



UNIVERSITY OF LEEDS

A RESPONSE TO THE LECG AND
HORTON 4 CONSULTING
REPORTS ON THE ITS/ORR
INTERNATIONAL
BENCHMARKING STUDY
(BASED ON THE LICB DATASET)

Andrew Smith and Phill Wheat

Institute for Transport Studies

University of Leeds

October 2008

ITS

Table of Contents

EXECUTIVE SUMMARY	3
1. Introduction	9
2. The LECG report.....	11
2.1 Model documentation	11
2.2 Data issues.....	12
2.3 Steady-state assumptions	13
2.3.1 Our underlying assumptions	14
2.3.2 Evidence put forward by LECG on steady-state	15
2.3.3 Summary of this section.....	17
2.4 Possible omitted variables.....	17
2.5 Functional form.....	19
2.6 Specification of the preferred model	21
2.6.1 LECG comments on the preferred model	22
2.6.2 Justification for our preferred model	23
2.6.3 A summary of the model selection process	25
2.6.4 The BC92 model put forward by LECG	28
2.6.5 Sensitivity analysis.....	29
2.6.6 ORR's use of the econometric work in PR2008.....	31
2.7 The regional international benchmarking study	32
3. The Horton 4 Consulting report.....	33
3.1 The use of a frontier	33
3.2 Data quality and currency conversion issues.....	34
3.3 The omission of variables relating to the capital stock.....	34
3.4 The implications of extrapolating the time trend from the preferred model	35
3.5 The use of panel data.....	35
3.6 Split between error and inefficiency	36
3.7 The regional international benchmarking study	37
4. Conclusions	39
REFERENCES	44

EXECUTIVE SUMMARY

Introduction and overview

During 2007 and 2008, The Institute for Transport Studies (ITS), University of Leeds, undertook econometric work to inform ORR's judgement on an appropriate efficiency target for the infrastructure manager over Control Period 4 (CP4). This work was conducted in conjunction with ORR and Network Rail, with input also from UIC, and culminated in a joint ITS / ORR report published in June 2008 (see ITS/ORR (2008)). During this period we have welcomed the involvement of and interaction with both institutions (in respect of Network Rail, mainly via Charles Robarts, David Rayner, Matthew Clements and David Smallbone; and Gerard Dalton and Teodor Gradinariu at UIC).

The work was peer reviewed by Dr Michael Pollitt from the Judge Business School, University of Cambridge. This work was used by ORR, alongside other evidence, in setting out its Periodic Review 2008 (PR2008) draft determination on efficiency published in June 2008.

Post June 2008, Network Rail commissioned LECG and Horton 4 Consulting to challenge the econometric work conducted by ITS/ORR (based on the UIC dataset) and ORR's judgement on Network Rail's efficiency more generally. The consultants' reports formed part of Network Rail's response to ORR's draft determination, which was submitted to ORR, and published, in early September 2008. This report sets out ITS's response to the challenges raised by both sets of consultants, and clearly demonstrates the robustness of the econometric study. This demonstration is based on:

- the validity of the preferred econometric models in themselves both from a statistical and economic theory viewpoint;
- the vast array of supporting econometric evidence for the preferred models provided by the estimation of a wide range of alternative efficiency measurement methodologies; and
- the supporting econometric evidence provided by the regional international benchmarking study.

The challenges raised by Network Rail's consultants are therefore emphatically refuted. The key arguments are set out below. It should be noted that during our work some areas of analysis were outside the scope of ITS's remit (for example, collating evidence on the extent to which companies were above or below steady-state). In these areas we took advice from ORR, and where this was done it is indicated in the relevant section of the report. We also comment, in general terms, on the way in which ORR used the econometric results in arriving at its draft efficiency determination.

Functional form and theoretical properties of the cost model

First of all, LECG's assertion that the Cobb-Douglas functional form adopted in our study violates economic theory in the multiple output case is shown to be incorrect. LECG's assertion is based on a single quote from one textbook, which

has been taken out of the context of the wider theoretical and empirical literature, and indeed even the book from which the quote is taken. It is clear from the literature that the multiple output Cobb-Douglas cost function does not violate any required theoretical property of a cost function in a regulated industry, such as railways, where output levels are typically assumed to be exogenously determined (see, for example, Klein (1953), Nerlove (1965), and Coelli and Perelman (2000)).

Indeed, we have shown that this functional form is widely used in both academic and regulatory studies. The appropriate functional form is, instead, an issue for econometric testing and, as noted in the ITS/ORR June 2008 report, we have tested the Cobb-Douglas model and found it to be preferred to the alternatives. Furthermore, LECG have themselves utilised a multiple-output Cobb-Douglas cost function in their recent (2005) study of postal delivery office efficiency, so it is extremely puzzling that they have raised this issue in criticism of our work. This functional form has also been used by OFWAT, as LECG note in their 2005 study.

Statistical properties of the preferred model

Secondly, we have demonstrated that the preferred econometric model is robust, both in its own right, and in the context of the vast array of other methods that we have applied to this dataset. LECG's assertion that a "fix" is required in order for the model to produce an estimate is shown to be incorrect. Furthermore, we have shown that the method used to derive the variance co-variance matrix (from which the standard errors and hence the means of determining the precision of the estimates are derived) is an accepted and widely used approach, and that alternative testing procedures also provide support for the method we have used.

The preferred model produces plausible estimates for the model parameters, which are also statistically significant at the usual levels of significance¹. It also produces an extremely plausible time path of efficiency for Network Rail over the period: that is, improving after privatisation², deteriorating after Hatfield, before improving during CP3. As noted, our preferred model also produces similar efficiency estimates for Network Rail to those from the other methods that we have tested, and there is also strong conformity of efficiency rankings (for all firms) across the different methods applied.

Comparison of our preferred model against the model suggested by LECG

Thirdly, we have shown why we selected our preferred model over the alternative time varying efficiency model (BC92³) put forward by LECG. Indeed, the BC92 model is shown to fail the relevant statistical tests that are normally applied. In addition, the model, which forces all firms (by assumption) to have the same

¹ As demonstrated by the standard errors derived from the variance co-variance matrix, and alternative testing procedures as noted.

² Of course, the low costs experienced during the Railtrack period may have been indicative of inadequate work being carried out, rather than efficient operation. This issue is addressed via the steady-state adjustment carried out prior to estimation discussed below.

³ Battese and Coelli (1992).

direction of efficiency change over time, produces the surprising result that Network Rail's efficiency is improving every year over the post-Hatfield period (or over the whole 11 year period, depending on the precise model formulation).

The BC92 model results are clearly shown to be the outlier when compared against the alternative models that we have considered. LECG have put the BC92 model forward without reporting the results of any statistical testing of its properties, or commenting on the reasonableness or otherwise of the results in themselves and as compared to the preferred or other relevant models.

Data quality issues

Fourthly, we have shown that the assertions of both consultants regarding the quality of the data do not fit with the facts. The LICB dataset used in our work has been developed by UIC over a number of years now (starting in 1995), and forms the basis for its own benchmarking methodology. In its ten year report on benchmarking, UIC describes the development of its approach over the period of the analysis and notes that:

“Phase 5 [of the work] provided considerable insights into cost levels and mechanisms and gave useful advice to Infrastructure Managers in Europe and overseas”, and “A “lasting benchmarking function” was established to guarantee a platform for continuous comparison of costs and the tracking of trends”; see UIC (2007), page 19.

UIC also produces guidance on data definitions to aid harmonisation. We therefore disagree that the data should be viewed as being at an “experimental stage” as Horton 4 Consulting state (Horton 4 Consulting (2008), page 9⁴), since this is not borne out by the statements contained in UIC's own report on the data's use in its own benchmarking approach.

Furthermore, the consultants ignore the fact that ITS/ORR and Network Rail have had access to the UIC dataset since February 2007. The dataset was discussed early on in the project, and at no point during the period of our work has Network Rail expressed any serious concerns in this respect. We also noted in our June 2008 report that ORR had carried out detailed inspection work on the dataset prior to analysis. The only concrete examples that LECG are able to produce in respect of data quality relate to just two data points for one variable (number of switches) that is not included in the preferred model specification. More generally, it is puzzling at this stage of the project for Network Rail's consultants to argue that the dataset is such that it is unsuitable for analysis when Network Rail has not made that point earlier.

⁴ They state this to be Network Rail's view.

Omitted variables and steady-state assumptions

Fifthly, the consultants point to possible omitted variables, in particular in respect of input prices and capital quality / variables relating to railways being above or below steady-state in respect of their renewal volumes (where steady-state is defined as the level of expenditure that is broadly required to maintain the assets in a stable condition).

All of these points were raised during our discussions with Network Rail. ORR and Network Rail (via the BSL (2008) study) therefore conducted / commissioned parallel studies to understand the likely impact of omitted variables in respect of Network Rail (further work in this area was outside the scope of ITS's remit). Ultimately ORR concluded that there was no reason to believe that incorporating such variables would necessarily lead to a significant change in the model results and be favourable to Network Rail, since there will be factors which disadvantage Network Rail as well as benefiting it. ORR has also stated, however, that further work would be helpful to try to enhance the modelling process and improve our understanding of the cost differences in future.

ITS also took advice from ORR in respect of the specific arguments surrounding steady-state. ORR expressed the view that it was not convinced that Network Rail was significantly above steady-state by the end of the period under analysis. Nevertheless, a downward adjustment was made to Network Rail's costs during the post-Hatfield period, which ORR considered to be a conservative assumption (i.e. it benefited Network Rail in terms of its relative efficiency score). For example, in 2006, this meant that the total (maintenance and renewal) cost data for Network Rail was reduced by roughly 10% as compared to the raw data prior to estimation.

Since ORR did not have sufficient data to make the same adjustment to other firms, it was assumed that the leading firms were broadly in steady-state. ORR looked at the evidence and concluded that there was no reason for doubting this assumption (see, for example, UIC (2007), which does not suggest a picture of systematic under-renewal, with renewal costs generally rising over the period covered by the dataset)⁵. We also understand that during the summer of 2008 ORR has undertaken some further analysis based on the available data on relative renewal levels for some of the countries in the UIC's LICB dataset which supports the original analysis.

Furthermore, the use of the stochastic frontier approach itself (which allows for random noise effects), and the fact that we have analysed costs over an 11-year period, and not for just a single year, provide further safeguards against the risk of mis-interpreting low costs (due to a company being below steady-state) as evidence of efficient operation.

It should also be noted that the stochastic frontier approach gives greater weight to the leading firms in estimation. As a result, it is only if the leading firms are below steady-state that we would have serious cause for concern. Indeed, even if one of the leading firms was found to be below steady-state, we still have the benchmark of the remaining leading firms against which to compare Network

⁵ See UIC (2007), page 46.

Rail. Thus, we would expect the model to be reasonably robust even to changes in the costs of one of the leading firms.

We have also shown that the evidence put forward by LECG on steady-state uses quotes from a UIC report in a selective and unbalanced way. In one case, a quote is given that omits even the first half of the same sentence from which it is derived and thus totally changes the meaning of the quote - from suggesting a picture of falling renewals to rapidly increasing renewals. Ultimately, LECG does not put forward any clear evidence that the leading firms are below steady-state.

At this point it should be noted that, as stated in our June 2008 report, we (and ORR) recognise that ideally additional variables would be included in the cost function. We therefore accept and have always said that there is some uncertainty here, and that the distance from the frontier may reflect both inefficiency and the impact of omitted variables. However, as discussed above, we have taken advice from ORR in this regard and ORR, having looked at the evidence, has concluded that there is no reason to believe that incorporating additional variables would necessarily lead to a significant change in the model results and be favourable to Network Rail, since there will be factors which disadvantage Network Rail as well as benefiting it.

In addition, as noted, a “steady-state” adjustment has been made to Network Rail’s costs. We therefore consider that appropriate supporting work has been done in parallel to the econometric study to address the concerns raised. Furthermore, as discussed below, ORR has applied discount factors to the raw results of the econometric models to reduce the level of savings required during CP4 (by aiming off the frontier, and requiring two thirds of the gap to be delivered over CP4), and also combined the results with other evidence.

ORR’s use of the econometric work in its efficiency determination

Finally, as noted above, it is clearly shown in this report that there are good reasons for selecting the preferred model in its own right, based on the statistical tests applied and the model’s underlying assumptions. Indeed, the general consensus of evidence here is one of a substantial efficiency gap across all methods and the preferred model produces results in the middle of the range of models estimated (see Table 4 from the main body of the report, repeated below). As noted earlier, there are strong reasons for rejecting the BC92 model put forward by LECG.

In our view it is therefore appropriate for ORR to use the results of this model as the starting point for its efficiency determination. The model implies an efficiency gap against the frontier of 40%. Indeed, ORR uses the smaller gap of 37% measured against upper quartile. Since the computation of efficiency scores relative to upper quartile is normally only applied in the case of deterministic frontier approaches (in particular, corrected ordinary least squares, or COLS), which do not take account of random noise, the use of an efficiency gap measured against upper quartile in this case (where the preferred model uses the stochastic frontier method) reflects ORR’s aim to use a conservative estimate of

Network Rail's efficiency gap as its starting point (see ORR (2008)⁶). Indeed, we note that in its work for Postcomm, LECG does not make any adjustment to the efficiency scores coming out of its stochastic frontier models (see LECG (2005)).

Table 4

Model includes cost drivers for passenger and freight density	COLS	Random Effects GLS**	Random Effects MLE**	Random Effects MLE	Random Effects MLE	Random Effects MLE
	Time varying	Time invariant	Time invariant	Time varying	Time varying	Time varying
				LECG***	Cuesta00 (simple)	Cuesta00 (flexible)****
No steady-state adjustment						
Network Rail score 2006	0.56*	0.51	0.54	0.64	0.50	0.57
Network Rail rank in 2006	13	12	12	9	12	9
Rank correlation (2006 rankings)	0.91	0.86	0.75	0.38	0.97	1.00
With steady-state adjustment						
Network Rail score 2006	0.63*	0.58	0.65	0.70	0.54	0.60
Network Rail rank in 2006	12	11	7	7	12	10
Rank correlation (2006 rankings)	0.93	0.83	0.71	0.40	0.98	1.00

* The COLS score is shown against the upper quartile. All other scores are relative to the frontier

** For these models, the 2006 score is the same as for all other years (time invariant efficiency model).

*** This is the model put forward by LECG in their challenge to the econometric work

**** This is a more flexible version of Cuesta (2000) that allows for a possible turning point in efficiency during the post-Hatfield period for Network Rail

Shading represents preferred model

In our view, ORR's starting point for its efficiency determination is therefore a reasonable one, based on the econometric work carried out. The econometric results are also supported by the regional international econometric study (see ITS/ORR (2008)). From the starting point of a 37% efficiency gap, ORR then makes a further discounting assumption that two thirds of the gap can be closed over CP4. Furthermore, ORR has combined the results of the econometric work with other evidence in arriving at its draft efficiency determination.

We therefore consider that, in general terms, ORR has made appropriate use of the econometric work in its analysis, although ITS did not review the other evidence commissioned / produced by ORR, and was not involved in the details of the process by which ORR reached its draft efficiency determination. This process resulted in the efficiency gap from the preferred econometric model of 37% being scaled down to an efficiency target for maintenance and renewals in CP3 of 22% (see ORR (2008)⁷) and, of course, required ORR to exercise its regulatory judgement.

⁶ See for example page 115 and 116.

⁷ See page 141.

1. INTRODUCTION

During 2007 and 2008, The Institute for Transport Studies (ITS), University of Leeds, undertook econometric work to inform ORR's judgement on an appropriate efficiency target for the infrastructure manager over Control Period 4 (CP4). This work was conducted in conjunction with ORR and Network Rail, with input also from UIC, and culminated in a joint ITS / ORR report published in June 2008 (see ITS/ORR (2008)). During this period we have welcomed the involvement of and interaction with both institutions (in respect of Network Rail, mainly via Charles Robarts, David Rayner, Matthew Clements and David Smallbone; and Gerard Dalton and Teodor Gradinariu at UIC).

The work was peer reviewed by Dr Michael Pollitt from the Judge Business School, University of Cambridge. This work was used by ORR, alongside other evidence, in setting out its Periodic Review 2008 (PR2008) draft determination on efficiency published in June 2008.

Post June 2008, Network Rail commissioned LECG and Horton 4 Consulting to challenge the econometric work conducted by ITS/ORR (based on the UIC dataset) and ORR's judgement on Network Rail's efficiency more generally. The consultants' reports formed part of Network Rail's response to ORR's draft determination, which was submitted to ORR, and published, in early September 2008. This report sets out ITS's response to these consulting studies.

This report responds to the challenges raised by both sets of consultants. Our focus is on the econometric work, since ITS was not involved in the details of the process by which ORR reached its June 2008 draft determination of Network Rail's efficiency target for CP4. Nevertheless, at the time of the June 2008 determination we expressed our opinion that ORR's use of the econometric evidence was appropriate, in general terms, since it aimed off the frontier and then discounted the results further before arriving at an efficiency target; and also combined the results with other evidence. However, ITS did not review the other evidence commissioned / produced by ORR, and of course the process of arriving at its draft efficiency determination required ORR to exercise its regulatory judgement.

This report clearly demonstrates the robustness of the econometric study. This demonstration will be based on:

- the validity of the preferred econometric models in themselves both from a statistical and economic theory viewpoint;
- the vast array of supporting econometric evidence for the preferred models provided by the estimation of a wide range of alternative efficiency measurement methodologies; and
- the supporting econometric evidence provided by the regional international benchmarking study.

The challenges raised by Network Rail's consultants are therefore emphatically refuted. It should be noted that during our work some areas of analysis were outside the scope of ITS's remit (for example, collating evidence on the extent to which companies were above or below steady-state). In these areas we took

advice from ORR, and where this was done it is indicated in the relevant section of the report. In our response, we also comment, in general terms, on the way in which ORR used the econometric results in arriving at its draft efficiency determination.

The remainder of this report is structured as follows. In sections 2 and 3 we respond to the points raised by LECG and Horton 4 Consulting respectively. Section 4 concludes.

2. THE LECG REPORT

LECG challenge the econometric work done by ITS/ORR in the following areas:

- model documentation;
- data issues;
- steady-state assumptions;
- possible omitted variables;
- functional form; and
- specification of the preferred model.

In their report LECG argue that the international benchmarking analysis is based on just a single model, that has serious data, theoretical and statistical limitations, whilst stating that they had no access to information regarding the model selection procedure or any other models run that might support the preferred model. As will be clear below, our response emphatically refutes LECG's assertions on every point.

2.1 Model documentation

In the introduction to their report, LECG make a number of incorrect statements in respect of the model documentation. Firstly, they suggest that:

“there is no written documentation on (i) how the costs and cost drivers were calculated; ii) the rationale for the method of calculation adopted; iii) how the final cost drivers were selected; iv) how the functional form was chosen; and v) how the final econometric model was selected”; see LECG (2008), page 1.

However, the ITS/ORR June 2008 report provided information on all of these points. For example, the model formulation, including details of how the cost drivers were calculated, is shown on page 14 of the June report. Information on the dataset and the various adjustments made to it is provided on pages 11 and 12. Model selection and choice of functional form, including information on the testing of other methodologies and functional form is dealt with in a number of places (pages 14-16; 22; and 52).

Perhaps more importantly, the LECG report makes no reference to the fact that ITS and ORR worked jointly with Network Rail on this project between February 2007 and June 2008. As noted in the introduction, we have welcomed the involvement of and interaction with Network Rail in this work throughout the period. During this period we held numerous meetings where the results of our analysis were presented and explained, including the alternative models run, and these meetings provided an opportunity for Network Rail to challenge the analysis. Similar meetings were also held with Gerard Dalton and Teodor Gradinariu at UIC and again we found their input to be extremely useful in

developing our analysis (though we note that UIC have raised questions regarding our work and its application as part of PR2008, and of course Network Rail has not accepted the results).

Post June 2008, ITS and ORR held a conference call with Network Rail and LECG (in July 2008) during which the approaches were further explained and any additional information requested was provided by email within a reasonable timescale. We therefore question why LECG have made these incorrect statements concerning the documentation and provision of information in their report.

Later in the introduction to their report, LECG claim that:

“We have not, however, had sight of an explanation of the process that was used to select the ITS/ORR preferred model and reject the alternative model”; see LECG (2008), page 1.

This point is a recurring theme in the LECG report. However, in addition to the published document information on model selection and alternative methodologies, noted above, in our email and telephone correspondence we referred to the existence of numerous other models that had been run. LECG could at any time have asked for more detail and this information would have been provided. Indeed, we provided LECG with the data set, so they could have run a range of other models themselves (they report only one in their report, and this model is discussed further below in our response). Finally, Network Rail could have provided LECG with further information on the modelling process, as the results of other models had been shared with Network Rail at various points throughout the project.

2.2 Data issues

LECG’s report challenges the quality of the data used in the study in regard definitional differences between companies, staff turnover within companies leading to data being compiled differently in different years, and some anomalies in the data and concerns over the adjustments made.

With regard to definitions, as the consultants themselves recognise, the UIC publishes guidance to members on data definitions to aid harmonisation. Indeed, UIC uses the data employed in our study in its own (LICB) international benchmarking work. The dataset has been developed by UIC over a number of years now (starting in 1995), and forms the basis for its own (LICB) benchmarking methodology. UIC also produces guidance on data definitions to aid harmonisation. In its ten year report on benchmarking, UIC describes the development of its approach over the period of the analysis and notes that:

“Phase 5 [of the work] provided considerable insights into cost levels and mechanisms and gave useful advice to

Infrastructure Managers in Europe and overseas”, and “A “lasting benchmarking function” was established to guarantee a platform for continuous comparison of costs and the tracking of trends”; see UIC (2007), page 19.

Furthermore, LECG do not produce any examples of what the definitional differences might be. Their point about different staff being involved in data collection over time could and does apply to any panel data set used for analysis and therefore carries little weight. Finally, the only concrete examples that LECG are able to produce in respect of data quality relate to just two data points for one variable (number of switches) that is not included in the preferred model specification.

It should also be noted that ORR carried out detailed inspection work on the dataset prior to the analysis, as detailed in the June 2008 report (page 11). As noted in that report, the LICB dataset was taken largely as given, although checks were made for unusually large changes in data values from year to year. A judgement was made on whether these changes were justified by changes in other variables correlated to the trend examined and whether the trend appeared to be confirmed by other published sources or data collected. As a result of this analysis, a small number of data points were amended where the evidence strongly suggested that an input error had been made.

In addition, a small number of changes to the data where gaps existed in the dataset (e.g. where data for a variable for a particular company was missing for a single year), and the approaches to infilling are outlined in the June 2008 report. It was also noted that the impact of this data cleaning should be small given the approach adopted and the fact that only a small number of data points were amended. Whilst no dataset would be expected to be perfect, the data on costs, route length and characteristics (single track; electrification), and passenger and freight train-km – the variables used in our preferred model – appeared to be well behaved.

The LECG report also ignores the fact that ITS/ORR and Network Rail have had access to the UIC dataset since February 2007. The dataset was discussed early on in the project, and at no point has Network Rail expressed any serious concerns in this respect. It should further be noted that the UIC uses this dataset to inform its benchmarking approach. It is puzzling at this stage of the project for Network Rail’s consultants to argue that the dataset is such that it is unsuitable for analysis, when Network Rail has not made that point earlier. At any stage since February 2007 Network Rail could have conducted its own analysis of the data to check its reliability.

2.3 Steady-state assumptions

On pages 6 to 10 of their report LECG discuss the valid concern that at a point in time railways may be out of steady-state in respect of their expenditure, particularly renewals (where steady-state is defined as the level of expenditure that is broadly required to maintain the assets in a stable condition). They argue that the approach adopted by ITS/ORR for addressing this issue is inappropriate.

However, we disagree with LECG's description of the underlying assumptions implied by the approach taken. We also note that LECG's use quotations from a UIC benchmarking report in a selective and unbalanced way to back up their assertion that other railways are below steady-state.

2.3.1 Our underlying assumptions

In our June 2008 report, we discussed the fact that potential swings in railway expenditure from year to year (especially for renewals) could impact on our analysis. In this area we took advice from ORR, and the June report outlined the approach taken to deal with the problem, which was to make an adjustment to Network Rail's track and signalling expenditure (as it could be argued renewal activity in these areas is presently at above steady-state levels). The underlying assumptions and data required to make this adjustment was supplied to ITS by ORR⁸. Since there was insufficient data to make similar adjustments for other railways, this means that the approach assumes that the leading firms are broadly in steady-state (see ITS/ORR (2008), page 17).

During PR2008 ORR has expressed the view that it is not convinced that Network Rail is significantly out of steady-state by the end of the time period under analysis. Indeed, the BSL report prepared for Network Rail as part of the PR2008 process reached a similar conclusion, although their analysis appeared to be based on looking at the whole period, rather than the post-Hatfield years of the sample. As such, ORR considered the adjustment made to costs (that is, for the post-Hatfield years, to reduce Network Rail costs substantially prior to modelling) to be a conservative assumption. This adjustment means that, for example, in 2006, Network Rail's total cost is reduced by roughly 10% as compared to raw cost data. As noted above, in respect of other railways it is assumed that the leading firms are broadly in steady-state. In reaching its draft determinations, ORR advised us that there was no evidence to suggest this is not the case.

LECG state in a number of places in their report that the approach taken implies that:

"all [emphasis added] the other companies were in steady-state during the 1996-2006 period"; see LECG (2008)⁹.

However, in stochastic frontier analysis, it is the frontier (or leading firms) that define the benchmark against which Network Rail is judged, and the method ensures that greater weight is given to the data from the frontier firms in

⁸ In particular, Network Rail's renewals data was amended 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 increases the renewals cost data used for Network Rail in the years up to 2000 and reduces it thereafter.

⁹ For example, paragraphs 3.6, 3.7 and 3.8. Surprisingly, LECG nevertheless correctly import the actual quote from the ITS/ORR report, which refers to leading firms also in paragraph 3.6, but then refer to all firms elsewhere in the report.

determining the position and shape of the frontier¹⁰. As a result, if other firms (non-leading firms) are below steady-state, then adjusting their costs upwards should not be expected to alter Network Rail's efficiency score materially.

It is therefore only if the leading firms are below steady-state that we would have major reason for concern in terms of the parameter estimates and the position of the frontier against which Network Rail (after having its cost reduced by the steady-state adjustment) is judged. As noted above, ORR advised us that there was no evidence to suggest that the leading firms were significantly out of steady-state. Furthermore, for the preferred model, the frontier is defined by three firms over the period of the sample, and thus even if one of these was out of steady-state, we still have the benchmark of the other two firms against which to judge Network Rail's efficiency performance. Thus, the model should be reasonably robust to changes to even one of the leading firms¹¹.

However, this argument does not necessarily hold the other way round. Therefore, whilst the model should be fairly robust to one of the leading firms being below steady-state, since the other leading firms continue to define the frontier, if just one of the leading firms was found to be above steady-state cost, then the resulting downward cost adjustment would be expected to shift the frontier outwards, thus increasing the relative inefficiency of Network Rail and the other firms in the sample.

At this point it is worth noting also, 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 model should not interpret unusually low costs in a given year (perhaps due to budget constraints) 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 which should further guard against the risk of counting low investment in a single year or over part of the period as a sign of efficient operation.

2.3.2 Evidence put forward by LECG on steady-state

Consideration of the evidence regarding the steady-state was not in ITS's remit. Nevertheless, we make some comments below on the way LECG has interpreted the evidence in this regard.

In their report, LECG set out a number of quotes - from a UIC publication¹² - that are intended to show numerous railways to be below steady state. However, one of the quotes included omits the first part of the same sentence from which it is derived, and thus totally changes the meaning of the original report.

¹⁰When estimated by maximum likelihood. See Lovell (1993), page 22.

¹¹ Of course, although the frontier may comprise more than one firm, an adjustment to the costs of one of the frontier firms in a relatively small sample could impact on the parameter estimates and thus affect the efficiency scores of firms in the sample to some extent.

¹² UIC (2007).

LECG's quote is as follows:

"... in recent years budget problems of the government forced us to cut renewal expenditures"; see LECG (2008), page 8.

The full quote, including the first part of the sentence is as follows:

"Following our strategy renewal expenditures rose very strongly over the last decade; in recent years budget problems of the government forced us to cut renewal expenditures"; see UIC (2007), page 79.

Omitting the first part of the sentence totally changes the meaning from painting a picture of underinvestment, to one of very strong growth in renewal activity which fell off to some extent in the last couple of years. Indeed, the data shown on the same page in UIC report¹³ shows renewals for this company to be rising strongly over the period and to be above the average in absolute terms (and in line with the average on a per route-km basis).

In addition, there are numerous other quotes in the report that indicate that renewals are increasing or that there is no problem in respect of steady-state – see, for example, Table 1 below. It is further worth noting that the UIC (2007) report shows a general picture among the participant railways of rising renewal costs per track-km over the period of our analysis (see UIC (2007), page 48¹⁴). Though this is a trend, rather than indicating absolutely high levels of renewals, it does not suggest a situation in which railways are cutting back on renewal expenditure.

¹³ UIC (2007), page 79.

¹⁴ Furthermore, it is not correct to say that this is entirely due to Network Rail, as suggested by Horton 4 Consulting (2008), page 10.

Table 1

“Renewal and maintenance expenditures are of the same order of magnitude and appear to be a good balance leading to optimised cost” (UIC (2007), page 103.

“On the one hand we managed to cut maintenance costs by introducing new procedures, technologies etc. On the other hand, the renewal expenditures were raised” (UIC (2007), page 78.

“In 1998 expenditures for maintenance increased, one reason was the new organisation (client/contractor) leading to focusing on the fact that the share of corrective maintenance was too high. More funds were then spent on preventative maintenance in order to decrease the level of corrective maintenance” (UIC (2007), page 73.

2.3.3 Summary of this section

To summarise on this section, our approach for dealing with the possibility that Network Rail and perhaps other railways are out of steady-state consists of making a substantial downwards adjustment to Network Rail’s costs during the post-Hatfield period (e.g. by roughly 10% in 2006)¹⁵, whilst assuming that the leading firms are broadly in steady-state. The stochastic frontier approach itself, and the analysis of costs over an 11-year period, provide further safeguards against the risk of mis-interpreting low costs (due to a company being below steady-state) as evidence of efficient operation. ORR has also looked at the evidence and has concluded that there is no reason for doubting the second assumption, whilst ORR has expressed the view that the downward adjustment to Network Rail’s costs is conservative.

In terms of responding to LECG’s points, we have explained why the approach assumes that it is the leading firms, and not all firms, which are in steady-state. Furthermore, we have demonstrated that the evidence put forward by LECG sources information from a UIC report in a selective and unbalanced way, and ultimately does not put forward clear evidence that the leading firms are below steady-state.

2.4 Possible omitted variables

The LECG report suggests three categories of omitted variables:

- input prices (for example, unit labour costs and materials prices);
- variables relating to allocation of funds (steady-state); and
- intensity of rail usage.

¹⁵ Whilst increasing costs during the Railtrack period prior to Hatfield.

With respect to input prices, as LECG note, all of the cost data is converted to a common currency by means of Purchasing Power Parity (PPP) exchange rates. This is a common approach. As LECG note, this allows us to take account of general (economy wide) differences in price (and wage) levels between countries, but not necessarily differences across countries in wage rates in the rail infrastructure sector (unless they mirror the general differences in wage rates between countries).

We accept that ideally the cost function would include input prices. However, input prices are captured in the ORR's overall determination of Network Rail's efficiency target in two ways. Firstly, as noted, in the econometric work, we have adjusted for differences between countries in general wages, and this adjustment will only be inadequate to the extent that the ratio of rail wage rates between different countries differs from that between general wage rates.

Secondly, in making its draft determination on Network Rail's target, ORR took account of differences in input price trends in rail compared to the economy as a whole in Britain via a downward adjustment to the company's efficiency target. Of course, this adjustment focuses on future trends in input prices, rather than differences in levels, so does not wholly deal with the problem. However, ORR advised us that there was insufficient evidence to judge whether Network Rail faced higher relative wage rates than in other countries (or at least that if differences did occur, wages might be thought of as partially endogenously determined and thus under Network Rail's control). We understand that over the summer of 2008 ORR carried out further work on this question using available data which supports the original analysis.

With respect to variables relating to annual allocations of funds (steady-state), the previous section outlines our approach to addressing this issue, so it is puzzling that LECG comment that:

"..ITS/ORR does not appear to have considered these factors"; see LECG (2008), page 10.

Our response to LECG's challenge to the approach in respect of steady-state is set out in section 2.3 above. Finally, LECG refer to the intensity of rail usage as an important variable. However, they do not define this variable, or explain how it might be expected to impact on the comparison. We note that our model already includes both passenger and freight density volume measures.

In concluding our response on this section, as we noted in our June report, we recognise that ideally additional variables would be included in the cost function¹⁶. We therefore accept that there is some uncertainty here, and that the distance from the frontier may reflect both inefficiency and the impact of omitted variables. It should be noted that this issue was raised by Network Rail during the model development process.

¹⁶ We would hope that obtaining data on additional cost drivers would be a priority direction for future research.

Further work in this area was outside the scope of ITS's remit. However, ORR (and Network Rail, via the BSL study) sought to mitigate this uncertainty by conducting / commissioning parallel studies to understand the likely impact of omitted variables in respect of Network Rail. Ultimately, ORR concluded that there was no reason to believe that incorporating such variables would necessarily lead to a significant change in the model results and be favourable to Network Rail, since there will be factors which disadvantage Network Rail as well as benefiting it.

It should also be noted that LECG are essentially arguing that unless one can include all possible variables and develop a "perfect" model, then efficiency analysis should not be conducted. This conclusion is not consistent with regulatory practice or the wider academic literature. In common with other regulators, it should also be noted that ORR has made conservative adjustments to the raw results coming out of the econometric models in arriving at its judgement on an appropriate efficiency target for Network Rail (by aiming off the frontier, and requiring that only two thirds of the gap is closed over CP4), and has also based its determination on a range of evidence from other studies, so as not to rely solely on the results of a single, econometric study (see section 2.6 below). We therefore consider that it is wholly appropriate to use the econometric models developed in this context.

2.5 Functional form

In their report LECG note the following:

"A great virtue of the Cobb-Douglas functional form is that its simplicity enables us to focus our attention where it belongs, on the error term, which contains information on the cost of inefficiency. As an empirical matter, however, the simplicity of the Cobb-Douglas functional form creates two problems. As Hasenkamp (1976) noted long ago, in a commentary on Klein's (1947) famous railroad study, a function (or frontier) having the Cobb-Douglas form cannot accommodate multiple outputs without violating the requisite curvature properties in output space"; see LECG (2008), page 11, taken from Kumbhakar and Lovell (2000), page 143.

LECG then question why ITS/ORR "have selected a functional form that is incompatible with economic theory", without giving any further explanation as to what they perceive the problem to be.

In response we make three points. First, LECG have taken this quote out of the context of the wider theoretical and empirical literature, and indeed even the book from which the quote is taken. The multiple output Cobb-Douglas cost function does not violate any required theoretical property of a cost function in a regulated

industry, as further reading of the earlier sections of the same textbook from which LECG derive the above quote demonstrates¹⁷.

The problem applies only for profit maximising firms in a purely competitive industry (where firms can choose output levels to maximise profits), since the conditions for profit maximisation are violated. It does not apply in a regulated industry such as railways, where railway output levels are typically assumed to be exogenously determined. This point is clearly made in Klein (1953):

“This same problem does not arise in a model of a regulated industry....”; see Klein (1953), page 227.

Coelli and Perelman (2000) also quote Klein (1953) and make a similar point¹⁸. The distinction between the perfectly competitive and regulated case is also made by Nerlove (1965) – this source being referenced in the Hasenkamp (1976) paper referred to in the paragraph quoted by LECG above. Nerlove (1965) dedicates a whole chapter to discussing Klein’s work, and notes the following:

*“The study is especially interesting for the techniques employed; and while these are primarily applicable only in the case of a **regulated industry** [emphasis added], several lessons may be drawn of more general interest”; see Nerlove (1965), page 61.*

Secondly, the error in LECG’s interpretation of the quote in their report is made clear by the fact that multiple output cost functions have been estimated in the literature in numerous cases – either in their own right, or as part of a statistical testing approach alongside other functional forms¹⁹. Some key regulatory studies have also adopted this approach, for example the recent and highly regarded study commissioned by the German Network Agency in respect of gas and electricity distribution benchmarking in Germany (see Sumicsid (2007))²⁰.

Indeed, a well acknowledged practical advantage of the Cobb-Douglas function, over the translog form, is its parsimony²¹. Thus the functional form for the cost function is in fact an issue for econometric testing, rather than being a required theoretical property. As noted in our June report we have tested the Cobb-Douglas model and found it to be preferred to the other alternatives (e.g. linear and translog).

¹⁷ See Kumbhakar and Lovell (2000), pages 20 and 34. See also Coelli, Rao, O’Donnell and Battese (2005), page 23.

¹⁸ Coelli and Perelman (2000), page 1969.

¹⁹ The following are just a few examples: Farsi and Filippini (2006), Barros (2005), Mulatu and Crafts (2005) and Eeckaut, Tulkens and Jamar (1993).

²⁰ For example, pages 35-38.

²¹ Coelli, Rao, O’Donnell and Battese (2005), page 212.

Finally, LECG themselves estimate a multiple output cost function in their work on postal cost efficiency (see LECG (2005)). For example, in their analysis of postal delivery office costs they estimate a Cobb-Douglas cost function:

“Another common functional form is the Cobb-Douglas form... We tested a number of alternative functional forms and found that the Cobb-Douglas form provided the best empirical fit to the data”; see LECG (2005), page 351.

and list the following variables as “measures of scale and output”:

“...measures of scale and output include: number of delivery points; percentage of delivery points that are business; and weighted and disaggregated volumes”; see LECG (2005), page 342.

In their 2005 study, LECG also list more than a dozen additional variables that may be thought of as output or output characteristic variables (for example, volume of mail redirected and length of road per delivery point)²² for possible inclusion in the model. They also note the OFWAT use the Cobb-Douglas form in their analysis²³. We therefore question why LECG consider our approach to be incompatible with economic theory, as it would imply that the approach they have adopted elsewhere suffers from the same problem.

2.6 Specification of the preferred model

LECG refers to a technical problem with our preferred model and the supposed fix that we have employed to make the model work. In this section we first respond to this point and show that the “fix” is in fact an accepted and very widely used approach to estimating the variance co-variance matrix for stochastic frontier models. We go on to argue why we think the model is robust (on its own terms), as well as against the background of the vast array of other methods (including the model put forward by LECG) that we have applied to this dataset.

²² See pages 343-344. The full list of variables included in the delivery office Cobb-Douglas cost function model are as follows: volume per delivery point, number of delivery points, length of road per delivery point, percentage of business delivery points, mail re-direction, average wage rate, wage competitiveness index, major city centre dummy, urban dummy, sub-urban dummy, rural dummy, number of RM2000 frames – see page 352. Of course, the precise definition of what is an output, as opposed to being an output characteristics is often open to debate, but it is clear that the variables listed above are similar in nature to the inclusion of route length and passenger and freight density in our preferred model.

²³ See page LECG (2005), page 351. It is clear that some of the OFWAT models include more than one output variable.

2.6.1 LECG comments on the preferred model

In their report, LECG state that:

“The ITS/ORR preferred model fails to produce an estimate unless a particular change is made to the solution algorithm normally used by the LIMDEP 9 software”; see LECG (2008), page 12.

This is not correct. The so called “fix” that we have used is an augment option to the standard stochastic frontier command which was provided to ITS by Econometric Software, the company that licenses the LIMDEP 9 software (in relation to an efficiency analysis conducted on a different data set for an unrelated project during 2007; and which is explained further below). When the model is run using the standard option (so without the augment option (as LECG has done)), LIMDEP gives the message “normal exit from iterations”, which indicates that the model has estimated. It is then a matter of requesting LIMDEP to produce the output from that estimation²⁴. As will be seen from Table 2 below, the parameter estimates are identical to those obtained when using the augment option²⁵. It is clear from Table 2 that these models are identical; and the software output contains all the information required to produce efficiency estimates.

Table 2

	Parameter estimates	
	LIMDEP software with augment option*	LIMDEP software standard option
ONE	6.245	6.245
ROUTE	1.074	1.074
PASSDR	0.335	0.335
FRDR	0.179	0.179
SING	-0.918	-0.918
ELEC	-0.037	-0.037
TIME	0.056	0.056
TIME2	-0.005	-0.005
Log likelihood	86.230	86.230

* Provided to ITS by Econometric Software (the company that licenses the LIMDEP software)

The issue that LECG raises concerns not the estimation of the model and its parameters and efficiency scores, but rather the estimation of the variance co-

²⁴ This involves the addition of the standard “OUTPUT=3” term to the LIMDEP code. The resulting output also confirms that the model has converged.

²⁵ The parameters relating to efficiency change for each firm are also the same but are omitted for brevity.

variance matrix. In this respect, LECG are correct in saying that without the augment option in LIMDEP, the software judges that the variance co-variance matrix is singular, so that standard errors cannot be computed.

However, we do not agree that our use of the augment option can be regarded as a “fix”. The augment option enables LIMDEP to estimate the variance co-variance matrix in a way that is accepted and widely used; and produces low standard errors, and thus statistically significant parameter estimates. For example, this is the method employed by the efficiency estimation programme, FRONTIER 4.1. The FRONTIER software was developed by Professor Tim Coelli and has been very widely used in the academic literature²⁶. A further comparison with another application (using code developed in the Matlab programming language by Professor Robin Sickles) produces almost identical results to those resulting from both the software using the augment option and the FRONTIER software (for a simpler time varying efficiency model)²⁷.

2.6.2 Justification for our preferred model

It should be noted that both approaches to estimating the variance co-variance matrix (that is, with and without the augment option) are accepted approaches. However, in the present context, one method implies that the parameters of the model are not estimated precisely, whilst the other suggests (with the augment option) that we can obtain quite precise estimates. Further comment is therefore warranted.

We are confident in the results from our preferred model for the following reasons. First of all, as noted above, the approach we have used to estimate the variance co-variance matrix has been widely adopted in the academic literature. Secondly, where there is uncertainty concerning the reliability of the variance co-variance matrix it is possible to use alternative testing procedures to determine whether certain parameters are estimated precisely. In particular, the likelihood ratio (LR) test is computed without relying on the standard errors coming out of the variance co-variance matrix, so is not affected by the potential uncertainty concerning the reliability of this matrix. Indeed, stochastic frontier researchers often prefer the LR test for this reason²⁸. We have computed likelihood ratio tests for each of the parameters in the model and the results are summarised in Table 3 below.

²⁶ See Coelli (1996).

²⁷ Professor Robin Sickles makes available code in the Matlab language for the purpose of efficiency analysis, which enables us to estimate a Battese and Coelli (1992) model. This model produces identical parameter estimates and almost identical standard errors to those obtained from LIMDEP (with the augment option) and FRONTIER. This finding suggests that this code is adopting the same method that we have used in our study.

²⁸ See Coelli, Rao, O'Donnell and Battese (2005), page 258.

Table 3

	LR test statistic (note 1)	Z-statistic based on standard errors (note 2)
Variable		
ROUTE	NA (note 3)	38.7253 ***
PASSDR	5.44 **	4.6103 ***
FRDR	6.58 **	2.85349 ***
SING	21.88 ***	-10.2273 ***
ELEC	2.3	-0.464107
TIME	13.84 ***	3.91652 ***
TIME2	16.14 ***	-4.18794 ***

* Significant at the 10% level. ** Significant at the 5% level

*** Significant at the 1% level

Note 1: this test statistic is distributed $\chi^2(1)$

Note 2: this test statistic is distributed as a standard normal N (0,1)

Note 3: the model produces an error when the key scale variable, route length, is excluded from the model

From this table it is clear that all of the variables that are deemed to be statistically significant based on the z-statistics (computed from the standard errors contained in the variance co-variance matrix) remain so at the 5% or 1% level when the LR test is used. An LR test also confirms that the efficiency effects are statistically significant at the 1% level. This additional testing gives us added confidence in the findings concerning the significance of the variables derived from the z-statistics and suggests that the approach we have used has produced a reasonable estimate of the variance co-variance matrix.

Thirdly, the point parameter estimates appear to be plausible in terms of the signs of the coefficients and their magnitude. The only exception to this is the electrification variable, which takes an unexpected negative sign, although it is close to zero and is statistically insignificant. It may be possible that the third rail electrification system in Britain – which will have a different impact on costs – could be impacting on the coefficient in respect of this variable, although the British third rail network will be a comparatively small share of total European electrified track. In any case, as described further in section 2.6.5 below, dropping the electrification variable either has little impact on Network Rail's score, or reduces it (implying greater relative inefficiency) depending on the model used²⁹.

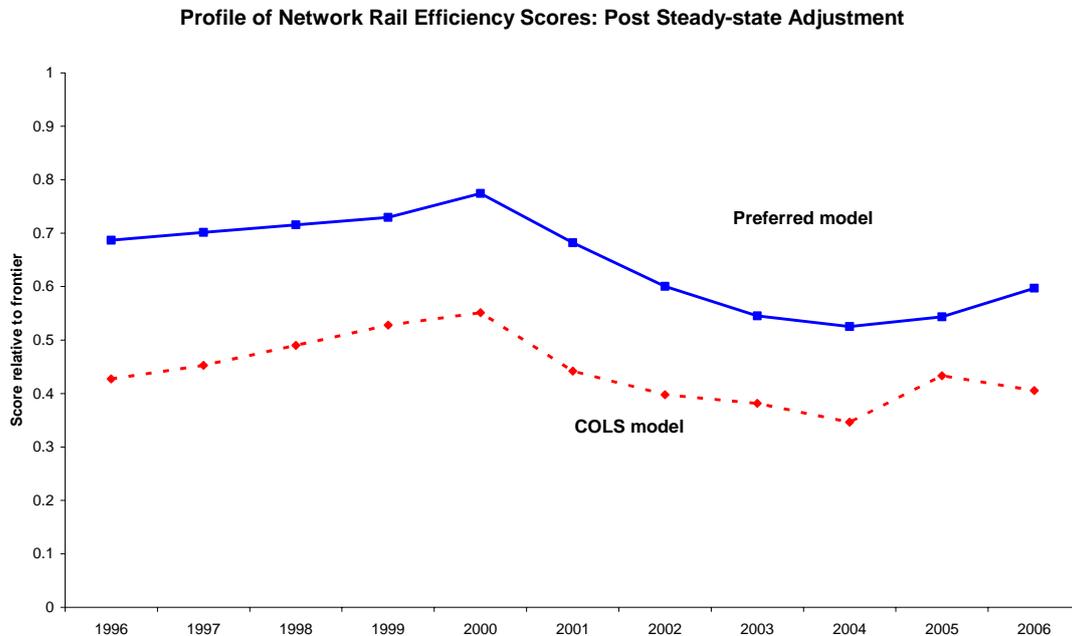
Fourthly, the results produce a plausible pattern of efficiency change over time for Railtrack / Network Rail, as illustrated by Figure 1 below. Figure 1 shows efficiency improving modestly in the early years after privatisation (even after making an adjustment for possible under-renewal during that period³⁰), before deteriorating sharply during the post-Hatfield period, before starting to improve once the efficiency savings being delivered by Network Rail during CP3 start to

²⁹ Therefore we prefer to retain this variable in the model for theoretical reasons. That is, we expect this variable to impact on costs – and the variable is statistically significant in the maintenance cost regression. It is possible that its effect in the M&R model is being obscured by its correlation with some of the other explanatory variables.

³⁰ As noted earlier, the steady-state adjustment increases costs during the early years and reduces them in later years.

have an impact. The 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). At this point it should be noted that the alternative time varying model put forward by LECG does not produce a believable time path for efficiency (see section 2.6.4 below).

Figure 1



Finally, we note that the preferred model produces similar conclusions in respect of efficiency to the results from other approaches to efficiency estimation (described in section 2.6.3. below). It also produces similar conclusions in respect of the efficiency gap and potential for efficiency savings to the other studies that ORR has drawn on in making its draft determination. Section 2.6.3 provides a summary of the basis of our model selection process (a full description of the model selection process is provided in a separate document).

2.6.3 A summary of the model selection process

The key models that informed our choice of preferred model – and gave us added confidence in that model - are shown in Table 4 (a full description of the model selection process is provided in a separate document; see Smith (2008)). Though the alternative time varying model (attributed to Battese and Coelli (1992); hereafter BC92) suggested by LECG was rejected for various reasons (see section 2.6.4 below), it is shown for comparative purposes given that LECG have raised it in their report. Note that in their report, LECG incorrectly reference this model as Battese and Coelli (1995), which is a totally different model.

In arriving at our preferred model we considered the underlying model assumptions and carried out appropriate statistical testing. With regard to the underlying assumptions, the preferred model allows efficiency to vary over time in a flexible but structured way that recognises the structure of the data (namely that

the dataset consists of thirteen³¹ firms over 11 years). Furthermore, it allows efficiency to change in a different direction and to different extents for each firm, so that efficiency can be improving for one firm, and deteriorating for another. Finally, it allows for a possible turning point in inefficiency during the post-Hatfield period, to allow for the model to pick up the improvements in efficiency achieved by Network Rail during CP3.

Furthermore, ORR carried out or commissioned other work to verify the results of our preferred model as noted elsewhere in the report. In addition, we compared the efficiency scores against a range of other frontier approaches (see Table 4) and on this basis judged that the preferred model was in line with the results of other approaches.

Table 4 shows Network Rail's efficiency scores for 2006. The stochastic frontier model scores are shown relative to the frontier as is usual, whilst the COLS scores are shown relative to upper quartile. The latter adjustment is commonly performed by economic regulators, and is a prudent adjustment to put the results on what is probably a more comparable basis to the results of stochastic frontier models (since the COLS model does not distinguish inefficiency from noise). Network Rail's 2006 ranking is also shown, together with the correlation coefficients between the scores for each model against the preferred model (which is shown, with shading, in the final column).

Since ORR expressed the view that it is not convinced that Network Rail's activity is significantly above steady-state by the end of the time period under analysis, the scores before and after the steady-state adjustment are shown. The models in Table 4 include both passenger and freight density as cost drivers since we expect these to impact on costs differently (see also the discussion on sensitivity analysis in section 2.6.5). It should also be noted that, whilst two of the models are described as time invariant models, there is a time variant element for Network Rail since the British data has been split to create two firms (relating to the pre- and post-Hatfield period)³².

The efficiency scores in Table 4 puts Network Rail's efficiency score in the range 0.50 to 0.70, implying an efficiency gap of 30-50% (calculated as one minus the efficiency score), with the average gap being 41%. As explained further below, we have strong reasons for rejecting the BC92 model suggested by LECG. Taking BC92 out of the comparison, puts the gap in the range 35-50%, with an average gap of 43%. If we focus just on the models that include a steady-state adjustment, then the average gap is 38% (including BC92) and 40% (excluding BC92).

³¹ Or 14 firms if the British data is treated as two firms (before and after Hatfield).

³² See our more detailed model selection report for further details.

Table 4: Scores based on models with passenger and freight train-km densities as separate cost drivers (all scores, except COLS, are relative to the frontier)

Model includes cost drivers for passenger and freight density	COLS	Random Effects GLS**	Random Effects MLE**	Random Effects MLE	Random Effects MLE	Random Effects MLE
	Time varying	Time invariant	Time invariant	Time varying	Time varying	Time varying
				LECG***	Cuesta00 (simple)	Cuesta00 (flexible)****
No steady-state adjustment						
Network Rail score 2006	0.56*	0.51	0.54	0.64	0.50	0.57
Network Rail rank in 2006	13	12	12	9	12	9
Rank correlation (2006 rankings)	0.91	0.86	0.75	0.38	0.97	1.00
With steady-state adjustment						
Network Rail score 2006	0.63*	0.58	0.65	0.70	0.54	0.60
Network Rail rank in 2006	12	11	7	7	12	10
Rank correlation (2006 rankings)	0.93	0.83	0.71	0.40	0.98	1.00

* The COLS score is shown against the upper quartile. All other scores are relative to the frontier

** For these models, the 2006 score is the same as for all other years (time invariant efficiency model).

*** This is the model put forward by LECG in their challenge to the econometric work

**** This is a more flexible version of Cuesta (2000) that allows for a possible turning point in efficiency during the post-Hatfield period for Network Rail

Shading represents preferred model

As noted above, there are good reasons for selecting the preferred model in its own right, based on the statistical tests applied and the model's underlying assumptions. Indeed, some of the other models in Table 4 are nested within the preferred model and the implied restrictions in the alternative models can be rejected (this applies to the other three random effects MLE models shown). However, even if we look at the results of the preferred model in the context of a wide range of alternative efficiency estimation methods, with an efficiency gap against the frontier of 40%, it is clear that the preferred model produces an efficiency gap for Network Rail more or less exactly in line with the average of the models which include a steady-state adjustment (and of course those that do not). Furthermore, this conclusion holds even if we retain the BC92 model as one of the comparator models.

Given the above results, we consider that it was appropriate for ORR to use the preferred model (including a steady-state adjustment) from Table 4 above as the starting point for its efficiency determination. As discussed further in section 2.6.6, if we look at the efficiency gap against upper quartile for the preferred model, the efficiency gap comes down to 37% and this is the figure that ORR uses as the starting point for its efficiency determination. This is a conservative assumption, as discussed further below.

It is worth noting here that LECG could at any point have asked us for the supporting information contained in this section, rather than simply commenting that they had not had any sight of the results of other models run or the model selection procedure. It should also be noted that ITS and ORR held numerous meetings with Network Rail during the course of the work showing results from

various models, so LECG's comments regarding lack of information on the model selection process are therefore somewhat puzzling.

2.6.4 The BC92 model put forward by LECG

As already noted, we find that the preferred model produces results in line with (and roughly equal to the average of) other models, even if the BC92 model is included as one of the comparators. However, since this model does produce a higher efficiency score than the preferred model, and since LECG put it forward as an alternative in their report, we briefly explain why we have good reasons to select our preferred model over the BC92 model, and indeed to have strong doubts about the BC92 results more generally.

Firstly, the BC92 approach is restrictive in that it forces the same direction of efficiency change for all firms over time. This could be a serious limitation if one company has a very different direction of efficiency change over the period (which could be the case for Network Rail). Indeed, the BC92 model is nested within our preferred model - which does allow efficiency to vary in different directions for different firms – and the restricted (BC92) model can be very clearly rejected based on appropriate statistical testing³³. Further statistical testing shows that, starting from the BC92 model, it is not possible to reject the null hypothesis of time invariant inefficiency, so that the BC92 model cannot be considered a robust model of time varying inefficiency³⁴.

As further evidence that the BC92 model may be inappropriate we note that it produces the unexpected result that Network Rail's efficiency increases every year over the post-Hatfield years (2000 to 2006), which seems unrealistic given the very sharp cost rises over that period (Figure 2). There seems to be little doubt that is was during the post-Hatfield period – that is, during the period of Railtrack's administration, and the early years after the formation of Network Rail - that efficiency got a lot worse.

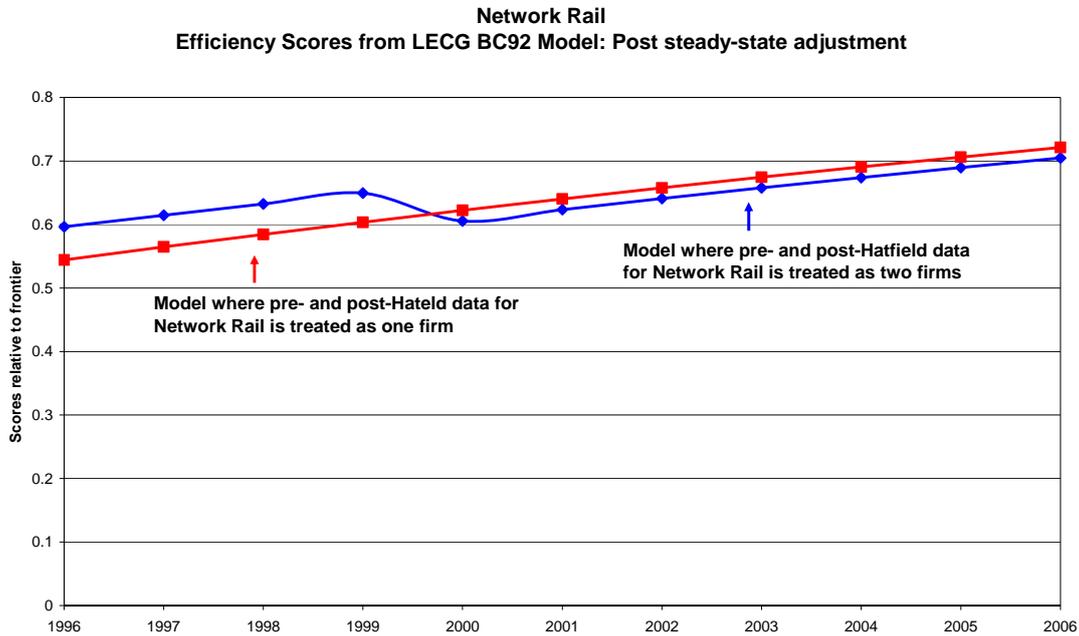
Indeed all other firms in the sample see increasing efficiency over the whole period as well (the model assumption forces the same direction of change for all firms). In addition, the key traffic density parameters are not statistically significant, and the BC92 model produces a large negative point estimate for the coefficient on the electrification variable. The latter point is important since Network Rail has previously expressed concern over the inclusion of this variable in the model³⁵. The BC92 model is also rather sensitive to small changes in the model specification³⁶.

³³ An LR test rejects the restricted (BC92) model at the 1% level, which is an indication of a very clear rejection of the BC92 model.

³⁴ In contrast, our preferred time varying efficiency model is clearly preferred over the time invariant alternative (statistically).

³⁵ Whilst we do not regard a negative coefficient as necessarily being a problem, the electrification variable takes a larger negative value than in our preferred model (although it remains statistically insignificant). Since Network Rail have made a point in their Strategic Business Plan Update (see Network Rail (2008)) of noting the "problem" with including an electrification variable, it is surprising that LECG have put forward an alternative model that suffers from the "problem" to a much greater extent.

Figure 2



It can be clearly seen then, that the weight of evidence from a wide range of models supports the results of the preferred model and that there are strong reasons for selecting the preferred model over the BC92 alternative suggested by LECG. The BC92 model clearly stands out as an outlier. The correlation coefficients between rankings of all firms across models indicate that the BC92 in general produces different rankings for all firms (not just Network Rail), whereas the correlation between the scores and rankings of the preferred model and the other models is reasonably high.

The model also produces some rather surprising results in terms of the efficiency scores for Network Rail over time, and can also be rejected based on numerous statistical tests. LECG have put the BC92 model forward without reporting the results of any statistical testing of its properties, or commenting on the reasonableness or otherwise of the results in themselves and as compared to the preferred or other relevant models.

2.6.5 Sensitivity analysis

To complete our discussion of model selection, we note that we also ran a number of sensitivities to test the robustness of our preferred model. The results

As discussed below, we have carried out sensitivity analysis in respect of excluding this variable from our models.

³⁶ The reported efficiency scores for Network Rail change markedly, for example, when the passenger and freight density measures are replaced with a total density measure or the electrification variable is dropped from the model.

from running the various models and sensitivities discussed in this section demonstrate that our modelling and model selection procedure is robust. First of all, we re-ran all of the models shown in Table 4, replacing the passenger and freight density variables with a measure of total train density. This sensitivity was run in response to a request from Network Rail raised during the various meetings held between ORR, Network Rail and ITS to discuss the modelling process. The results are shown in Table 5 below.

Table 5: Scores based on models with total train-km density as the traffic cost driver (all scores, except COLS, are relative to the frontier)

Model includes single traffic related cost driver (total train density)	COLS	Random Effects GLS**	Random Effects MLE**	Random Effects MLE BC92	Random Effects MLE Cuesta00 (simple)	Random Effects MLE Cuesta00 (flexible)***
No steady-state adjustment						
Network Rail score 2006	0.56 *	0.51	0.55	0.46	0.52	0.59
Network Rail rank in 2006	12	12	11	12	12	9
Rank correlation (2006 rankings)	0.73	0.89	0.82	0.86	0.97	1.00
With steady-state adjustment						
Network Rail score 2006	0.640 *	0.59	0.65	0.75	0.61	0.67
Network Rail rank in 2006	12	10	8	7	9	7
Rank correlation (2006 rankings)	0.68	0.88	0.82	0.69	0.98	1.00

* The COLS score is shown against the upper quartile. All other scores are relative to the frontier

** For these models, the 2006 score is the same as for all other years (time invariant efficiency model).

*** This is a more flexible version of Cuesta (2000) that allows for a possible turning point in efficiency during the post-Hatfield period for Network Rail

The results in Table 5 produce slightly higher efficiency scores (and therefore imply a slightly lower efficiency gap) than those in Table 4, putting the efficiency gap in the range 25% to 54%, with an average gap of 41%. Excluding the BC92 models, which we have reason to doubt as outlined above³⁷, the efficiency gap ranges from 33% to 49% (average of 41%), or 33% to 41% (average of 37%) if only the most favourable models (i.e. those that include a steady-state adjustment) are included. The preferred model, with an efficiency gap of 40% against the frontier, is again broadly in the middle of the range of estimates in Table 5.

At the time of the June 2008 draft determination, we judged that there were reasons for selecting the preferred model shown in Table 4 as compared with the alternatives shown in Table 5. Whilst the “total train density” model is more parsimonious, this comes at a cost, since it does not distinguish between passenger and freight traffic in terms of their impact on costs, whereas our

³⁷ The comments above apply also to the total train density version of the BC92 model. Indeed, we note here that the BC92 model defines the extreme upper and lower bounds of the efficiency gap, depending on the cost variable used, which is a further indication of its instability.

preferred model does. Standard model selection procedures (based on the Log-Likelihood values and the AIC and BIC criteria³⁸) also point towards the preferred model. Overall, we judged that all models, however specified, were indicating a substantial efficiency gap, and that the preferred model was in the middle of the range of the models selected. We therefore consider that our model selection process is robust.

In addition to the models shown, we also ran a number of other sensitivities, including dropping the electrification variable (as noted earlier, Network Rail had some concerns over the inclusion of this variable), as well as numerous other model specifications, for example, including switch and stations density (though noting the possible data quality issues with these variables). In all cases the resulting efficiency scores were very close to those of the preferred model, and in many cases were substantially lower, thus implying larger efficiency gaps (and the models tended to perform less well in terms of the size and significance of the parameter estimates).

2.6.6 ORR's use of the econometric work in PR2008

It can be seen from the discussion above that there are good reasons for selecting the preferred model in its own right, based on the statistical tests applied and the model's underlying assumptions. Indeed, the general consensus of evidence here is one of a substantial efficiency gap across all methods and the preferred model produces results in the middle of the range of models estimated. In our view it is therefore appropriate for ORR to use the results of this model as the starting point for its efficiency determination.

Furthermore, since the computation of efficiency scores relative to upper quartile is normally only applied in the case of deterministic frontier approaches (in particular COLS), which do not take account of random noise, the use of an efficiency gap measured against upper quartile in this case (for a stochastic frontier model) reflects ORR's aim to use a conservative estimate of Network Rail's efficiency gap as its starting point (see ORR (2008)³⁹).

In our view, ORR's starting point for its efficiency determination is therefore a reasonable one, based on the econometric work carried out. The econometric results are also supported by the regional international econometric study (see ORR/ITS (2008)). From the starting point of a 37% efficiency gap, ORR then makes a further discounting assumption that two thirds of the gap can be closed over CP4. Furthermore, ORR has combined the results of the econometric work with other evidence in arriving at its draft efficiency determination.

We therefore consider that, in general terms, ORR has made appropriate use of the econometric work in its analysis, although ITS did not review the other evidence commissioned / produced by ORR, and was not involved in the details of the process by which ORR reached its draft efficiency determination. This

³⁸ Akaike Information Criteria (AIC) and Bayesian (or Schwartz) Information Criteria (BIC); see Greene (2003), pages 159-160.

³⁹ See, for example, page 115-116.

process resulted in the efficiency gap from the preferred econometric model of 37% being scaled down to an efficiency target for maintenance and renewals in CP3 of 22% (see ORR (2008)⁴⁰) and, of course, required ORR to exercise its regulatory judgement.

2.7 The regional international benchmarking study

LECG does not make any reference to the corroborating evidence provided by the results based on the regional international benchmarking study. We consider this to be disappointing, given the time and effort devoted to this work by ORR, ITS and the participating countries as part of PR2008. Not only does the regional international work produce similar efficiency scores for Network Rail as the LICB study, it also produces very similar rankings for the relevant countries included in both studies.

Whilst the data is confidential, when the request was made by LECG for the data to be released we offered to contact participants to ask permission for the data to be shared. However, this offer was never followed up by LECG. The regional international benchmarking study is discussed further in section 3 below.

⁴⁰ See page 141.

3. THE HORTON 4 CONSULTING REPORT

Horton 4 Consulting challenge the econometric work done by ITS/ORR in the following areas:

- justification of the use of a frontier;
- data quality and currency conversion issues;
- the omission of variables relating to the capital stock;
- the implications of extrapolating forward the time trend estimated by the preferred model;
- the use of panel data;
- the split between error or inefficiency; and
- the regional international benchmarking study.

We respond to each of these points in turn and demonstrate that none of these undermine the econometric analysis and its use by ORR in informing the regulator's judgement on an appropriate efficiency target for Network Rail. The consultants report also challenges other aspects of the ORR's efficiency analysis which ITS has not been involved in, and so we restrict our comments to the areas relevant to the international econometric analysis only.

3.1 The use of a frontier

In their report, Horton 4 Consulting argue the following:

"It is hard to justify the use of a frontier, rather than a mean, because one would expect a company at the frontier to earn supernormal profits rather than merely the cost of capital.";
see Horton 4 Consulting (2008), page 8.

As their report acknowledges, the above point is not relevant since in this study we are conducting an efficiency analysis of state owned⁴¹, regulated companies that are not assumed to maximise profits. Indeed, there exists a vast literature on the comparison of efficiency between different railways and other regulated or public sector bodies by reference to an efficiency frontier (not against the mean). Economic regulators either benchmark against the frontier (for stochastic frontier models) or against upper quartile (for COLS models), but not against the mean. As noted earlier, ORR compares Network Rail against upper quartile (even though the preferred model is based on the stochastic frontier method) which is therefore a conservative assumption.

Finally, it might further be noted that Network Rail is committed to being a world class company, so aiming for the average would not appear to be consistent with that aim.

⁴¹ Or, in the case of Network Rail, a company limited by guarantee with debt underwritten by government.

3.2 Data quality and currency conversion issues

As noted earlier in section 2.2, the LICB dataset has been developed over a number of years, based on published guidance from UIC on data definitions to aid harmonisation. It is used by UIC in its approach to international benchmarking. This does not seem to fit with the consultant's statement of Network Rail's view that the data is at an "experimental stage" (Horton 4 Consulting (2008), page 9)).

Furthermore, as noted earlier, at no point during our discussions with Network Rail were serious concerns over the quality of the data raised. Following analysis of the data, we consider the data on the variables included in our preferred specification to be well behaved and, given the points noted above, we had no serious grounds for concern regarding the quality of the data. Horton 4 Consulting further note that there is considerable variation in expenditure per route-km for different railways. This is to be expected for railways with different levels and types of traffic, and also with different proportions of single versus multiple track. Indeed, the aim of the econometric exercise was to explain this variation.

We have already discussed the issue regarding the PPP adjustment in section 2.4 above. Using PPP exchange rates is a common approach in the literature. As noted earlier, ORR has separately made a judgement on the evidence concerning differences in rail specific wage rates across different countries.

3.3 The omission of variables relating to the capital stock

Horton 4 Consulting raise the valid point that ideally the cost model would incorporate variables that capture differences in capital stock between railways; see Horton 4 Consulting (2008), section 4.2. Our model includes variables capturing some elements of the capital stock (route length, and its characteristics in terms of single track and electrification). However, our model does not contain a measure of capital stock quality, and, as noted in section 2.4 above more generally, we recognise that ideally additional variables would be included in the cost function. As a result, as part of its analysis, ORR conducted other studies in parallel to understand the likely impact of omitted variables in respect of Network Rail, and concluded that there is no reason to believe that incorporating such variables would be favourable to Network Rail.

Of course the question of the impact of the capital stock is closely related to the issue of whether Network Rail and other railways are out of steady-state or not. As noted above, we have made an adjustment to Network Rail's costs accordingly. As discussed above, we do not agree that our approach requires the assumption that all other firms are in steady-state, but rather only that the leading firms are in steady-state.

As noted earlier in our response to LECG's comments, Horton 4 Consulting's challenge to our work in this area implies that if it is not possible to develop a "perfect model", then the alternative is to do nothing. Certainly, we recognise that

there is some uncertainty here, and that the distance from the frontier may reflect both inefficiency and the impact of omitted variables. However, we consider that the results are suitable to be used to inform ORR's efficiency judgement, given that parallel work has been done by ORR to look at the impact of possible omitted variables, a steady-state adjustment has been made for Network Rail, various adjustments have been made to the raw outputs of the models, and that ORR has utilised other evidence to inform its judgement on an appropriate efficiency target for Network Rail.

3.4 The implications of extrapolating the time trend from the preferred model

In section 4.3 of their report, Horton 4 Consulting argue that the time trend variables are inappropriately specified and that extrapolation of the model to 2015 would suggest that frontier costs would fall by 88% over that period⁴².

We do not consider this to be a valid point for the following reasons. Firstly, the inclusion of a time trend (possibly including a squared term to allow for turning points) in the model is a standard approach modelling of frontier shift over time (see Coelli, Rao, O'Donnell and Battese (2005)⁴³). Secondly, the purpose of the model is to capture frontier productivity growth over the sample period (11 years) and not to extrapolate forward.

Thirdly, as Horton 4 Consulting state in their report, the overall shift in the frontier over the ten years to 2005 has been to reduce costs by just 5% (or roughly 0.5% per year). It is therefore clear that a projection that sees costs fall by approximately a further 83% over the next ten years (to 2015; to give 88% in total) is not a sensible extrapolation of the model⁴⁴. Finally, we note that Network Rail's efficiency scores are little affected if only a simple time trend is included or if no time trend is included at all⁴⁵.

3.5 The use of panel data

In section 4.4, Horton 4 Consulting challenge our use of pooled cross-section and time series data. However, the analysis that they show contains no information regarding standard errors and so the consultants are unable to say anything about the statistical significance of their analysis. Furthermore, as the consultants themselves acknowledge, panel data analysis is a common approach in efficiency modelling. It should also be noted that the modelling approach

⁴² See Horton 4 Consulting (2008), page 13.

⁴³ Page 213.

⁴⁴ The only way to generate a higher number for the cost fall from an extrapolation is to extrapolate the time element of the frontier equation from the preferred model $\text{Ln Cost} = 0.055615 \text{ time} - 0.00476 \text{ time}^2$. Though this would not be a sensible approach, as noted above, since it requires the annual rate of change of the frontier shift to increase substantially each year, even then the frontier would be 57% lower by 2015, and not 88% as reported in the Horton 4 Consulting study.

⁴⁵ Network Rail's score is 0.62 in both cases in 2006 as compared with 0.60 in the preferred model. The correlation coefficients between the efficiency scores in general for these models as compared with the preferred model are 0.9922 and 0.9918 respectively.

underlying our preferred model explicitly recognises the panel structure of the data (that is, that it consists of a number of railways over a number of years). The consultants suggest that “*Further dynamic specification may be necessary*⁴⁶”, but without giving further details as to what this would entail. The approaches that we have applied to this dataset are standard and are backed by a wide literature (see section 2.6 above).

3.6 Split between error and inefficiency

In section 5 Horton 4 Consulting suggest that:

“.it is unreasonable to suggest that the bulk of the error of each observation can be attributed to inefficiency rather than to equation error”; see Horton 4 Consulting (2008), page 15.

There follows a set of regression results and series of error plots based on applying ordinary least squares (OLS) to what appears to be country averages (based on just 13 data points), and annual averages (based on either 10 or 11 data points)⁴⁷. This approach leaves very few degrees of freedom, and we expect therefore that the parameter estimates are imprecise. In this respect we note that no standard errors are provided in the report (see Horton 4 Consulting (2008), page 15). It is not clear therefore whether the consultants statements regarding the changes (or otherwise) in coefficients across models can be regarded as precise.

For example, the statement that:

“The relationship does not seem to provide an explanation over time when estimated over annual averages”; see Horton 4 Consulting (2008), page 15

seems to be based on creating a single, European railway and carrying out a regression over 10 or 11 years using OLS. This analysis can hardly be regarded as convincing, particularly without the inclusion of standard errors on the parameter estimates.

The statement underneath Figure 4 in their report:

“ORR’s stochastic frontier analysis attempts to separate the errors into a symmetrical error with a significantly negative mean and an asymmetric inefficiency component that explains all the positive results”; see Horton 4 Consulting (2008), page 16

⁴⁶ Horton 4 Consulting (2008), page 14.

⁴⁷ It is not clear whether the consultants have used the data up to 2005 or 2006.

is simply wrong, since stochastic frontier analysis assumes that the noise error term has a symmetrical, zero mean distribution (not a significantly negative mean). The remainder of the section repeats earlier comments regarding omitted variables⁴⁸ and it is unclear precisely what conclusions can be drawn from the OLS plots shown.

In summary, we do not consider that it is possible to draw robust conclusions from the analysis presented in this section, and the analysis certainly does not provide evidence in support of the consultants' assertion that there is a problem regarding the split of the error term between inefficiency and random noise.

3.7 The regional international benchmarking study

We do not consider that section 6 of the Horton 4 Consulting report represents a robust assessment of the regional international benchmarking study which ITS and ORR have developed.

The regional international benchmarking study was developed by ITS and ORR over the period 2006 to 2008, in conjunction with the participating railways, and represents an interesting and innovative means of obtaining improved estimates of efficiency by utilising multiple observations within each country (that is, regions within each country). ITS and ORR staff visited most of the countries involved and held meetings to discuss the range of variables which might be included, as well as data definitions. Where face-face meetings were not possible, teleconference calls were held to discuss these issues.

Whilst the approach is relatively new in its development, it provides a useful cross-check against the UIC econometric work. As noted earlier, not only does the regional international work produce similar efficiency scores for Network Rail as the LICB study, it also produces very similar rankings for the relevant countries included in both studies.

The consultants make a number of largely unsubstantiated assertions. First, they question the reliability of the data without providing any supporting evidence or without ever having requested information from ORR or ITS with regard to this matter. Second, the consultants note that the model does not include a time trend. However, the reason for the exclusion of the time trend variable (that is, that for some countries, including Network Rail, there is only a single year's data) was clearly stated in our June report⁴⁹.

Third, the consultants note that the models produce differing implications in regard to economies of scale, but without noting the different interpretation of the scale variables in the different models. One considers what happens when the size of railway increases, the other looks at what happens when the size of the regional units responsible for maintenance / renewal increase. They further do

⁴⁸ It is unclear precisely what conclusions can be drawn from the OLS plots shown.

⁴⁹ Page 30.

not refer to the previous literature in this respect⁵⁰, and do not provide any information on whether the differences between the models may be regarded as different (statistically)⁵¹.

⁵⁰ As noted in our June report, previous studies (for other European rail infrastructure providers) based on disaggregated data have shown economies of scale (see ITS/ORR (2008)). Horton 4 Consulting do not provide a source for the supposed statement that ORR has made regarding the economies of scale in other studies; see Horton 4 Consulting (2008), page 19, penultimate paragraph.

⁵¹ In this respect we note that the confidence intervals for the scale parameter in the UIC econometrics permit the possibility of modest economies of scale.

4. CONCLUSIONS

In this response document we have emphatically refuted the challenges raised by LECG and Horton 4 Consulting. First of all, LECG's assertion that the Cobb-Douglas functional form adopted in our study violates economic theory in the multiple output case is shown to be incorrect. LECG's assertion is based on a single quote from one textbook, which has been taken out of the context of the wider theoretical and empirical literature, and indeed even the book from which the quote is taken.

It is clear from the literature that the multiple output Cobb-Douglas cost function does not violate any required theoretical property of a cost function in a regulated industry, such as railways, where output levels are typically assumed to be exogenously determined (see, for example, Klein (1953), Nerlove (1965), and Coelli and Perelman (2000)).

Indeed, we have shown that this functional form is widely used in both academic and regulatory studies. The appropriate functional form is, instead, an issue for econometric testing and, as noted in the ITS/ORR June 2008 report, we have tested the Cobb-Douglas model and found it to be preferred to the alternatives. Furthermore, LECG have themselves utilised a multiple-output Cobb-Douglas cost function in their recent (2005) study of postal delivery office efficiency, so it extremely puzzling that they have raised this issue in criticism of our work. This functional form has also been used by OFWAT, as LECG note in their 2005 study.

Secondly, we have demonstrated that the preferred econometric model is robust, both in its own right, and in the context of the vast array of other methods that we have applied to this dataset. LECG's assertion that a "fix" is required in order for the model to produce an estimate is shown to be incorrect. Furthermore, we have shown that the method used to derive the variance co-variance matrix (from which the standard errors and hence the means of determining the precision of the estimates are derived) is an accepted and widely used approach, and that alternative testing procedures also provide support for the method we have used.

The preferred model produces plausible estimates for the model parameters, which are also statistically significant at the usual levels of significance⁵². It also produces an extremely plausible time path of efficiency for Network Rail over the period: that is, improving modestly after privatisation⁵³, deteriorating after Hatfield, before improving during CP3. As noted, our preferred model also produces similar efficiency estimates for Network Rail to those from the other methods that we have tested, and there is also strong conformity of efficiency rankings (for all firms) across the different methods applied.

⁵² As demonstrated by the standard errors derived from the variance co-variance matrix, and alternative testing procedures as noted.

⁵³ Even after making an adjustment for renewals potentially being below steady-state during that period.

Thirdly, we have shown why we selected our preferred model over the alternative time varying efficiency model (BC92⁵⁴) put forward by LECG. Indeed, the BC92 model is shown to fail the relevant statistical tests that are normally applied. In addition, the model, which forces all firms (by assumption) to have the same direction of efficiency change over time, produces the surprising result that Network Rail's efficiency is improving every year over the post-Hatfield period (or over the whole 11 year period, depending on the precise model formulation).

The BC92 model results are clearly shown to be the outlier when compared against the alternative models that we have considered. LECG have put the BC92 model forward without reporting the results of any statistical testing of its properties, or commenting on the reasonableness or otherwise of the results in themselves and as compared to the preferred or other relevant models.

Fourthly, we have shown that the assertions of both consultants regarding the quality of the data do not fit with the facts. The LICB dataset used in our work has been developed by UIC over a number of years now (starting in 1995), and forms the basis for its own benchmarking methodology. In its ten year report on benchmarking, UIC describes the development of its approach over the period of the analysis and notes that:

“Phase 5 [of the work] provided considerable insights into cost levels and mechanisms and gave useful advice to Infrastructure Managers in Europe and overseas”, and “A “lasting benchmarking function” was established to guarantee a platform for continuous comparison of costs and the tracking of trends”; see (see UIC (2007), page 19.

UIC also produces guidance on data definitions to aid harmonisation. We therefore disagree that the data should be viewed as being at an “experimental stage” as Horton 4 Consulting state (Horton 4 Consulting (2008), page 9⁵⁵), since this is not borne out by the statements contained in UIC's own report on the data's use in its own benchmarking approach.

Furthermore, the consultants ignore the fact that ITS/ORR and Network Rail have had access to the UIC dataset since February 2007. The dataset was discussed early on in the project, and at no point during the period of our work has Network Rail expressed any serious concerns in this respect. We also noted in our June 2008 report that ORR had carried out detailed inspection work on the dataset prior to analysis. The only concrete examples that LECG are able to produce in respect of data quality relate to just two data points for one variable (number of switches) that is not included in the preferred model specification. More generally, it is puzzling at this stage of the project for Network Rail's consultants to argue that the dataset is such that it is unsuitable for analysis when Network Rail has not made that point earlier.

⁵⁴ Battese and Coelli (1992).

⁵⁵ They state this to be Network Rail's view.

Fifthly, the consultants point to possible omitted variables, in particular in respect of input prices and capital quality / variables relating to railways being above or below steady-state in respect of their renewal volumes (where steady-state is defined as the level of expenditure that is broadly required to maintain the assets in a stable condition).

All of these points were raised during our discussions with Network Rail. ORR and Network Rail (via the BSL (2008) study) therefore conducted / commissioned parallel studies to understand the likely impact of omitted variables in respect of Network Rail (further work in this area was outside the scope of ITS's remit). Ultimately ORR concluded that there was no reason to believe that incorporating such variables would necessarily lead to a significant change in the model results and be favourable to Network Rail, since there will be factors which disadvantage Network Rail as well as benefiting it. ORR has also stated, however, that further work would be helpful to try to enhance the modelling process and improve our understanding of the cost differences in future.

ITS also took advice from ORR in respect of the specific arguments surrounding steady-state. ORR expressed the view that it was not convinced that Network Rail was significantly above steady-state by the end of the period under analysis. Nevertheless, a downward adjustment was made to Network Rail's costs during the post-Hatfield period, which ORR considered to be a conservative assumption (i.e. it benefited Network Rail in terms of its relative efficiency score). For example, in 2006, this meant that the total (maintenance and renewal) cost data for Network Rail was reduced by roughly 10% as compared to the raw data prior to estimation.

Since ORR did not have sufficient data to make the same adjustment to other firms, it was assumed that the leading firms were broadly in steady-state. ORR looked at the evidence and concluded that there was no reason for doubting this assumption (see, for example, UIC (2007), which does not suggest a picture of systematic under-renewal, with renewal costs generally rising over the period covered by the dataset)⁵⁶. We also understand that during the summer of 2008 ORR has undertaken some further analysis based on the available data on relative renewal levels for some of the countries in the UIC's LICB dataset which supports the original analysis.

Furthermore, the use of the stochastic frontier approach itself (which allows for random noise effects), and the fact that we have analysed costs over an 11-year period, and not for just a single year, provide further safeguards against the risk of mis-interpreting low costs (due to a company being below steady-state) as evidence of efficient operation.

It should also be noted that the stochastic frontier approach gives greater weight to the leading firms in estimation. As a result, it is only if the leading firms are below steady-state that we would have serious cause for concern. Indeed, even if one of the leading firms was found to be below steady-state, we still have the benchmark of the remaining leading firms against which to compare Network Rail. Thus, we would expect the model to be reasonably robust even to changes in the costs of one of the leading firms.

⁵⁶ See UIC (2007), page 46.

We have also shown that the evidence put forward by LECG on steady-state uses quotes from a UIC report in a selective and unbalanced way. In one case, a quote is given that omits even the first half of the same sentence from which it is derived and thus totally changes the meaning of the quote - from suggesting a picture of falling renewals to rapidly increasing renewals. Ultimately, LECG does not put forward any clear evidence that the leading firms are below steady-state.

At this point it should be noted that, as stated in our June 2008 report, we (and ORR) recognise that ideally additional variables would be included in the cost function. We therefore accept and have always said that there is some uncertainty here, and that the distance from the frontier may reflect both inefficiency and the impact of omitted variables. However, as discussed above, we have taken advice from ORR in this regard and ORR, having looked at the evidence, has concluded that there is no reason to believe that incorporating additional variables would necessarily lead to a significant change in the model results and be favourable to Network Rail, since there will be factors which disadvantage Network Rail as well as benefiting it.

In addition, as noted, a “steady-state” adjustment has been made to Network Rail’s costs. We therefore consider that appropriate supporting work has been done in parallel to the econometric study to address the concerns raised. Furthermore, ORR has applied discount factors to the raw results of the econometric models to reduce the level of savings required during CP4 (by aiming off the frontier, and requiring two thirds of the gap to be delivered over CP4), and also combined the results with other evidence.

Finally, as noted above, it is clearly shown in this report that there are good reasons for selecting the preferred model in its own right, based on the statistical tests applied and the model’s underlying assumptions. Indeed, the general consensus of evidence here is one of a substantial efficiency gap across all methods and the preferred model produces results in the middle of the range of models estimated (see Table 4 from the main body of the report, repeated below). As noted earlier, there are strong reasons for rejecting the BC92 model put forward by LECG.

In our view it is therefore appropriate for ORR to use the results of this model as the starting point for its efficiency determination. The model implies an efficiency gap against the frontier of 40%. Indeed, ORR uses the smaller gap of 37% measured against upper quartile. Since the computation of efficiency scores relative to upper quartile is normally only applied in the case of deterministic frontier approaches (in particular COLS), which do not take account of random noise, the use of an efficiency gap measured against upper quartile in this case (where the preferred model uses the stochastic frontier method) reflects ORR’s aim to use a conservative estimate of Network Rail’s efficiency gap as its starting point (see ORR (2008)⁵⁷). Indeed, we note that in its work for Postcomm, LECG does not make any adjustment to the efficiency scores coming out of its stochastic frontier models (see LECG (2005)).

⁵⁷ See for example page 115 and 116.

Table 4

Model includes cost drivers for passenger and freight density	COLS	Random Effects GLS**	Random Effects MLE**	Random Effects MLE	Random Effects MLE	Random Effects MLE Cuesta00 (flexible)****
	Time varying	Time invariant	Time invariant	Time varying	Time varying	Time varying
				LECG***	BC92	Cuesta00 (simple)
No steady-state adjustment						
Network Rail score 2006	0.56*	0.51	0.54	0.64	0.50	0.57
Network Rail rank in 2006	13	12	12	9	12	9
Rank correlation (2006 rankings)	0.91	0.86	0.75	0.38	0.97	1.00
With steady-state adjustment						
Network Rail score 2006	0.63*	0.58	0.65	0.70	0.54	0.60
Network Rail rank in 2006	12	11	7	7	12	10
Rank correlation (2006 rankings)	0.93	0.83	0.71	0.40	0.98	1.00

* The COLS score is shown against the upper quartile. All other scores are relative to the frontier

** For these models, the 2006 score is the same as for all other years (time invariant efficiency model).

*** This is the model put forward by LECG in their challenge to the econometric work

**** This is a more flexible version of Cuesta (2000) that allows for a possible turning point in efficiency during the post-Hatfield period for Network Rail

Shading represents preferred model

In our view, ORR's starting point for its efficiency determination is therefore a reasonable one, based on the econometric work carried out. The econometric results are also supported by the regional international econometric study (see ITS/ORR (2008)). From the starting point of a 37% efficiency gap, ORR then makes a further discounting assumption that two thirds of the gap can be closed over CP4. Furthermore, ORR has combined the results of the econometric work with other evidence in arriving at its draft efficiency determination.

We therefore consider that, in general terms, ORR has made appropriate use of the econometric work in its analysis, although ITS did not review the other evidence commissioned / produced by ORR, and was not involved in the details of the process by which ORR reached its draft efficiency determination. This process resulted in the efficiency gap from the preferred econometric model of 37% being scaled down to an efficiency target for maintenance and renewals in CP3 of 22% (see ORR (2008)⁵⁸) and, of course, required ORR to exercise its regulatory judgement.

⁵⁸ See page 141.

REFERENCES

- Barros, C.P. (2005), 'Governance and Incentive Regulation in Defence Industry Enterprises: A Case Study', *European Journal of Law and Economics*, vol. 20, pp. 87-97.
- 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.
- Battese, G.E. and Coelli, T.J. (1995), 'A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data', *Empirical Economics*, vol. 20, pp. 325-332.
- BSL (2008), *Rail Infrastructure Cost Benchmarking: Brief LICB-gap analysis and cost driver assessment, final report* (April 2008).
- Coelli, T. (1996), 'A Guide to FRONTIER Version 4.1: A Computer Program for Frontier Production Function Estimation', *CEPA Working Paper 96/07*, Department of Econometrics, University of New England, Armidale.
- Coelli, T. and Perelman, S. (2000), 'Technical Efficiency of European Railways: A Distance Function Approach', *Applied Economics*, vol. 32, pp. 1967-1976.
- Coelli, T., Rao, D.S.P, O'Donnell, C.J. and Battese, G.E. (2005), *An Introduction to Efficiency and Productivity*, 2nd edition, New York, Springer.
- 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.
- Eeckaut, P.V., Tulkens, H. and Jamar, M. (1993), 'Cost Efficiency in Belgian Municipalities', in Fried, H.O., Lovell, C.A.K. and Schmidt, S.S., eds., *The Measurement of Productive Efficiency*, New York, Oxford University Press.
- Farsi, M. and Filippini, M. (2006), 'An Analysis of Efficiency and Productivity in Swiss Hospitals', *Swiss Journal of Economics and Statistics*, vol. 142, issue 1, pp. 1-37.
- Greene, W.H. (2003), *Econometric Analysis: fifth edition*, Prentice Hall.
- Hasenkamp, G. (1976), 'A Study of Multiple-Output Production Functions: Klein's Railroad Study Revisited', *Journal of Econometrics*, vol. 4, pp. 253-262.
- Horton 4 Consulting (2008), *International Efficiency Benchmarking in the Network Rail 2008 Periodic Review: ORR's Draft Determination*, report produced as part of PR2008, 31 July 2008.

- ITS/ORR (2008), *International Benchmarking of Network Rail's Maintenance and Renewal Costs*, joint report written by Andrew Smith and Phill Wheat (ITS) and Hannah Nixon (ORR) as part of PR2008, June 2008.
- Klein, W.H. (1953), *A Textbook of Econometrics*, Row, Peterson and Company.
- Kumbhakar, S.C. and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge UK.
- LECG (2005), *Future Efficient Costs of Royal Mail's Regulated Mail Activities*, August 2005.
- LECG (2008), *Assessment of the ITS/ORR Econometric Models*, report produced as part of PR2008, 4 September 2008.
- Lovell, C.A.K. (1993), 'Production Frontiers and Production Efficiency', in Fried, H.O., Lovell, C.A.K. and Schmidt, S.S., eds., *The Measurement of Productive Efficiency*, New York, Oxford University Press.
- Mulatu, A. and Crafts, N.F.R. (2005), 'Efficiency Among Private Railway Companies in a Weakly Regulated System: The Case of Britain's Railways in 1893-1912', *Department of Economic History Working Paper No. 08/05*, London School of Economics.
- Nerlove, M. (1965), *Estimation and Identification of Cobb-Douglas Production Functions*, Rand McNally and Company, Chicago.
- Network Rail (2008), *Network Rail: Response to ORR's Draft Determinations: Periodic Review 2008*, September 2008.
- ORR (2008), *Periodic Review 2008: Draft Determinations*, June 2008.
- Smith, A.S.J. (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 Regulation, report written as part of PR2008, October 2008.
- Sumicsid (2007), *Project Gerner / AS6: Development of Benchmarking Models for German Electricity and Gas Distribution: Final Report*.
- UIC (2007), *Lasting Infrastructure Cost Benchmarking (LICB): 10 years of Benchmarking 1996-2005*.