Article:

Smith, ASJ, Wheat, PE and Smith, G (2010) The role of international benchmarking in developing rail infrastructure efficiency estimates. Utilities Policy, 18 (2). 86 - 93. ISSN 0957-1787

https://doi.org/10.1016/j.jup.2009.06.003

The role of international benchmarking in developing rail infrastructure efficiency estimates

Dr Andrew Smith and Phill Wheat*, Institute for Transport Studies, University of Leeds

and

Gregory Smith, Office of Rail Regulation*

Abstract

International cost efficiency benchmarking played a central role in informing the Office of Rail Regulation's (ORR) determination of Network Rail's future funding during the 2008 periodic review (PR08) of the company's finances. This paper sets out how international benchmarking can inform a regulator's decisions on efficiency and, in particular, how international econometric studies can be used alongside other evidence in the regulatory context. We start by reviewing the use of previous international benchmarking work. We then set out the data, methodology and results in respect of the two separate econometric studies carried out as part of PR08. The further work that was done in support of the econometric results is then described. The paper shows that top-down econometric techniques, combined with bottom-up engineering analysis produced comparison between Network Rail and its peers. We conclude by outlining how the econometric results were used, in conjunction with other evidence, to reach a final efficiency determination, and how we consider that international benchmarking can be applied by other regulators.

* Dr Andrew Smith is a lecturer in transport economics at the Institute for Transport Studies, University of Leeds. Phill Wheat is a research fellow at the same institution. Greg Smith is a regulatory economist at the Office of Rail Regulation.

We would like to thank Gerard Dalton of the International Union of Railways (UIC), Network Rail and the other infrastructure managers who provided data to us and worked with us during the course of this study.

1. Introduction

This paper constitutes part of the output for the Centre for Competition and Regulatory Policy workshop hosted by City University on 23 January 2009. It is based primarily on the econometric analysis completed by the Institute for Transport Studies (ITS), University of Leeds and the Office of Rail Regulation (ORR) as part of ORR's 2008 periodic review (PR08) of Network Rail's outputs and funding for 2009-2014. As part of the periodic review, ITS and ORR undertook extensive international benchmarking work. This work formed an integral part of ORR's judgement regarding the scope for Network Rail, the owner and operator of the British rail infrastructure, to improve its efficiency performance, and in turn on the company's allowed funding. International benchmarking is particularly important in the case of Network Rail, as there are no domestic comparators against which to judge its efficiency.

This paper summarises the policy and technical aspects of ORR's work on international benchmarking and how this work informed ORR's final determinations on efficiency. Although we present our preferred econometric results in this paper, we reference Smith, Wheat and Nixon (2008) and Smith (2008) for a full explanation of the technical aspects behind the work done by ITS and ORR and the wide range of other econometric models estimated. As well as discussing the role of international benchmarking in PR08, we also generalise and comment on the circumstances in which international benchmarking is likely to be most appropriate and practical for other economic regulators.

The remainder of the paper is organised as follows. Section 2 places international benchmarking into context and outlines previous work done by ORR to benchmark Network Rail against its peer infrastructure managers. Sections 3 and 4 describe the dataset and methodology used for the main econometric work based on the International Union of Railways' (UIC) Lasting Infrastructure Cost Benchmarking (LICB) dataset, which was provided by the UIC for use in this work. Section 5 presents the results from that analysis. Section 6 shows the results of a supporting econometric study conducted using geographically disaggregated data (within each country) for a different sample of infrastructure managers. This data was collected directly by ORR and ITS.

Section 7 provides examples of additional, supporting studies conducted or commissioned by ORR and discusses how ORR combined the results of the different studies to form its overall judgement on the scope for Network Rail to improve its efficiency in the five year control period starting on 1 April 2009. Section 8 offers our conclusions.

2. International benchmarking in context

Assessing efficiency is an important part of an economic regulator's work. Regulators are required to take a view on the regulated industry's potential for efficiency savings as part of any price control or periodic review. ORR, like other regulators, faces a problem in that Network Rail is a natural monopoly. In order to ensure that Network Rail operates, maintains and renews the network efficiently, ORR regulates Network Rail using a form of conventional 'RPI-X' incentive regulation.

Assessing the X factor is not a straightforward task. In the most basic sense, X represents the total factor productivity (TFP) growth that a regulated firm can achieve over and above that which is achieved by the economy as a whole (which is implicit within RPI).[1] Potential TFP growth may be further broken down into that resulting from technical progress (e.g. the introduction of new technology), changes in scale, and catch-up efficiency relative to some appropriate benchmark (for example, international best practice) if the company is not at the efficiency frontier.

The scope for catch-up efficiency as compared to relevant benchmarks has been the main focus of the analysis during PR08, given the very sharp rise in costs after the Hatfield accident[2] in 2000. There are various methods that a regulator uses to take a view on potential catch-up efficiency. Using 'bottom-up' analysis, individual initiatives are identified, and their efficiency impact aggregated into an overall efficiency target. Within 'top down' analysis the regulator forms a view on potential efficiency by benchmarking the company against its peers in the sector, usually either other regulated firms in the same country, or internationally, typically using econometric analysis. Internal benchmarking between different parts of the company can also by employed.

Usually the 'bottom-up' approach is expected to underestimate the potential for improvement. Within the top-down approach, internal benchmarking is likely to show lower potential efficiency than methods based on external comparisons (e.g. other regulated firms or international best practice).

During previous periodic reviews, ORR conducted work to establish the scope for employing international benchmarking as part of its efficiency assessment (see NERA (2000)). This work largely focused on North American, Australian and East Asian railways, although the report concluded that there was insufficient data in the public domain to draw meaningful conclusions on comparative efficiency levels. However, the report did produce some evidence of trend total factor productivity growth which was used by ORR alongside other evidence on productivity and unit cost trends in UK regulated industries.

ORR also conducted work on international benchmarking as part of the 2003 access charges review though from a bottom-up perspective (see LEK, Halcrow, TTCI (2003)).

In late 2005, ORR and ITS began the process of developing a 'top-down' international benchmarking methodology with a view to producing useable results for PR08 and the longer term. The first study, carried out in

---

[1] Regulators also may make adjustments to the extent that a regulated firm faces different input price trends to those experienced in the general economy.

[2] A train derailment, resulting from defective track, which resulted in four people being killed.

conjunction with Network Rail and the UIC, undertook work to benchmark the company against 12 other West European infrastructure managers.

One of the biggest hurdles to overcome in international benchmarking is ensuring that one is comparing like for like. The UIC has gathered data on these infrastructure managers for 11 years, and during that time work has been done to ensure the comparability of the data through, for example, the specification of common data definitions. Representatives from each participating company also meet regularly and the data is used as part of UIC's own benchmarking analysis. UIC agreed to provide this data for the purpose of econometric work in support of PR08, on the basis that only results for Network Rail would be published and the confidentiality of other infrastructure managers' data and relative efficiency respected. The UIC data set covers: costs; network size; outputs such as passenger and freight train kilometres; and network characteristics such as the proportion of track electrified and numbers of points units per track km.

ORR and ITS also worked with five other rail infrastructure managers in Europe and North America to develop a new dataset. This includes data on costs, outputs, and network characteristics at the regional level within each country. Thus, although the number of companies included is smaller than in the LICB dataset, the sample size is expanded via the use of regional data within companies. It is still new and emerging and we are working to expand the coverage and improve the harmonisation of definitions.

Having access to an existing, good quality dataset from UIC allowed the ORR to give international benchmarking an important position in its final decision making on the scope for Network Rail's efficiency improvement. Likewise, given time, the use of the regional international dataset, which ITS and ORR collected from scratch, demonstrates that good progress can be made if regulators and companies aim to develop a benchmarking framework over a number of years, working between periodic reviews, rather than having to rely on studies commissioned at each review within a constrained timeframe.

3. The LICB dataset

As part of its own benchmarking analysis the UIC has developed a very useful dataset. It consists of data for 13 national rail infrastructure companies in western Europe, or infrastructure divisions within integrated companies over a period of eleven years.

As noted above, we have reason to be relatively confident in the consistency of the LICB data, given the efforts made to standardise definitions. UIC uses this dataset in its approach to international benchmarking (see UIC (2007) and ORR's 2008 final determinations for further details). The availability of multiple years for the same companies is also highly advantageous, as it avoids the danger of focusing on a single year snap-shot which might be impacted by year-to-year fluctuations in expenditure unrelated to efficiency. Furthermore, the dataset contains a wide range of variables in addition to the key measures of track length and traffic volumes.

Below we list the key variables from the LICB dataset used in the econometric analysis[3]. The data provided by UIC contained in excess of thirty variables. The variables that were ultimately included in our analysis are listed below. These were the variables for which there was sufficient coverage of the data across the different companies and years. Table 1 summarises the coverage of the UIC dataset.

[3] The dataset was taken largely as given, although a small amount of data cleaning was carried out. See Smith (2008).

Table 1: Coverage of the UIC dataset

Cost Data	Network Size	Final Outputs	Network Characteristics
Maintenance costs Total costs (Maintenance + renewals)	Track kilometres Route kilometres Single track kilometres Electrified track kilometres	Passenger train kilometres Passenger tonne kilometres Total tonne kilometres Freight train kilometres Freight tonne kilometres Total train kilometres	Ratio of single track to route kilometres (as a measure of the extent of single / multiple track) Proportion of track electrified Number of stations per route km Number of points units per track km

Whilst the above list includes a wide range of potential cost drivers, ideally we would have wanted to include “quality” measures (e.g. track geometry), as well as asset age. The issue of track quality / age is also related to the question of whether Network Rail’s renewal volumes are above steady-state (discussed in section 5 below). It would also be useful to explore the impact of different safety standards and possessions regimes on cost; however data is not currently available.

The cost data was converted to a common currency using Purchasing Power Parity (PPP) exchange rate data from the OECD. In this way, differences in national price levels which affect costs were controlled for. To control for inflation, the data was then deflated to a common year price level (with German Euros in 2006 as the numeraire). These real costs were then the final figures used in our analysis. We did not have sufficient information to separately include a wage rate variable in the model, although general (economy wide) wage rate differences will be accounted for via the PPP adjustment.

4. Methodology

In this section, the econometric methods used to relate maintenance and renewals costs to the relevant cost drivers included in the LICB database are explained. These methods allowed us to test the extent to which costs can be explained by the relevant cost drivers and, once those relationships were estimated, to use that information to estimate the relative efficiency of a company compared to other companies in the dataset.

In broad terms there are two approaches to conducting efficiency analysis; econometric methods and data envelopment analysis (DEA). In line with general practice in efficiency analysis (as for example, carried out by economic regulators, governments or academic studies) it is important to avoid over reliance on one particular method in arriving at results.

In the present context we consider that econometric methods are more suitable, since the (stochastic frontier) econometrics techniques permit the error term to be decomposed into random noise and inefficiency (unlike DEA, which is deterministic). Furthermore, the econometric model produces estimates of the elasticity of costs with respect to the cost drivers included in the model and, if these estimates are considered plausible, this gives us confidence in the efficiency scores derived. We also note that the DEA models that we did run produced similar results to the econometric methods (see Smith (2008)).

There are a range of econometric methods that might be applied in the present context. Our preferred model enables efficiency to vary over time, with firm-specific variation in the extent and direction of efficiency change. It has the following error structure:

$$\varepsilon_{it} = v_{it} + u_{it} \tag{1}$$

where the *i* and *t* subscripts refer to the firms in the sample and time respectively. The (v_{it}) term is a random, stochastic, component representing unobservable factors that affect the firm's operating environment (often referred to as random noise). This term is distributed symmetrically around zero. A further one sided error term is then added to capture inefficiency (u_{it}), where the inefficiency term in turn has the following time varying structure (and is based on the model proposed by Cuesta (2000)):

$$\begin{aligned} u_{it} &= u_i \cdot g(t) \\ g(t) &= \exp(\eta_i \cdot (T - t)) \quad t = 1, \ldots, T \end{aligned} \tag{2}$$

where the η_i are a set of firm specific parameters to be estimated, u_i has a one sided normal distribution with zero mean and variance σ_u^2, and T is the number of years covered by the sample. If η_i is positive for an individual firm, this indicates that efficiency is improving for that firm over time, and vice versa for a negative η_i.

This model therefore captures a number of important and desirable features for efficiency estimation. First of all, it is one of a class of models, referred to as stochastic frontier models, that distinguishes between random noise and inefficiency, and therefore is not overly influenced by outliers. Second, it allows efficiency to vary over time and in flexible manner, so that a time varying efficiency parameter is estimated for each firm. Therefore, the direction and extent of efficiency variation over time can be different for each firm. Third, the variation in efficiency over time is nevertheless structured, and not random - and the model thus recognises the panel structure of the data (so it recognises that we have a dataset of thirteen firms over 11 years, and not simply a dataset of 143 firms).

Furthermore, the preferred model incorporates additional flexibility in respect of the time path of efficiency for Network Rail. First, the sample was split, such that the observations for Britain are treated as two firms, one for the pre-Hatfield period (1996 to 1999 inclusive), and one for the post-Hatfield period (2000-2006)[4].

Second, an additional squared term was incorporated into equation (2) for Network Rail which allows the model to potentially capture a turning point in inefficiency during the post-Hatfield period. This was considered important as efficiency deteriorated during the early post-Hatfield years, before improving during Control Period 3 (CP3)[5]. In general, incorporating additional flexibility into the efficiency time path for Network Rail seems desirable given the substantial changes in costs that have occurred over the period under analysis.

As described in Smith (2008), we also applied a wide range of other efficiency measurement approaches in order to validate the preferred model. These include the corrected ordinary least squares (COLS) model, which is deterministic, and so does not distinguish between noise and inefficiency. In the panel data context, it also treats each observation (across firms, and over time) as a separate firm, which is unrealistic.

The other panel data stochastic frontier models estimated include models that assume inefficiency to be constant over time (Pitt and Lee (1981); and Schmidt and Sickles (1984)), and a simpler time varying efficiency model, where all firms are forced (by assumption) to have the same direction of efficiency change over time (Battese and Coelli (1992)); an assumption that could be particularly restrictive in the present context, where not all firms in the sample might be assumed to have followed a similar profile of efficiency change as Network Rail. It is clear therefore that our preferred model offers considerably greater flexibility, and more plausible assumptions, compared to the alternatives.

A key point to note here is that several techniques were estimated prior to selecting the preferred model. This process ensured that the most appropriate model was selected, and also resulted in a range of estimates that could be used as a cross-check against the preferred model. Cross checking the results of benchmarking analysis of this sort in economic regulation by using alternative techniques is now considered regulatory best practice. There are also various test statistics that can help us in choosing our preferred models. Further details regarding the choice of methods are provided in Smith (2008).

It should also be noted that all of these methods have been used in the academic literature therefore providing a precedent. In a regulatory context,

[4] We note that the models with and without this separation produce very similar results.

[5] The five year period ending 2008/09. PR08 is concerned with Network Rail's funding over the subsequent five year period, starting 2009/10.

the COLS method has been used in other UK regulated industries (for example by Ofgem, Ofwat and Oftel (now Ofcom) and stochastic frontier analysis has also been used (e.g. telecoms and postal services). UK academics and water companies have applied panel data techniques to estimate relative efficiency in the water industry. However, whilst the advanced time varying efficiency modelling techniques adopted in this study have been used extensively in the academic literature, we are not aware of their use in regulatory studies.

5. Results – based on LICB data

This section sets out the results for our preferred model. It starts by explaining the key assumptions underpinning the model. It also explains the approach to dealing with potential swings in railway expenditure caused by renewal expenditure being above or below steady-state levels. The results are then presented.

5.1 Preferred model: key assumptions

Our preferred approach was to benchmark Network Rail based on total costs (maintenance and renewals (M&R)) together. This is more appropriate than considering maintenance and renewals separately, as both the trade-offs between M&R and any residual accounting differences that may exist between countries (in the way in which they record maintenance and renewals costs) are taken into account. We additionally modelled maintenance and renewals costs separately to act as a crosscheck. This produced similar results.

It should be noted that potential swings in railway expenditure from year to year (especially for renewals) could impact on our analysis. Following additional work carried out by ORR, an adjustment was made to Network Rail's costs to address the possibility that Network Rail is currently renewing at above steady-state levels, and ORR and Network Rail have also undertaken parallel analysis to assess the impact of other possible omitted variables on the analysis.[6] As such, ORR considered the adjustment made to Network Rail's costs to be a conservative assumption.

More specifically, Network Rail's renewals data was adjusted to make it consistent with 2.5% of total track and signalling assets being renewed in each year, implying an average life of 40 years for these assets. This adjustment increases the renewals cost data used for Network Rail in the years up to 2000 and reduces it thereafter. Since there was insufficient data to make similar adjustments for other railways, the approach also requires the assumption that the leading firms are broadly in steady-state. ORR carried out

[6] Steady state expenditure is defined as the level of expenditure needed so that the assets, on average, are retained in a constant and serviceable condition. Put another way, this reflects the average annual expenditure needed over the life of the assets, stripping out fluctuations caused by the lumpy nature of renewal activity and other factors affecting expenditure in individual years.

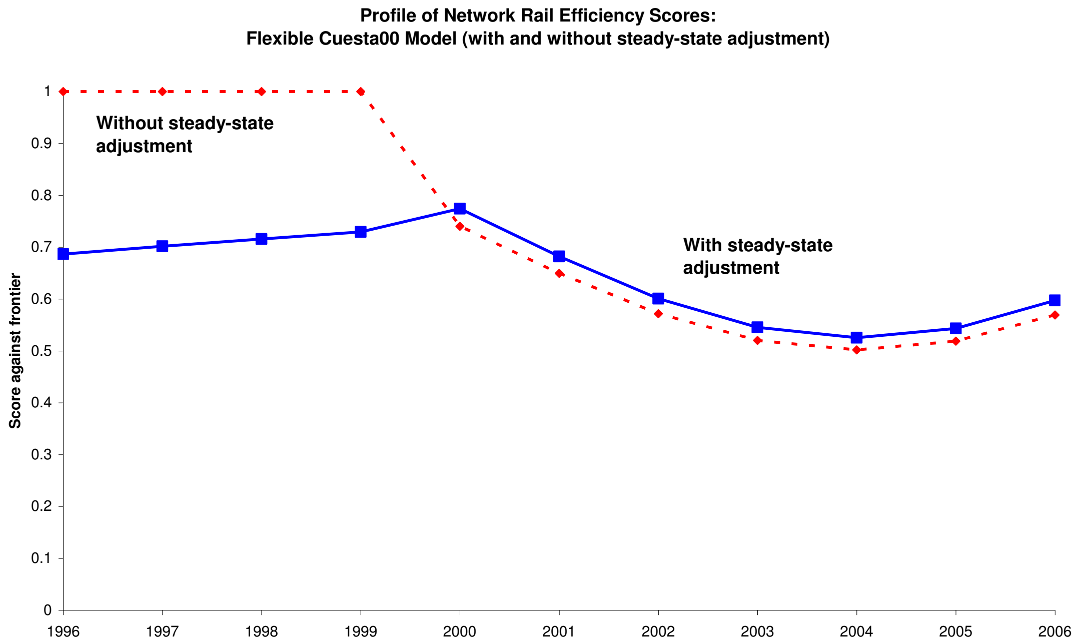
additional work in this area and concluded that there was no evidence to suggest that this was not the case.

It should also be noted that the stochastic frontier method makes allowance for noise in the data, such as might be caused by year-to-year fluctuations in expenditures relating to steady-state issues. As a result, the method guards against interpreting unusually low costs in a given year as reflecting efficient operation. In addition, we have not relied on analysis of data at a snap shot in time, but have looked at evidence over an 11 year period.

5.2 Preferred model: results

Based on the above discussion we go on to report our results for the preferred approach, both with and without the steady-state adjustment (Figure 1). The preferred model (with a steady-state adjustment to costs) produces a plausible pattern of efficiency change over time for Railtrack, the owner and operator of the rail infrastructure between 1995 and 2002, and Network Rail (the company that took over from Railtrack in 2002). Figure 1 shows efficiency improving modestly in the early years after privatisation (even after the steady-state adjustment), before deteriorating sharply during the post-Hatfield period, and then starting to improve once the efficiency savings being delivered by Network Rail during Control Period 3 (CP3; 2004-2008) start to have an impact.

Figure 1: Network Rail efficiency scores for the preferred model



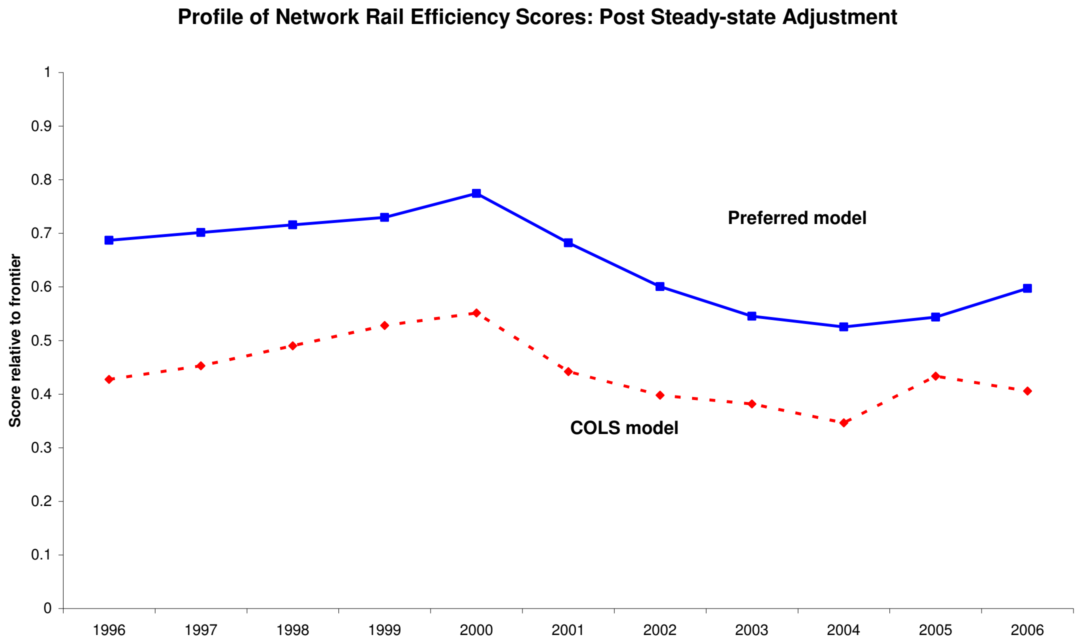
A slightly different picture emerges for the model without the steady-state adjustment, where Railtrack is found to be on the frontier during the first four years of the period, before efficiency starts to deteriorate sharply and then pick up again towards the end. This difference is to be expected, since the steady-state adjustment increases the costs for the British data during the early years of the dataset.

The models perform well statistically. The parameter estimates appear to be generally well behaved in terms of their signs, magnitudes and significance levels (broadly in line with engineering expectations and previous econometric work; see Smith, Wheat and Nixon (2008) and Smith (2008) for further details). With respect to the chosen functional form, we tested the Cobb-Douglas functional form against the linear and translog alternatives and found that for both versions of the total cost model the Cobb-Douglas form could not be rejected. In empirical work the Cobb-Douglas functional form is often seen as a practical, parsimonious alternative to the translog which, in the present case, is also supported by the relevant test statistics.

Furthermore, the efficiency element of the model specification is also shown to be robust. The null hypothesis of no inefficiency effects can clearly be rejected based on the standard Likelihood Ratio (LR) test (at the 1% level). Likewise, the simpler, Battese and Coelli (1992) time varying model, which is nested within our preferred model, can clearly be rejected (again at the 1% level). Thus it is important to allow different time paths of efficiency change for different firms, as in our preferred model. The null hypothesis that efficiency is time invariant for all firms is also clearly rejected at the 1% level. Thus we can be confident in the selection of our preferred model as compared with the relevant alternatives from a statistical perspective.

It should also be noted that the much simpler, COLS model produces a similar pattern, although with a lower absolute level of efficiency (as expected since the COLS model does not distinguish between efficiency and noise); see Figure 2 below. Thus, although we have put forward a relatively complex model as our preferred model, the time path of inefficiency produced is similar to that of a much simpler model.

Figure 2: Network Rail efficiency scores across method (costs adjusted for steady-state)



The preferred model produces an efficiency score for Network Rail in 2006 of 0.57 (unadjusted) and 0.60 (steady-state adjusted; where a score of 1 represents the frontier). This puts the efficiency gap in the range 40% to 43% compared to the frontier. In using the results of these models, ORR benchmarked Network Rail against upper quartile (not the frontier), which is a conservative assumption; and reduces the starting efficiency assumption to 37% (to which ORR then made further adjustments, as discussed in Section 7). It should also be noted that the results of this preferred model are in line with those of a wide range of alternative efficiency estimation approaches, giving us added confidence in the results (see Smith (2008)).

6. International regional benchmarking

Between autumn 2005 and autumn 2007 ORR and ITS held discussions with and collected data from a number of rail companies with a view to carrying out an international benchmarking study. The participants were Infrabel (Belgium), ProRail (the Netherlands), Irish Rail, Amtrak (USA) and Network Rail. Data was provided for a number of geographic areas in each country and, for some countries, a number of years. Our goal was to verify the results of our work with the LICB dataset.

This was a relatively new approach in that it sought to utilise regional data within countries as part of an international comparison across countries. It did not form the main part of the econometric evidence used as part of PR08, but was used as supporting evidence to the main analysis based on the LICB data.

There are numerous advantages associated with this dataset. First, it provided an opportunity for ORR to collate a new data source direct from the companies concerned. Furthermore, the use of regional data within each country substantially increased the number of observations, which offers benefits in respect of modelling precision. We were able to acquire a consistent set of volume and network characteristic variables for all companies. However, the analysis was limited by the fact that for some companies only one year's data was available. We expect the benefits of this approach, in terms of the size and extent of the dataset, to improve over time. It also provides useful information for the companies concerned on their relative efficiency.

We applied similar techniques in assessing relative efficiency as those in the analysis of the UIC LICB dataset. This dataset was collected independently of the LICB data, and was based on an alternative aggregation of data; that is company level data was disaggregated into regions, in a similar way to the analysis that has been conducted in the past by Ofwat (except this time in an international context). The results are summarised below (further details are shown in Smith, Wheat and Nixon (2008)).

Table 2 below shows the efficiency results for the preferred models, based on a time invariant efficiency method. The preferred models are shown for two different cost categories and two different sets of output variables, but were selected from a wider range of models. The time invariant efficiency approach was considered reasonable, since we only had one year's data for some companies. The model also assumes that the regions within each country had the same efficiency. This approach was taken because it was considered preferable to the alternative of pooling the data with no panel structure, and because the focus was on company level and not internal efficiency for this study. We intend to apply more advanced alternative approaches that allow for internal inefficiency in subsequent analysis.

Table 2: Network Rail efficiency scores (regional international benchmarking work) compared to the frontier

Model	Output	Maintenance only model	Maintenance and track renewal model
Random effects (MLE)	TTKM	0.52	0.55
Random effects (MLE)	PTKM and FTKM	0.57	0.61
Average all models			**0.56**

Note: TTKM=total tonnage density (passenger plus freight tonnage per track-km); PTKM=passenger tonnage density and FTKM=freight tonnage density

The results are shown based on the time invariant efficiency maximum likelihood (MLE) method for two models: the first which is based on total traffic (tonnage) density as the key measure of volume on the network, and the second which included separate variables for passenger and freight tonnage density. The model also included a measure of network size and a variable capturing the extent of electrification on the different networks. We aim to incorporate further variables in the future.

The results show an average efficiency gap across the different models shown in Table 2 of 44%, with Network Rail ranked 4th out of the 5 companies. A wider set of methods was also applied with similar results. The efficiency gap is therefore in line with that resulting from the analysis of the LICB dataset, which produced a gap against the efficiency frontier of 40-43%. The preliminary analysis of the regional (international) data has produced a reasonable set of results. It is expected that future development of the dataset and methodology will further strengthen the robustness of the analysis.

7. International benchmarking in context – evidence in support of the econometric results

In this section, we summarise the additional work undertaken by ORR aimed at verifying the results of the econometric study. ORR considered that it is important to approach any application of results from international benchmarking with a degree of caution and to verify the results. There are clear and robust statistical reasons to apply the preferred model in its own right, and the econometric work thus forms a good starting point to assess efficiency. However, in reaching its final determination, ORR wanted to satisfy itself further by looking at a range of other evidence (all of which showed significant scope for catch-up efficiency). Several examples of further top-down and bottom-up evidence supporting the results of the econometric study are presented below.[7]

A degree of uncertainty exists in using international benchmarking in a policy context. The econometric evidence itself was however robust to a variety of methods and sensitivity tests, and it consistently demonstrated a 30%-50% efficiency 'gap' between Network Rail and the upper quartile of infrastructure managers in our peer group. Further work undertaken by ORR identified areas where technologies and working methods explain the difference in the cost 'gap' between Network Rail and European practice. Additionally, a variety of other top-down approaches undertaken by ORR, Network Rail and their consultants confirmed the results of our econometric study.

A brief outline of their results from our work with the UIC dataset is presented in the Table 3 below.

[7] Extensive work was undertaken to confirm the cost 'gap' between Network Rail and European best practice including but not limited to studies presented in this paper. For a review of ORR's evidence in support the PR08 efficiency analysis, see: Periodic review of Network Rail's outputs and funding for 2009-2014, chapter 7. http://www.rail-reg.gov.uk/server/show/category.1917, 30 October 2008.

Table 3: Comparison of results

Study	Efficiency gap (value / range)*	Comments
ORR/ITS international benchmarking (gap at 2006-07)	43% / 36 – 50%	No steady-state adjustment; to frontier of peer group**; M&R
	40% / 30 – 46%	With steady-state adjustment; to frontier of peer group*; M&R
	42% / 38 – 49%	No steady-state adjustment; to upper quartile of peer group; M&R
	37%*** / 24 – 43%	With steady state adjustment; to upper quartile of peer group; M&R

Source: Office of Rail Regulation (2008), *Periodic review of Network Rail's outputs and funding for 2009-2014*.

Extensive work was undertaken in order to support our analysis. Network Rail commissioned BSL Management Consultants to examine the UIC dataset and conduct work on international benchmarking, reflecting the importance of international benchmarking both in the PR08 efficiency debate and to the industry as a whole. BSL's work encompassed a unit cost analysis and found a 50%-70% expenditure 'gap' between Network Rail and its European peers based on their analysis.[8] BSL then explained part of this gap leaving a smaller, though significant residual efficiency gap.

ORR also conducted several international visits aimed at increasing understanding of international best practice. The visits revealed significant scope for efficiency improvements. EWS, the freight train operator, compared Network Rail's efficiency to North American railroads. ORR also commissioned RailKonsult, part of Balfour Beatty Rail, to conduct further bottom-up work. The results of their study are discussed below.

7.1 RailKonsult Study

In order to understand the more detailed differences in the level of cost between Network Rail and European practice, ORR commissioned a study by RailKonsult.[9] The objective was to examine which technologies and working methods used in Europe could help account for the differences in the cost gap between Network Rail and the LICB comparators. It was not the purpose of the study to identify and analyse all technologies or working methods used in Europe that could be introduced in Great Britain, but rather to identify some

[8] See: 'Rail infrastructure cost benchmarking – Brief LICB-gap analysis and cost driver assessment'. 2008. 'www.**networkrail**.co.uk/browse%20documents/StrategicBusinessPlan/Update/Cost%20benchmarking%20assessment%20(**BSL**).pdf

[9] *Review of European renewal and maintenance methodologies – overview,* RailKonsult, May 2008. The overview may be accessed at http://www.rail-reg.gov.uk/upload/pdf/pr08-konsovw-290508.pdf. Detailed appendices may be accessed at http://www.rail-reg.gov.uk/server/show/ConWebDoc.9145 - consult.

methods which are currently being investigated by Network Rail, and several others which are not. Seven initiatives were chosen for detailed study.

Table 4: Examples of European best practice

Asset inspection and asset management.	In general best practice European railways undertake fewer track inspections but inspections are generally of higher quality. It is estimated that similar techniques applied in Britain could reduce foot patrolling inspection costs by around 75% and tamping expenditure by 20%
Recycling components	This is common European practice. In Switzerland, for example, rail, point motors, sleepers and signal heads are regularly refurbished then cascaded from higher to lower category routes. Cascaded rail on lines re-laid with steel sleepers could lead to savings. Additionally ballast cleaning (partial renewal) as opposed to traxcavation (complete renewal) could reduce ballast renewal cost in Britain by 40%
High output rail stressing	Stressing continuously welded rail by heating it rather than physically stretching it is a process discontinued in Britain in the 1960s and 1970s. Some European networks (using modern equipment) have re-introduced this method which doubles on-site productivity and, if applied to the renewals re-railing workbank in CP4, could lead to significant annual savings for Network Rail
Formation rehabilitation trains	Modern high output European plant is regularly used to undertake formation and also ballast renewals. If applied to Network Rail's CP4 category 7 and 12 track renewals RailKonsult estimate that it could reduce unit costs for both activities by around 40%
Lightweight station platforms	The use of modular construction polystyrene station platforms in the Netherlands could provide opportunities in Britain, given the substantial CP4 platform extension workbank. Analysis suggests a unit cost saving of around 25% in Britain
Efficient European re-railing techniques.	This particular study brought together many themes from the previous RailKonsult work by focussing upon the Swiss re-railing method. Bespoke plant, high output welding techniques and dedicated teams are applied routinely. Put together for basic re-railing work alone this method is around 40% more efficient than current Network Rail practice
Use of dedicated teams	Contractors are widely used by most continental railways, as they are in Britain. However there is generally a greater degree of specialisation by activity in Europe (such as S&C renewal or tamping). This ensures a highly skilled and productive workforce dedicated to particular tasks in contrast to the situation in Britain where contractors are often not even dedicated to rail.

ORR found that this work provided strong support for the findings of the econometric analysis that the cost gap between Network Rail and the comparators used in the LICB dataset is due to inefficiency. ORR considered that most of the practices described in the RailKonsult report are readily applicable to the British railway network and point towards greater efficiency savings than those identified by Network Rail as part of the 2008 periodic review.

7.3 Use of the econometric results as part of PR08

This section summarises the way in which ORR utilised the econometric international benchmarking study and other evidence in arriving at its final conclusions. First of all, ORR ultimately made its comparisons of Network Rail against the upper quartile of the peer group, rather than the frontier, thus meaning that the starting efficiency gap for its analysis – based on the preferred econometric model from the analysis of the LICB data- was 37% rather than 40%.

ORR's judgement was informed by many pieces of evidence suggesting a significant level of inefficiency, and a cautious interpretation of the results of the econometric analysis was used. Given the robustness of the international benchmarking, ORR used this as the 'numeric' basis for its determination on efficiency. ORR believes that the steady-state adjustment was favourable to Network Rail. Additionally, whilst it is common to benchmark against upper quartile when using deterministic frontier approaches, in particular, COLS, which do not take account of random noise, this is not so where a stochastic frontier model is used (as in this case). These points underscore the point that ORR reached its final conclusions on efficiency using a conservative approach.

In its final determinations, ORR also concluded that Network Rail could close two-thirds of the 'gap' during control period 4 (CP4; 2009-2014). ORR argued that Network Rail should be well on the way to closing the 'gap' by the end of CP4 as it aspires to be a world class company. Thus the starting efficiency gap of 40% to 43% against frontier resulting from the preferred econometric model, both with and without the steady-state adjustment, was ultimately reduced to an efficiency target of 22% for CP4. Reflected in this figure is also an assumption about Network Rail's specific input price growth, as well as frontier shift efficiency. As part of this process some assumptions were also made about future frontier shift and input price changes that largely cancelled each other out.

Other regulators, including OFWAT, set targets based on a 60% catch-up assumption. ORR felt that requiring Network Rail to catch-up to two thirds of the gap was in line with regulatory practice. Thus, in our view, ORR adopted a conservative approach, which was supported by extensive bottom-up evidence, and therefore provides an accurate and robust measure of Network Rail's potential for efficiency improvements during CP4.

8. Conclusions

International benchmarking has proved to be a useful and important tool in informing ORR's judgements on the scope for efficiency improvement. The econometric work based on the UIC data proved to be robust in respect of different methodologies and various sensitivities. The supplementary, regional international benchmarking econometric study also provided additional support for the main econometric work. Furthermore, ORR interpreted the results conservatively, and also drew on a wide range of other analyses, including engineering-based bottom-up studies.

To what extent, however, can this type of approach be generalised and applied by other economic regulators? A key success factor in this case was the fact that a high quality dataset already existed (the UIC's LICB dataset);. Furthermore, the UIC was interested in considering new approaches to analysing this data for its own purposes, separate from the objectives of ORR in respect of PR08. UIC thus agreed to provide the data for our study.

It is worth noting that the international benchmarking workstream of PR08 started three years before the determination was published, emphasising the fact that the development of an international benchmarking framework can take many years to develop, even under the relatively favourable conditions in this case, with the existence of a "ready-to-go" dataset.

Having access to an existing, good quality dataset from UIC therefore clearly impacted on the ability of ORR to place international benchmarking at the forefront of its efficiency analysis. That said, the use of the regional international dataset, which ITS and ORR collected from scratch, demonstrates that good progress can be made if regulators and companies aim to develop a benchmarking framework over a number of years, working between periodic reviews, rather than having to rely on studies commissioned at each review within a constrained timeframe. The preliminary results from the international regional benchmarking work confirmed the results from the LICB study.

Economic regulators should therefore not necessarily be deterred by challenges and work involved in international benchmarking; but the lesson from our study is that work needs to start early, and that the some of the benefits will not be felt until future price reviews. ORR is currently considering how the current approach can be retained, and developed, for use in CP5 and beyond.

Of course, regulators need to recognise the inherent uncertainties involved in undertaking international comparisons. It is therefore important that the econometric results are shown to be robust, and that the results are not overly sensitive to the methodology or particular model adopted (except where there are strong reasons to prefer one model or approach over another). The fact that the preferred model in this case was robust on its own terms, and also with respect to changes in the method and other sensitivities further increased improved confidence in the findings.

Furthermore, as was the case in PR08, given the special challenges posed by international benchmarking, it is prudent for regulators interpret the econometric results conservatively, and combine them with a wide range of other evidence, and in particular bottom-up engineering based studies.

Finally, we note that, for the case of rail infrastructure in Britain, the very large cost rises experienced since Hatfield meant that external benchmarks, based on hard data, were crucial to ORR's analysis. The lack of domestic comparators for obtaining comparisons of productivity levels thus increased the importance of international benchmarking in this case.

The availability of domestic comparators in other sectors means that the benefits of international benchmarking may not necessarily justify the investment in all cases, at least in the short term. However, obtaining an international perspective is important in any industry, and particularly where the ability to make domestic comparisons is being reduced by common (group) ownership of previously independent infrastructure companies. Whilst international benchmarking requires greater commitment by the regulator, we consider that the benefits can be significant. ORR intends to take forward and further develop its approach to international benchmarking for use in future regulatory price reviews.

References

Abrantes, P., Wheat, P., Iwnicki, S., Nash, C. and Smith, A. (2008), 'Review of Rail Track Cost Allocation Studies', Deliverable 1, EU Sixth Framework Project CATRIN (Cost Allocation of Transport Infrastructure Cost).

Battese, G.E. and Coelli, T.J. (1992), 'Frontier Production Functions and the Efficiencies of Indian Farms Using Panel Data from ICRISAT's Village Level Studies', *Journal of Quantitative Economics*, vol. 5, pp. 327-348.

Coelli, Tim; Estache, Antonio; Perelman, Sergio; Trujillo, Lourdes (2003). 'A Primer on Efficiency measurement for Utilities and Transport Regulators'. World Bank Institute Development Series. Washington, D.C. Pg. 6

Cuesta, R. A. (2000). 'A Production Model with Firm-Specific Temporal Variation in Technical Inefficiency: With Application to Spanish Dairy Farms' *Journal of Productivity Analysis*, 13, pp. 139-158.

LECG (2005), *Future Efficient Costs of Royal Mail's Regulated Mail Activities,* August 2005.

LEK, Halcrow, TTCI, 2003. Report to the Office of the Rail Regulator, Network Rail and SRA: International benchmarking. London

NERA. 2000. *Report prepared for the Office of Rail Regulation, Review of overseas railway efficiency.* London

Office of Rail Regulation. 2008. *Periodic review of Network Rail's outputs and funding for 2009-2014***.** London.

Office of Rail Regulation. 7 March 2008. *Expected Cost of Network Ownership: Network Rail and Key LICB Comparators: Initial Results*, London.

Office of Rail Regulation. 2008. *ORR Best Practice Study: A report on the programme of international visits carried out by ORR between July – October 2007 (Summary Report)*, London.

OFWAT (2005), 'Water and sewerage service unit costs and relative efficiency: 2004-05 report - Appendix 1: Econometric models'.

Pitt, M.M. and Lee, L.F. (1981), 'Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry', *Journal of Development Economics*, 9,43-64.

RailKonsult. 2008. *Review of European renewal and maintenance methodologies – overview*, London.

Schmidt, P. and Sickles, R.C. (1984), 'Production Frontiers and Panel Data', *Journal of Business & Economic Statistics*, vol. 2 (4), pp 367-374.

Smith A.S.J. 2008. Report prepared for the Office of Rail Regulation, International Benchmarking of Network Rail's Maintenance and Renewals Costs: An Econometric Study based on the LICB Dataset (1996-2006). Leeds.

Smith A.S.J., and Wheat, P.E. 2008. Report prepared for the Office of Rail Regulation, Response to LECG and Horton-4-Consulting's review of the ITS/ORR international benchmarking study. London.

Smith, A.S.J., Wheat, P.E. and Nixon, H. (2008), *International Benchmarking of Network Rail's Maintenance and Renewal Costs*, joint ITS, University of Leeds and ORR report written as part of PR2008, June 2008.

UIC (International Union of Railways). 2006. *Lasting Infrastructure Cost Benchmarking (LICB): Summary Report,* Paris..

Zarembka, P. (1968), 'Functional form in demand for money'*, Journal of the American Statistical Association*, vol. 63 (322): pp. 502-511