techniques. In PR08 both techniques were explored, with COLS used as a cross-check on the (preferred) SFA results. COLS results were adjusted by placing the benchmark at the upper quartile (instead of the frontier company) to make them more comparable to SFA estimates. This modification is designed to account for the inherent uncertainties in data. In a regulatory

---

[42] Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October, p. 13.
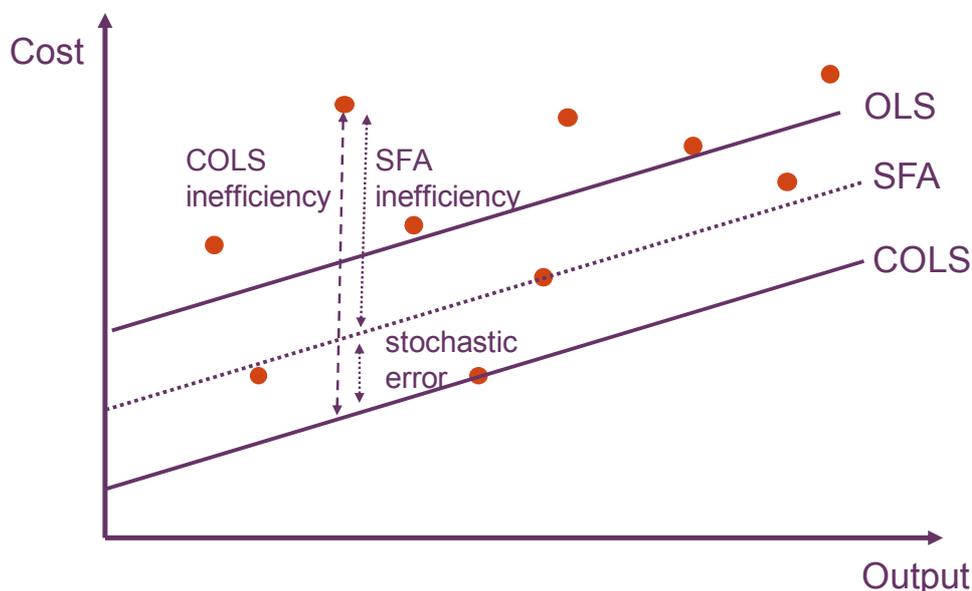
context, Ofgem makes similar adjustments to the efficiency scores estimated by COLS (see section 6.2 for further details).

COLS is a relatively simple methodology, based on OLS regression, which estimates an average cost function by fitting a 'line of best fit' through the data. It is 'corrected' in the sense that the cost function is shifted to the most efficient company, to form the cost benchmark for the industry (shown by the COLS frontier in Figure 4.1). The resultant gap between a company's actual cost and this predicted cost is interpreted as inefficiency.

A significant drawback of COLS analysis is that it assumes that the estimated residual for any company is attributable to relative (in)efficiency. As the presence of data issues could be a major concern in an international benchmarking exercise, the residuals from COLS will include unexplained cost differences resulting from data errors (eg, due to cost allocation inconsistency), as well as the impact of omitted variables and other factors (see section 3).

SFA overcomes this issue by decomposing the residual term into random noise and inefficiency. The difference between the COLS and SFA frontiers is illustrated in Figure 4.1.[43]

Figure 4.1   OLS, COLS and SFA

Note: The SFA frontier drawn in the figure is purely illustrative.
Source: Oxera

Efficiency scores from SFA can be regarded as more objective than those from COLS as they require no subjective judgement regarding how much noise is present in the estimates. Therefore, SFA is generally preferable to COLS, data permitting.

4.3   Data envelopment analysis

4.3.1   Review of PR08
DEA is a linear programming-based method for measuring efficiency. In the current context, DEA measures the efficiency of a company by first estimating the minimum cost level necessary to secure its output levels when it is compared with peer companies. The ratio of

---

[43] Figure 4.1 is intended to illustrate the difference in frontiers estimated by OLS, SFA and COLS. For the purpose of presentational clarity, a dotted line represents the SFA frontier. The actual SFA frontier varies by company and year, depending on the estimated value of the stochastic component.

minimum estimated cost level to the observed cost level gives a measure of a company's efficiency and of its scope for efficiency savings.

In PR08, DEA was not applied to the final 1996–2006 dataset, for the following stated reasons:[44]

–   preliminary DEA analysis, based on the dataset for the years 1996–2005, produced similar results to the COLS econometric results;

–   DEA has difficulties in dealing with scale and density effects without resorting in any case to a two-stage approach that involves econometric estimation in the second stage.

While it is not the focus of this report to review the approach used in PR08, it is worth considering whether DEA should be examined going forward. Oxera considers that DEA can provide an alternative to the econometric approach and serve two aims. First, it can address the issues highlighted in PR08 and is a valid and widely used approach to efficiency measurement. Second, it can provide a rich set of additional insights into performance in terms of technical efficiency, cost efficiency, productivity change, and comparisons of productivity across companies and within a company over time. Comparison of each company with its peers, as identified by DEA, is of particular importance in the present setting where, as discussed elsewhere in this report, uncertainty exists about certain assumptions (eg, functional form, input prices across countries). Furthermore, DEA does not need to make an assumption, for each company, about whether efficiency is constant, always rising or falling over time, and will reveal the extent to which that assumption accords with the data. DEA can also offer insights into how the mix of maintenance and renewal affects a company's efficiency. This could be important since decisions between renewal and maintenance of track may be beyond a company's control and determined by age of track.

The following discusses, at a broad level, how DEA modelling can be undertaken, together with its potential in addressing the issues noted in PR08 and providing additional insights into Network Rail's efficiency estimates.

4.3.2   Developing DEA model

Input–output variables
The first stage in the efficiency assessment is the identification of input–output variables. Since DEA is not an econometric approach, it is not possible to test which variables to include in the model. Instead, the econometric model can be used to inform the choice of variables. Thus, having first identified a large set of potential input and output variables, based on industry knowledge and economic intuition, OLS regression would then be used to determine the variables that have a significant impact on costs and these variables would be used in the DEA model. That is, DEA and econometric modelling are generally used in parallel.

In PR08 a number of density and ratio measures were used as proxies for network characteristics, such the proportion of single track and electrification. Ratios of output variables, which can be readily included in econometric modelling, cannot be directly used in DEA. However, these variables can be incorporated in DEA indirectly in a gross, rather than a ratio manner, and similarly allowed to influence the measured efficiency. For example, the analysis can split route-km into electrified and non-electrified and use the two variables as outputs. Clearly, route-km, passenger train-km and freight train-km, for example, are cost drivers and thus the efficiency measure should reflect the cost level at which a company can handle a given volume of these. However, the issue of a density or ratio measure is different in that it reflects the need to ensure that the specific proportion of total km that is, say,

---

[44] Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October, p. 10.

passenger-km (or, by implication, freight-km) influences costs given a certain volume of activity. This proportionality of what are otherwise expenditure drivers in their own right can be readily captured in the DEA model through restrictions. That is, DEA models can also incorporate the assumption that one type of output (eg, single track) is relatively cheaper than the other (eg, multiple tracks).

Returns to scale

The next step is to decide on the assumption of returns to scale (variable returns to scale, VRS, or constant returns to scale, CRS). This again could be based on the econometric model. The Cobb–Douglas cost function used in PR08 assumes that returns to scale remain the same for all companies across all years. In econometric modelling, the returns to scale assumption can be tested. While, in DEA, the shape of the cost function cannot be tested, a range of assumptions can be imposed based on evidence from the econometric modelling or on industry expertise. In most cases, CRS is a strong assumption. It does not disaggregate pure technical inefficiency and scale inefficiency, which is usually outside managerial control in the case of a regulated industry. If there is no evidence to suggest CRS, the DEA model should be run using a VRS specification, which allows for (dis)economies of scale. The assumption on returns to scale could have a considerable impact on the estimated efficiency of Network Rail given that it is a large operator, measured by the size of its network.

Input prices

The PR08 model using total cost arrives at an overall cost efficiency in which the following factors are conflated, in the sense that benchmark companies would be those that:

– minimise volumes of track maintained and renewed, given the mix (ratio) of these;

– have the most advantageous mix of maintained/renewed volumes given the 'prices' they face for each; and

– happen to face the lowest prices (eg, because labour and/or materials are cheaper).

The issue of potential inaccuracies in the PPP conversions (section 3.2) does, of course, have an impact on the second and third of the foregoing factors.

This conflation may lead to least-cost benchmarks that might be misleading about the potential for efficiency savings at a company because the second and/or the third of the above conflated factors may not be exploitable by the company. For example, the company may not be able to alter the mix of maintained/renewed track because of the age of its track; alternatively, the company may face higher material/labour costs for reasons beyond its control or the PPP factors are not sufficiently accurate. As such, an alternative approach that decomposes these factors warrants investigation.

A first step in decomposing the above factors in the absence of input prices would be to use a two-input DEA model in which the two inputs would be either renewal and maintenance costs or physical track maintained and renewed.[45] The former does not get over the problem of differing labour and material prices, but it goes some way to dealing with the issue of the company not controlling the mix of its track that needs renewal and that which needs maintenance.[46]

In suggesting this decomposition of cost inefficiencies, no ex ante view needs to be taken as to whether trade-offs in the foregoing sense exist or not. Oxera considers that it would be beneficial to decompose, as far as possible, the overall cost inefficiency of each company into components attributable to technical, allocative and price differentials. This would enable

---

[45] The decompositions cannot be carried out by separate modelling of renewal and maintenance costs.

[46] An SFA input distance function using the same two inputs as in DEA would be an alternative approach which could also be taken.

an informed view to be taken, through knowledge of the operating context or any special factors affecting each company, as to whether to discard in whole, or in part, some of these components. Potential differences between renewals and maintenance will affect the decompositions, but, on the other hand, total cost modelling discards potential inaccuracies in the PPP conversions, which, in the decomposition approach, can be identified more easily.

Uncertainties

A shortcoming of non-parametric approaches such as DEA (but also COLS in the set of parametric approaches) is that it constructs a deterministic frontier, which is more sensitive to noise and outliers in data than, for example, an SFA frontier. The basic DEA model, however, can be expanded in a number of ways to accommodate the impact of noise. These approaches include super-efficiency and bootstrapping (see section 6).[47]

4.3.3 Two-stage DEA

Second-stage and multi-stage DEA are among a variety of approaches to capturing the impact of environmental factors in DEA.[48] In brief, when environmental factors affect the operations of a unit in terms of how inputs can be converted to outputs, the assessment by DEA should reflect the impact of the environmental factors. For example, if data permits, potential factors to be included in the second-stage regression could be environmental and policy variables, covering:

– ownership structure;
– public policy;
– scope and density;
– topology; and
– differences in production processes.

There are alternative approaches to capturing this impact. One is the 'all-in-one' approach where the environmental conditions are part of the initial input–output set in the DEA assessment, obviating the need for a second- (or indeed multi-) stage regression. However, environmental factors may not be under the control of a company. Second-stage regression is an alternative to the all-in-one approach where 'internal' factors are used as inputs and outputs in a traditional DEA model and the resulting DEA scores are regressed on the environmental factors in order to adjust the original DEA efficiencies to reflect the impact of the environmental factors. Because the efficiency scores are a limited dependent variable, between zero and one, the Tobit model is considered an appropriate methodology to use in the second stage.[49]

Multi-stage DEA is an extension of two-stage DEA in order to capture, it is argued, in a more effective way the impact of environmental factors.[50] Implicit in adopting this approach is the assumption, generally difficult to justify, that the internal inputs and outputs are independent of the environmental factors. The all-in-one approach could be used as a way around the problem of lack of independence.

---

[47] *Super-efficiency*: Andersen, P. and Petersen, N. (1993), 'A Procedure for Ranking Efficient Units in Data Envelopment Analysis', *Management Science*, **39**:10, 1261–64. *Bootstrapping*: Simar, L. and Wilson, P. (2008), 'Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives', in H. Fried, K. Lovell and S. Schmidt (eds) (2008), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

[48] A brief exposition of the approaches can be found in Thanassoulis, E., Portela Maria, C.S., and Despić, O. (2008), 'DEA – The Mathematical Programming Approach to Efficiency Analysis', in H. Fried, K. Lovell and S. Schmidt (eds) (2008), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

[49] Tobin, J. (1958), 'Estimation of Relationships for Limited Dependent Variables', *Econometrica*, **26**, 24–36.
An example of Tobit regression in the rail sector is Oum, T.H. and Yue, C. (1994), 'Economic efficiency of railways and implications for public policy: a comparative of the OECD countries' Railways', *Journal of Transport Economics and Policy*, **28**:2, 121–38.

[50] For example, the four stages referenced to in Thanassoulis, E., Portela Maria, C.S., and Despić, O. (2008), 'DEA – The Mathematical Programming Approach to Efficiency Analysis', in H. Fried, K. Lovell and S. Schmidt (eds) (2008), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

However, for the model developed in PR08, it is debatable whether the need for second- or multi-stage regression will arise. As noted earlier, the presence of ratios does not necessarily, in itself, imply the need for two- or multi-stage DEA. This debate can only be settled at the point of actual implementation of the assessment of rail companies.

4.3.4 Comparison of results

The comparison of results estimated by SFA and DEA could provide additional insight into the source of inefficiency. How the results could be used together is discussed in section 4.5.

4.4 Non- and semi-parametric approaches

Within an econometric regression modelling framework, in order to obtain parameters that can be used for prediction purposes (eg, efficiency) and to test economic hypotheses, a functional form needs to be assumed for the relationship between parameters and the explanatory factors that describe the modelled variable. Approaches that rely on a limited number of parameters to describe these relationships are referred to as parametric approaches. Often-used examples in the econometric estimation of cost functions are Cobb–Douglas or translog functional forms.

At the last review there was some debate over whether a translog or Cobb–Douglas functional form was the most appropriate, given the dataset. The Cobb–Douglas was, at that time, deemed the most appropriate form for that particular dataset.

The validity of the functional form assumption can be verified using statistical tests. However, there may be instances where the choice of the parametric functional form is limited and therefore the appropriate functional form cannot be robustly determined. One approach would be to use a flexible functional form, such as a translog specification, and test for something that is less flexible than the translog, such as semi-translog that is more general than the Cobb–Douglas. The semi-translog may not include the square terms but it may include the cross-product terms to give some flexibility in estimated elasticities. This is because the economic implications of the Cobb–Douglas function are quite restrictive on economic theoretical grounds. Furthermore, since an incorrect inflexible functional form may introduce bias in the efficiency estimates (see further discussion below) alternative approaches need to be investigated. These may also be used to cross-check whether efficiency estimates are unduly influenced by potentially over-restrictive functional form assumptions.
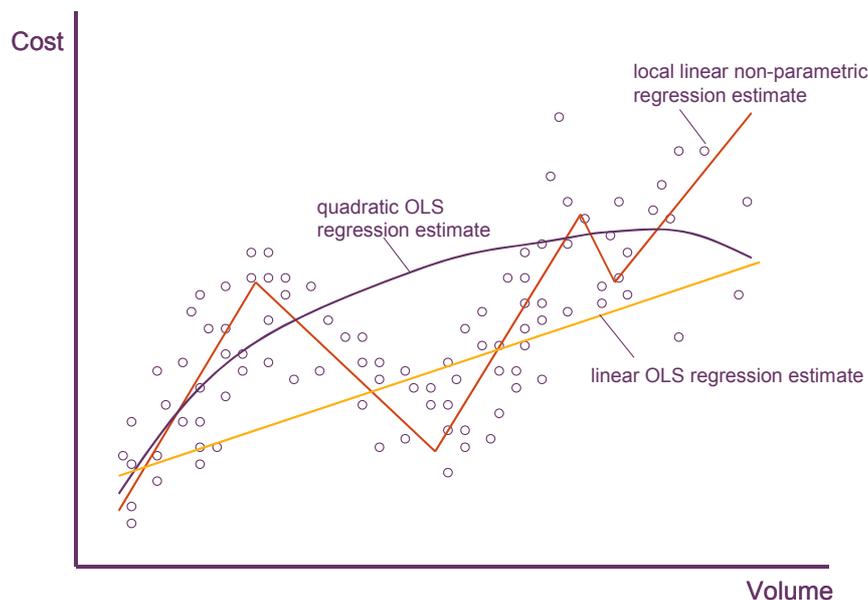
Approaches that place less or no reliance on the parameters of a cost function are semi-parametric and non-parametric approaches, with the latter dispensing with parameter and functional form assumptions altogether. DEA (discussed above) is a non-parametric approach, in the sense that it does not use any parametric functional form. However, DEA is not a non-parametric econometric approach, as discussed below.

To illustrate the non-parametric approach, consider Figure 4.2 below which shows a scatter plot of costs against route-km. The true relationship underlying this dataset is non-linear. There are several approaches that can be taken to obtain information about this cost–volume relationship.

In this example, fitting a simple linear regression would capture the true relationship ineffectively and systematically and significantly over- or underestimate the cost for a given route-km. Similarly, a quadratic functional form also does not appropriately capture the true relationship. As an alternative, a local linear regression approach could be adopted to uncover the true relationship. This approach amounts to running a linear regression at each point by taking into account the nearby points. By doing so the non-parametric regression

most closely captures the true relationship. Therefore, if the functional form assumption is in doubt, a non-parametric approach is likely to provide a better fit to the data.[51]

Figure 4.2   Comparison of linear and quadratic OLS estimates of regression with non-parametric regression local linear estimates

Source: Oxera.

Not having to specify a functional form has a further advantage from an economics perspective since, in most cases, economic theory provides some idea about the possible dependent and independent variables, but it rarely helps in choosing the functional form. The non-parametric approach is thus closer to economic theory.

The semi-parametric approach to efficiency modelling comprises two steps. The first step involves using non-parametric regression to model the relationship between costs and explanatory factors. The second step involves calculating inefficiency using SFA. (The term 'semi-parametric approach' is derived from using a non-parametric approach in the first step, and a parametric approach in the second step.)

Clearly, each modelling approach has its advantages and disadvantages. So far, only the benefits of non- and semi-parametric modelling have been discussed (namely, the relaxation of the rigidities of the traditional approach, as well as providing a way of testing the robustness of the traditional approach). The non- and semi-parametric approaches suffer from some drawbacks mainly in terms of the size of the dataset. However, since these models are econometric (which always include noise terms), it is possible to undertake statistical testing (although parameters are not directly estimated). For example, one can test these models against some standard parametric models (such as the Cobb–Douglas model). As such, a combined approach is proposed in this report and how to draw together the results from several approaches is discussed in section 4.5 below.

4.5   Cross-checking or synthesising the results

Regression-based methodologies such as COLS and SFA make use of industry-wide information to determine the position of the cost frontier consistent with a pre-specified functional form. On the other hand, the segment of the DEA frontier and semi-parametric frontier used as the referent to assess the efficiency of each unit is based on the unit's

---

[51] Another alternative, among others, would be to fit a spline function.

characteristics to a much greater extent. Thus, efficiency scores estimated by DEA/semi-parametric regression and COLS/SFA may reinforce or contradict each other. Having results from both parametric and non-parametric models, where the results agree, enhances the confidence in the findings. Where the results differ substantially, reasons for the differences should be investigated.[52]

The implication is that it is not just a simple case of combining DEA, semi-parametric regression, SFA and COLS results in a mechanistic way to arrive at final estimates of efficiency. Rather, it is using the comparison to build, possibly iteratively, better parametric and non-parametric models which ultimately yield mutually compatible results, with any differences clearly understood and judgement made as to which results to adopt. That is, where there are significant differences in the results, these should be investigated and understood. Based on an understanding of the different approaches, the models should then be further refined in order to develop more consistent results as far as possible.

---

[52] See, for example, Thanassoulis, E. (2002), 'Comparative Performance Measurement in Regulation: The Case of English and Welsh Sewerage Services', *Journal of the Operational Research Society*, **53**:3, 292–302.

5 Specific modelling issues in PR08 analysis

> **Recommendation 5.1**: Substitution between cost categories should be captured where possible.
>
> **Recommendation 5.2**: The appropriate functional form of the cost model requires detailed examination and testing.
>
> **Recommendation 5.3**: For SFA models, alternative assumptions for the distribution of noise and inefficiency should be examined.
>
> **Recommendation 5.4**: Explaining the efficiency term through explanatory factors assists in understanding inefficiency and improves the robustness of the modelling, and should be examined in further detail.
>
> **Recommendation 5.5**: Modelling the time profile of inefficiency assists in understanding inefficiency and improves the robustness of the modelling, and should be examined in further detail.
>
> **Recommendation 5.6**: A range of statistical and economic tests needs to be carried out to develop a robust cost model.

5.1 Explanatory factors

The importance and benefits of settling on a preferred modelling approach(es) only after extensive consideration of alternative approaches (ie, parametric, non-parametric and semi-parametric techniques) have been discussed above. This section *assumes* that SFA is the most appropriate methodology to assess Network Rail's comparative efficiency, and provides comments on taking the PR08 analysis forward, based on the range of models used in PR08.

While regulators often undertake a consultation on the relevant input and output variables (eg, cost definitions and the information used in modelling costs, such as outputs, input prices, and quality measures), in practice benchmarking exercises are often constrained by data availability. This was the case with the econometric study for PR08. The lack of potentially relevant, yet unavailable data leading to potential omitted variables bias was noted in PR08, and discussed in section 2. This section focuses on how the available information was used in the modelling process.

5.1.1 Modelled cost
It is possible that some degree of substitution exists between maintenance and renewals expenditure. In addition to economic trade-offs between cost categories, in the current context the added complication arises that maintenance and renewals are similar in nature and the classification follows an accounting rather than an economic convention, leading to possible inconsistencies in cost allocation. To address this, total cost modelling (ie, maintenance costs plus renewals costs) was adopted as the preferred approach in PR08. While this is a possible approach, using total cost as the modelled cost variable may not fully address potential input or output substitution.

The aggregation of maintenance and renewal costs results in the implicit modelling assumption that £1 additional spending on maintenance saves £1 in renewal. In practice, the

substitution between the cost categories (or sub-categories) does not follow on a one-to-one basis.

The substitutability between inputs, while recognised in a regulatory context, is often resolved using a pragmatic approach. Where cost allocation issues are sufficiently resolved, regulators often model separate functional or activities expenditure and then aggregate the results for benchmark selection. For example, Ofwat uses four functional models to determine the overall water service operating efficiency of water companies.[53] In the rail industry, however, similar conditions may not hold. Operationally, maintenance and renewal costs may not be separable, and companies in different countries may not account for these cost categories in a consistent manner. Joint modelling may therefore be more appropriate in this context.[54] However, this may not translate into total cost modelling, where total cost (maintenance plus renewals) is the dependent variable.

One possible approach is to specify a cost function which aggregates maintenance and renewals via a parametric aggregator function. Such a model can be represented as:

$$f \text{ (maintenance, renewals)} = \beta \cdot \text{cost drivers} + \text{noise} + \text{inefficiency}$$

where f(.) is an aggregator function for total cost. If f (maintenance, renewals) = maintenance + renewals = total cost, the trade-off between maintenance and renewals costs is one to one. If, on the other hand, f (maintenance, renewals) = a × maintenance + (1 − a) × renewals— where 0<a<1 is a parameter to be estimated—the trade-off is not one to one.

If the model is in logs, one can specify:

$$\log(\text{total cost}) = a \times \log(\text{maintenance}) + (1 - a) \times \log(\text{renewal})$$

The 'a' parameter determines the weight that is the trade-off. One can test whether a = 0.5, in which case the trade-off is one to one. Moreover, the model is linear and does not involve any complex estimation issues.

The other possibility is to use a Constant elasticity of substitution (CES) aggregator function. However, the CES function is non-linear and therefore can have estimation issues. If one wants more flexibility, log(total cost) can be specified as a translog function involving log(maintenance) and log(renewals). It is worth noting that the parameters of the aggregator function (whether linear or not) are identified and estimated jointly with the other parameters of the model.

This approach accounts for possible cost misallocation between the categories, without making the assumption of one-for-one substitution. Total cost modelling can be seen as a special case of the above, with the sum of maintenance and renewal as the variable on the left-hand side.

5.2   Functional form selection

In selecting a functional form that is theoretically appropriate and best describes the data, in PR08 the following were considered:

–   the linear function: all variables are used in the cost function without transformation;

–   the log-linear or Cobb–Douglas function: all dependent and independent variables enter the cost function in natural logarithms;

---

[53] Ofwat (2009), 'Relative Efficiency Assessments 2007-08: Supporting Information', January.

[54] If trade-offs between cost categories are deemed important, combined modelling is theoretically justified. Separate modelling may result in a misleading impression of a company's efficiency. For more details, see Oxera (2007), 'Assessing Approaches to Expenditure and Incentives', October, section 5.2, http://www.oxera.com/main.aspx?id=6549.

– translog function: similar to the log-linear form, but with square and cross terms for the main traffic and scale variables.
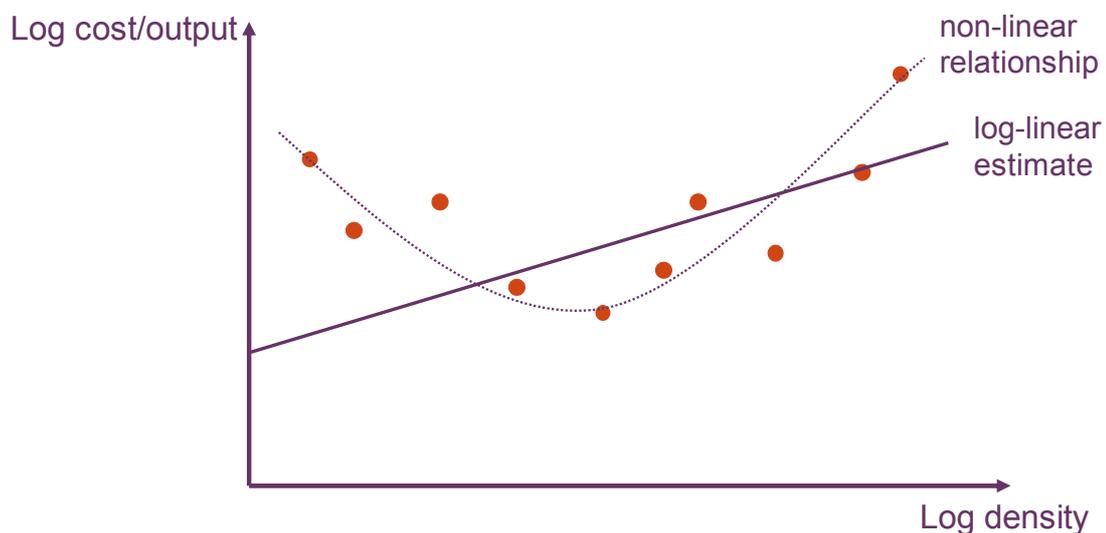
Based on statistical testing, the Cobb–Douglas functional form was adopted as the preferred specification (with a linear form being rejected on the basis of a Box–Cox test). Moreover, linear restriction tests indicated that the Cobb–Douglas function was preferred over the translog specification.[55]

Going forward, while statistical testing was undertaken on the PR08 dataset, the Cobb–Douglas functional form has several well-known limitations, and a translog functional form, which offers greater flexibility, may warrant further investigation before discarding it in favour of the Cobb–Douglas specification.

In PR08, it was noted that 'a well acknowledged practical advantage of the Cobb–Douglas function, over the translog form, is its parsimony' (ie, using fewer explanatory variables).[56] The ease of interpretation was noted as another appealing feature of a Cobb–Douglas functional form specification, with the coefficients representing the cost elasticities with respect to outputs. The potential downside of this relative simplicity, however, is that the Cobb–Douglas functional form forces the elasticities to be the same for all countries and all years. This may not be an adequate description of the variation in costs in an international comparison, given the heterogeneity across countries and time (for example, around specific events, such as Hatfield).

The issues associated with this can be illustrated by assuming that a non-linear relationship exists between costs and certain cost drivers. For example, unit rail maintenance costs could be higher both in densely populated countries [57] and in countries with a greater share of rural areas.[58] As illustrated in Figure 5.1, a log-linear function ignores the potential U-shaped relationship between cost and the cost driver, population density, and may bias the efficiency results (upwards or downwards) for some operators.

Figure 5.1 Illustrative comparison of non-linear relationship and linear cost function between costs and density

---

[55] Smith, A. (2008), 'International Benchmarking of Network Rail's Maintenance and Renewal Costs: An Econometric Study Based on the LICB dataset (1996-2006): Report for the Office of Rail Regulator', October, p. 28.

[56] Ibid., October, p. 16.

[57] For example, due to the requirement to carry out maintenance work during non-standard hours and the associated higher labour costs.

[58] For example, due to the longer access time to tracks.

Source: Oxera.

A translog function could be used to accommodate the above-described type of non-linearity.[59] Another implication is that it also allows the elasticity measures, which now depend on the actual values of the data, to vary by company and time.

The greater flexibility of the translog function comes at the cost of an increased number of parameters that need to be estimated, and hence a reduced parsimony. This could be a potential constraint on smaller datasets.

Thus, it is necessary to reconcile flexibility and parsimony. As such, a detailed examination of the appropriate functional form is required prior to estimating inefficiency.

5.3 Factors affecting inefficiency and noise

SFA requires assumptions to be made about the distribution of the inefficiency and noise terms. These distributional assumptions are used to estimate first the parameters of the frontier function and then the inefficiency.

5.3.1 Distribution of the inefficiency term
There are several distributional assumptions that can be used for the inefficiency term. It is a routine procedure to check the robustness of the results due to alternative distributional assumptions. The distribution of the inefficiency term—ie, whether one or more of the most common functional form assumptions is applicable (eg, half-normal, exponential, truncated normal, gamma)—and the impact on efficiency needs to be considered.

5.3.2 Distribution of the noise term
The noise term is, by definition, normally distributed with a zero mean. However, its variance, instead of being constant, could be affected by a number of factors. In PR08 no determinant of the noise term was specified, and thereby it was assumed that its distribution was homoscedastic (ie, its variance was assumed to be the same for all companies and for all years). In practice, this assumption may not hold. To improve the robustness of efficiency estimates, the possibility of heteroscedasticity in the noise term should be further investigated.[60] For example, an improvement in accounting standards could lead to more reliable data over time. In this illustrative example, the variance of the noise term is expected to decrease over time. SFA models allow variables that determine the distribution of the noise term to be specified.

5.3.3 Explanatory factors in the inefficiency and noise terms
In addition, SFA allows the specification of factors that affect the inefficiency term (but not modelled costs).[61] This offers the flexibility for SFA models to control, for example, for the effects of country-specific factors, as well as specific events that have an impact on inefficiency.

Variables that could be considered as determinants of inefficiency include the following.

–   **Time**: the technical efficiency of companies is generally expected to improve over time (although it may not, of course). (This is distinct from technological improvements, which should be reflected in the cost function rather than in the inefficiency term.) The time variable can be defined chronologically or according to the number of years/price review periods since privatisation or regulation. The latter is consistent with the theory that inefficiency is gradually driven out by incentive regulation. Clearly, this can be tested empirically. (See also section 5.3.4.)

---

[59] Technically, the greater flexibility is due to the inclusion of square and cross terms of explanatory variables.

[60] Kumbhakar, S. and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, chapter 3.4.

[61] Ibid.

- **Network operators**: this is another channel to control for network operator-specific factors. For example, company dummies in the inefficiency term capture the effect that certain companies have riskier operating environments than others. Alternatively, additional variables, other than dummies, can be included in the inefficiency term.

- **Specific events**: certain specific events may not only have an impact on the cost structure of a company/industry, but may also change the profile of inefficiency. Hatfield may be an example of such an event. (See also section 5.3.4.)

All of these are potential options for modelling efficiency. The actual approach will need to be developed and refined at the estimation stage.[62]

5.3.4 Time profile of inefficiency

The PR08 specification allowed flexibility in the temporal profile of the inefficiency term. In particular, the efficiency of all companies (except for Network Rail) was allowed to improve (or deteriorate) steadily over time, albeit at different rates. For Network Rail, efficiency was allowed both to increase and decrease over time (ie, a turning point was allowed for). Thus, while other factors explaining the time profile of inefficiency were not included, to some extent the effect of Hatfield was incorporated in the special treatment of Network Rail. Going forward, the effect of Hatfield could be tested explicitly by including dummy variables for the pre- and post-Hatfield periods. The estimates would provide information on whether the behaviour of the inefficiency term was different in pre- and post-Hatfield periods (see section 5.3.3). While the interpretation of the dummy for Hatfield in this instance would require further consideration, the key focus of the analysis is on Network Rail's efficiency at the end point of the period, so the critical issue is ensuring that this is estimated as robustly as possible.

5.4 Statistical and economic testing

Using the parametric approach, the ORR could generate a wide range of models, differing in functional form, explanatory variables and distributional assumptions. While all of these models are worth examining and useful for cross-checking the results, a definitive model (or a relatively small set of models) needs to be identified in order to be used as part of the price review. Model selection is an empirical issue, and statistical and economic testing is required to reach a conclusion.

The robustness of econometric models can be evaluated in light of a series of statistical tests (also known as diagnostic tests). The key issues to be tested are discussed below.

- **Functional form and model specification**: where possible, the choice of functional form should be based on empirical tests. It is recommended that, initially, several functional forms are considered, including flexible functional forms. The modelling process should then follow a general-to-specific approach and tests should be undertaken to identify insignificant variables that should be excluded from the regression. The model specification should then be tested in order to detect potential non-linear relationships between modelled cost and cost drivers. If the functional form is deemed to be an issue then non-parametric approaches should be considered (see sections 4.3 and 5.2).

- **Significance of explanatory variables**: a general-to-specific modelling process should be used to help in the identification of a model with significant explanatory variables.

---

[62] In this sub-section, the focus is on the specification of inefficiency. However, an important technical point to be aware of is the potential risk of over-parameterisation—for example, explaining inefficiency with country dummies may result in over-parameterisation when other, more parsimonious, specifications exist (eg, using explanatory variables rather than dummy variables). However, as discussed in section 3.3, country dummies should at least be used in order to allow for, and examine, the possibility of country-specific effects in the frontier technology. Unless this is undertaken, country-specific effects in inefficiency might be conflated with some effects of technology.

– **Outlier tests** should be undertaken at the modelling stage to identify potential data problems.

– **Economic implication**: diagnostic tests help to identify a model that is statistically robust. However, it is equally important that the model makes sensible predictions in light of economic intuition and industry-specific knowledge. This requires the estimated coefficients to exhibit the expected signs and size. Counterintuitive results would indicate potential issues, such as colinearity in explanatory variables. Network Rail might be in the best position to comment on this issue, based on its industry expertise. This suggests the need for a collaborative approach in the development of the models.

– **Structural break**: country-specific factors or shock events could lead to structural breaks in the model—ie, some of the coefficients may vary across time or company, instead of staying constant. A common approach to identify any potential structural break is to estimate a flexible model by including in the regression dummy variables for individual companies and time periods, and then to test these dummy variables for statistical significance. Significant dummy variables would indicate a structural break around a company or a certain period of time. This approach identifies where a structural break lies, and the significant dummy variables can be included in the final model in order to account for the impact of the structural break. There are several further approaches that can be used to examine this issue, such as the Chow test,[63] rolling regression and outlier tests.

– **Heteroscedasticity**: a model is heteroscedastic if the random component has a non-constant variance. A number of statistical tests can be used to detect heteroscedasticity. As discussed in section 5.3, heteroscedasticity may arise because data from different countries is not equally robust. The model can be more accurately estimated by taking heteroscedasticity into account.

---

[63] Chow, G.C. (1960), 'Tests of Equality Between Sets of Coefficients in Two Linear Regressions', *Econometrica* **28**:**3**, 591–605.

6　Accounting for the uncertainty of the efficiency scores

> **Recommendation 6.1**: The degree of uncertainty in the modelling should be quantified.
>
> **Recommendation 6.2**: Given the (quantified and unquantified) uncertainty, the ORR should use the modelling results with an appropriate degree of caution when making efficiency assumptions.

6.1　Sources of uncertainty

Regardless of whether efficiency scores are estimated using parametric, semi-parametric or non-parametric approaches, they are subject to some degree of uncertainty. This uncertainty can be due to:

– modelling errors, including, but not limited to:

  – omitted explanatory factors;
  – incorrect functional form (for parametric modelling);[64]
  – differences between companies and countries not accounted for in the modelling or through the use of special factors;

– data errors, including data adjustments such as the use of PPPs, etc;

– statistical uncertainty or noise, which may or may not account for some of the above factors.[65]

The following sections consider how this uncertainty can be accounted for.

6.2　Quantifying uncertainty

6.2.1　Econometric modelling

In econometric estimation, the output provides point estimates for parameters, such as the coefficients in the cost function, but also the inefficiency estimates themselves. However, the point estimates of these parameters are different from their true values. The uncertainty in the estimators is captured by standard errors. Estimates of inefficiency are based on estimated parameters and data, which makes efficiency estimates uncertain. Thus, similar to parameter estimates, when providing inefficiency estimates, the standard error or confidence interval of estimated inefficiency (which is observation-specific) should be reported. Depending on the approach taken, these can be calculated in different ways.

– SFA—the formulae for calculating these confidence intervals when using SFA can be found in Kumbhakar and Lovell (2000) and are provided in most software packages that run SFA models.'[66] Even if the focus is on Network Rail's efficiency for a particular year, the confidence interval of its estimated inefficiency can be calculated. In addition, uncertainty in the efficiency estimate depends on whether variances of the noise and

---

[64] DEA, for example, is more flexible than, say, a Cobb–Douglas functional form, in that it uses an implicit piecewise linear form which can approximate many underlying (unknown) functional forms.

[65] See, for example, Simar, L. and Wilson, P. (2008), 'Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives', in H. Fried, K. Lovell and S. Schmidt (eds) (2008), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

[66] Kumbhakar, S. and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press.

inefficiency components are constant. Ignoring heteroscedasticity in the noise and/or inefficiency components produces statistically biased estimates of inefficiency, as documented in section 3.4 of Kumbhakar and Lovell (2000). As a result, one should be careful not only in choosing the functional form and distributions of noise and inefficiency, but also in checking for heteroscedasticity.

– COLS—it is not tractable to produce confidence Intervals for the inefficiency estimate based on COLS, but one can provide some indication of the uncertainty using bootstrapping.[67]

6.2.2 DEA
Uncertainty for the DEA-based inefficiency scores can be calculated using bootstrapping.[68]

Other sources of uncertainty and regulatory precedent
Although the above discussion demonstrates that the uncertainty surrounding the inefficiency estimates can be quantified, these confidence intervals are predicated on the model being correct. However, there is still inherent uncertainty in the inefficiency estimates due to remaining issues, such as data errors, omitted variables, omitted special factors, incorrect functional form, that are not captured in the standard errors.[69]

For pragmatic purposes, regulators adopt a number of approaches to translating the efficiency results from their modelling into efficiency assumptions for a price review. For illustration purposes, the discussion below focuses on a number of regulators' approaches to accounting for this uncertainty, but not the subsequent adjustments they make to provide companies with incentives to outperform (as this is outside the scope of this study). The approaches adopted when making these adjustments a highly dependent on the context, including the quality of the data, the estimation approach adopted, etc. For example:

– to accommodate modelling errors, Ofwat reduces the residuals estimated by COLS by 10% for water services and 20% for sewerage services. To attempt to ensure that the benchmark company is robust, Ofwat requires that the benchmark passes a number of criteria, including that the company should represent at least 2–3% of industry turnover. Finally, Ofwat places each company into an efficiency band (eg, Band A is within 5% of the frontier, Band B between 5% and 15%, Band C between 15% and 25%, Band D between 25% and 35%, and Band E greater than 35%).

– Ofgem also makes adjustments to its benchmark selection. Its modelling is based on COLS. In its Initial Proposals, it set the benchmark at the upper quartile for indirect costs. However, for network operating costs, where Ofgem considered that there were more data issues, it used a hybrid approach, with the upper quartile taken as the benchmark for more efficient companies, and the average for less efficient ones.[70] Following its October update, Ofgem used the upper third as the benchmark for network operating costs and continued to use the upper quartile for indirect costs.[71]

---

[67] See, for example, Competition Commission, (2007), 'South East Water Limited and Mid Kent Water Limited: A report on the completed water merger of South East Water Limited and Mid Kent Water Limited'.

[68] See, for example, Simar, L. and Wilson, P. (2008), 'Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives', in H. Fried, K. Lovell and S. Schmidt (eds) (2008), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

[69] All of these issues should be mitigated as far as possible during the modelling phase.

[70] For further details, see Ofgem (2009), 'Electricity Distribution Price Control Review Initial Proposals - Allowed revenue - Cost Assessment', August 3rd, p. 77.

[71] Ofgem (2009), October update to Initial Proposals, October 5th.

6.3 Summary

To some extent, the uncertainty surrounding the efficiency estimates can be quantified, but other uncertainties remain which also should be accounted for. This residual uncertainty requires a degree of regulatory judgement and qualitative adjustment.

www.oxera.com