Article:
Smith, ASJ and Wheat, PE (2011) Estimation of cost inefficiency in panel data models with firm specific and sub-company specific effects. Journal of Productivity Analysis, 37 (1). 27 - 40. ISSN 0895-562X

https://doi.org/10.1007/s11123-011-0220-8

Estimation of Cost Inefficiency in Panel Data Models with Firm Specific and Sub-Company Specific Effects

ANDREW S.J. SMITH[a]

and

PHILL WHEAT[b]

January 2011

[a] Institute for Transport Studies and Leeds University Business School, University of Leeds, LS2 9JT, UK. Corresponding author. Telephone: +44 (0)113 3436654. Fax: +44 (0)113 3435334. Email: a.s.j.smith@its.leeds.ac.uk.

[b] Institute for Transport Studies, University of Leeds. Email: p.e.wheat@its.leeds.ac.uk.

Abstract

This paper proposes a dual-level inefficiency model for analysing datasets with a sub-company structure, which permits firm inefficiency to be decomposed into two parts: a component that varies across different sub-companies within a firm (internal inefficiency); and a persistent component that applies across all sub-companies in the same firm (external inefficiency). We adapt the models developed by Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995), making the same distinction between persistent and residual inefficiency, but in our case across sub-companies comprising a firm, rather than over time. The proposed model is important in a regulatory context, where datasets with a sub-company structure are commonplace, and regulators are interested in identifying and eliminating both persistent and sub-company varying inefficiency. Further, as regulators often have to work with small cross-sections, the utilisation of sub-company data can be seen as an additional means of expanding cross-sectional datasets for efficiency estimation. Using an international dataset of rail infrastructure managers we demonstrate the possibility of separating firm inefficiency into its persistent and sub-company varying components. The empirical illustration highlights the danger that failure to allow for the dual-level nature of inefficiency may cause overall firm inefficiency to be underestimated.

Keywords
Stochastic Frontier Model; Efficiency, Sub-company data; Panel Data

JEL codes

C23 ,C81, L51, D24

1. Introduction

The purpose of this paper is to propose, and illustrate via an empirical example, a dual-level inefficiency model that enables firm inefficiency to be separated into two components: a component that varies across different sub-companies within a firm (internal inefficiency); and a persistent component that applies across all sub-companies in the same firm (external inefficiency). Here we use the term "sub-company" to refer to sub-divisions within the firm based around, for example, regional or business unit structures. External inefficiency reflects the extent to which even the best performing sub-company unit in the firm fails to match best practice within the industry.

The proposed model is important for three reasons. First, in the regulatory context, where the global trend towards privatisation, and the associated development of RPI-X regulation, has led to the creation of numerous regulatory bodies with a direct interest in estimating the efficiency of firms under their jurisdiction. The proposed dual-level model should enable regulators to obtain a clearer understanding of both internal and external inefficiency. Separating out internal efficiency from the wider external inefficiency clearly has benefits to regulators (and firms) since they can use the analysis to determine the appropriate emphasis on two performance enhancing strategies. First, regulators will expect regulated firms to focus on implementing internal best practice across all sub-company units. This would aim to eliminate internal inefficiency. Regulators would also expect firms to learn from external best practice and apply this firm wide (thus eliminating external inefficiency)[1].

The second reason why we consider the proposed model to be important is the following. Economic regulators often have to work with small cross-sections. Whilst some regulators have sought to alleviate this problem by utilising panel data sets, these are often short (as a result of industry restructuring at the time of privatisation); or where longer panels exist, the ability to specify an appropriate, flexible model of time varying efficiency becomes critical, and may not be straightforward. The data structure under consideration in this paper sees the utilisation of sub-company data as an additional means of expanding cross-sectional datasets for the purpose of efficiency estimation. The utilisation of such data should therefore enable more precise measures of overall firm efficiency to be obtained.

This second benefit is directly analogous to that obtained from traditional panel data (see Schmidt and Sickles 1984). That is, the sub-company observations can be seen as multiple observations on the same firm in the same way as standard (time-based) panel data. Further, when a sub-company dataset is augmented with repeat observations over time, firm-specific time paths of inefficiency are likely to be more precisely estimated – as compared to the situation where only standard panel data (not augmented by sub-company data) is available – which again is important for economic regulators.

Thirdly, it is beneficial for both efficiency performance analysis and more generally cost analysis to analyse data at a level of geographical disaggregation that corresponds to how firms organise

[1] Even under incentive-based (RPI-X) regulation, regulators are interested in the sources of inefficiency in order to assess deliverability of savings (without compromising safety and quality), and to monitor progress. Understanding the split between internal versus external inefficiency thus provides important information for the regulator.

their activities. This allows both for any dual-level inefficiency to be captured, but also allows for the true scale and density properties of the cost frontier to be established.

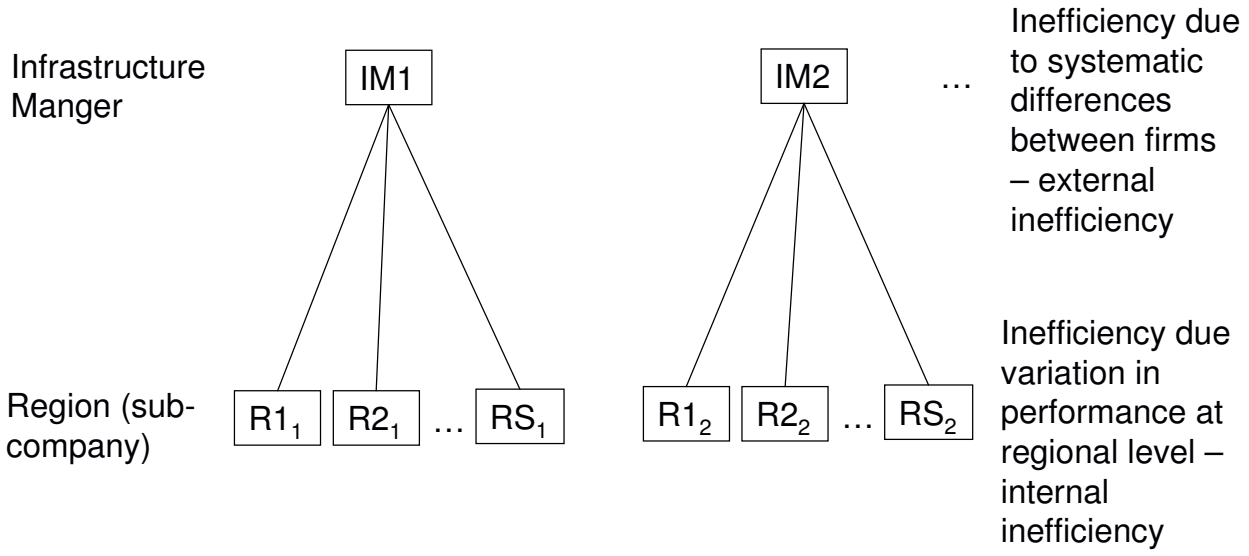
Datasets with a sub-company structure readily exist, either residing with economic regulators, or within the cost accounting systems of firms. As such it is sensible to ask how such data sets should best be exploited and this is the subject of this paper. Importantly we draw attention to some statistical tests which can be used to help identify the appropriate treatment of inefficiency effects when sub-company data is available. To our knowledge, the benefits and modelling issues associated with expanding datasets to include sub-company data – including the similarities to and differences from the standard panel case – have not been discussed in the literature.

The remainder of the paper is structured as follows. In section 2 we define what we mean by a sub-company structure. Section 3 describes the general modelling framework and interesting special cases. The possible estimation methods are then set out in section 4. Section 5 sets out the empirical example, which demonstrates some of the potential benefits from utilising sub-company data and the possible problems. The empirical application builds on an important international benchmarking study that the authors undertook in 2008, together with the British Office of Rail Regulation (ORR), aimed at estimating the efficient cost of sustaining and developing Britain's rail network. Finally, section 6 offers some conclusions.

2. The data structure

As noted in the introduction, the envisaged data structure under consideration in this paper is one which contains N firms, over T(i) years, with S(i) (i=1,…,N) sub-company units within each firm (see Figure 1). The N dimension of the data structure could either be viewed as comprising a number of regulated firms operating under the same regulatory regime (yardstick competition), or firms operating in the same industry but in different countries (international benchmarking). The precise nature of the S dimension depends on the industry, but in all cases should represent a level of disaggregation which has operational relevance and for which data is available. Figure 1 illustrates the data structure for the empirical example shown in section 5.

Figure 1 Sub-company data structure



Most regulated, network utilities, have some kind of de-centralised decision-making structure, comprising a corporate centre and separate business units, in many cases based on a regional organisational structure. The proposed model therefore potentially has very wide application in a regulatory context. As an example, the water industry in England and Wales consists of a number of water and sewerage firms, where each firm also collects data at the sub-company level, in this case derived from multiple observations on specific assets in different locations within the same firm (for example, sewage treatment plants). The water economic regulator, OFWAT, has utilised sub-company data across the regulated firms under its jurisdiction in its comparative efficiency work. However, importantly, the motivation in that case was simply to expand the size of the data set, with the data being pooled and treated merely as a larger cross-section (see OFWAT 1994; 2005).

More widely, economic regulators have commissioned internal benchmarking studies, for example in the case of rail infrastructure (see LEK, 2003) and gas distribution (see OFGEM, 2003), in order to understand variation in performance within-companies. The internal benchmarking approach is also recognised in the academic literature (e.g. Burns and Weyman Jones, 1998, and Kennedy and Smith, 2004).

The empirical application (section 5) is based on an international dataset of railway infrastructure firms. These firms are monopoly operators of the rail network in each country and therefore an external efficiency perspective cannot be obtained by looking at domestic comparators[2]. Within each network, operations are organised into smaller regional units, at which maintenance activity is organised, and these form the S dimension. The dataset also has a panel structure in time.

[2] At least in terms of efficiency levels. Some regulators have compared trends in efficiency / productivity between different industries however.

Whilst economic regulators have utilised sub-company data in a simple way and some have commissioned internal benchmarking studies as noted above, this paper is, to our knowledge, the first attempt in the literature to estimate a dual-level inefficiency model.

3. Sub-company model of inefficiency

In this section we develop a stochastic frontier cost[3] model which allows for both persistent, firm-specific and sub-company level inefficiency effects (external and internal inefficiency respectively). We also outline the interesting special cases which are used as (nested) comparator models in the empirical illustration that follows.

3.1 Dual-level inefficiency model

We consider a cost frontier transformed by taking logarithms. The inefficiency term(s), while initially multiplicative, are additive following the logarithm transformation:

$$(1)\ \ln C_{its} = \alpha + f(\mathbf{X_{its}};\boldsymbol{\beta}) + u_{its} + v_{its} \qquad i=1,\ldots,N,\ t=1,\ldots,T(i),\ s=1,\ldots,S(i)$$

where C_{its} is the cost for sub company unit s in firm i in time period t, α is a constant, $\mathbf{X_{its}}$ is a vector of logged outputs and input prices (and covariants if applicable), $\boldsymbol{\beta}$ is the conformable vector of parameters, v_{its} is a random variable representing statistical noise and u_{its} is a variable representing inefficiency. v_{its} is assumed to be distributed independently from the regressors and u_{its}.

In order to consider inefficiency effects at two levels within the firm , we decompose u_{its} into:

$$(2)\ u_{its} = \mu_{it} + \tau_{its} \text{ with } \mu_{it} \perp v_{its},\ \mu_{it} \sim \text{iid and } \tau_{its} \sim \text{iid}.$$

In this formulation u_{its} is split into two components: μ_{it}, which is the persistent element of inefficiency that applies across all sub-companies within the same company; and τ_{its}, which is the residual component that varies randomly across all sub-companies. Both inefficiency terms may either be fixed over time or vary in some way. In order to explain the economic interpretation of our model, and its position within the literature, we first drop the t subscripts from the model and focus on the sub-company structure of the data. We then briefly outline the different assumptions that may be made concerning the variation of inefficiency over time, although the time dimension is not central to this paper.

For ease of exposition we therefore now re-write equations (1) and (2) without the time subscripts as:

[3] As widely noted in the literature, the model can easily be translated into a production function by reversing the sign on u_{its}.

(3) $\ln C_{is} = \alpha + f(\mathbf{X_{is}};\boldsymbol{\beta}) + u_{is} + v_{is}$ i=1,…,N, , s=1,…,S(i)

(4) $u_{is} = \mu_i + \tau_{is}$ with $\mu_i \perp v_{is}$, $\mu_i \sim$ iid and $\tau_{is} \sim$ iid.

This formulation is analogous to that presented in Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995). In their formulation, applicable to standard panel data (i and t subscripts only), μ_i represents the persistent (over time) element of inefficiency, and τ_{it} is the residual component of inefficiency[4] (both of which are one-sided). Here we make the same distinction between persistent and residual inefficiency, but this time over sub-companies comprising a firm, rather than time. We note that in the standard panel literature the μ_i term has also been interpreted as a measure of unobserved heterogeneity (see Greene 2005; Kumbhakar 1991; Heshmati and Kumbhakar 1994). However, for the purpose of this paper we ignore unobserved heterogeneity and focus on the inefficiency interpretation.

The economic interpretation of the model outlined above is as follows. Inefficiency within an organisation is assumed to reside at two levels. First of all, there is a component due to the central management of individual firms, which sets company strategy, and various policies and standards. This is the persistent element that applies across all sub-companies within the same company (μ_i). The persistent element of inefficiency so calculated represents the best practice performance of the i-th firm relative to the best practice performance of the other firms in the sample.

Second, to the extent that the sub-company units have some degree of autonomy in how they interpret and then deliver the policies set out by central management, there is a second component that captures inefficiency at the sub-company level within each company – that is, the extent to which sub-company units fail to reach the best practice attained elsewhere within the same firm (τ_{is})[5]. Thus the model separates persistent, firm-specific inefficiency effects (or external inefficiency) from internal inefficiency at the sub-company level. It should be noted that since u_{is} is the inefficiency of each sub-company in the sample (comprising a persistent and random element), a further step is required to produce an overall measure of firm inefficiency. Overall inefficiency for an individual firm is computed therefore as the sum of the persistent element and a weighted average of the random component for each of the sub-companies within the firm:

(5) $$\bar{u}_i = \mu_i + \frac{\sum_{\forall s} C_{is} \cdot \tau_{is}}{\sum_{\forall s} C_{is}}$$

[4] We use the terms persistent and residual inefficiency as in Kumbhakar and Hjalmarsson (1995).

[5] Since the sub-company varying component is an absolute measure of inefficiency, the efficiency scores for each sub-company unit are measured relative to a theoretical frontier and for a given sample it will not necessarily be the case that one sub-company within each firm will be on the frontier.

Finally as noted in the introduction, the use of sub-company data has benefits for performance analysis and more generally cost analysis beyond the ability to measure dual level performance. It substantially increases the number of observations for analysis which addresses a common problem in economic regulation (small N). Further, aggregation bias can arise in an estimated cost function if data is aggregated at a level that is not equivalent to the level at which operational decisions are actually made (Theil, 1954). For example analysing infrastructure of railways using national data may lead to misleading estimates of economies of network size if the railway is in fact organised into zones. A more useful concept would be to look at the economies relating to network zone size. Obviously much depends on what the analyst is trying to understand in the first place, but we do note that for cost analysis, sub-company data provides a much richer dataset to investigate much more subtle distinctions regarding economies of size and density.

3.2 Sub-company inefficiency invariance model

One interesting special case that is nested within the model outlined in equations (3) and (4) is the sub-company inefficiency invariance model. Where it is reasonable to assume that $\tau_{is} = 0 \; \forall \, i,s$, that is, all inefficiency is persistent across sub-companies in a firm, and thus there is no additional inefficiency variation between sub-companies comprising a firm, then the model can be written:

$$(6) \quad \ln C_{is} = \alpha + f(\mathbf{X_{is}};\boldsymbol{\beta}) + \mu_i + v_{is} \qquad i=1,\ldots,N,\ s=1,\ldots,S(i)$$

In this case the model has reduced to a more conventional model, analogous to the time invariant inefficiency models of Pitt and Lee (1981) or Schmidt and Sickles (1984), but with inefficiency invariance in sub-companies comprising a firm rather than across time.

We note here that one of the weaknesses of the time invariant model in the standard panel inefficiency model literature is that it may not be appropriate to assume that inefficiency is invariant over time, particularly when panels are long (and of course it is exactly when panels are long that the benefits of the panel approach to inefficiency estimation are fully felt). Whilst the assumption of sub-company inefficiency invariance may likewise be challenged – in fact, the presence of sub-company effects is the motivation behind the dual-level efficiency model – this assumption may be a reasonable approximation in some circumstances (when there is little sub-company autonomy). Furthermore, the assumption does not necessarily become more implausible as the number of sub-company units is increased (as is the case for long panels). Importantly, since this model is nested within the dual-level inefficiency model, we can test for the absence of sub-company inefficiency variation.

3.3 The pooled model

The restriction $\mu_i = 0 \;\; \forall i$ yields a simple pooled model in which the inefficiency of each sub-company ($u_{is} = \tau_{is}$) is assumed to be identically and independently distributed across all sub-company units irrespective of which firm they belong to. In this case the central management in each firm plays no role at all from an inefficiency perspective. Since this model is nested within the dual-level inefficiency model, we can test for the absence of a persistent, firm-specific inefficiency component.

3.4 Assumptions about inefficiency variation over time

The empirical illustration shown in this paper comprises data both at sub-company level and over time. However, the focus in this paper is on the sub-company dimension of the panel structure. Therefore, the dual-level inefficiency model outlined in equations (1) and (2) makes a simple assumption concerning the variation in inefficiency over time ($\mu_{it} \sim \text{iid}$ and $\tau_{its} \sim \text{iid}$). The pooled model likewise makes a simple assumption regarding the variation in inefficiency over time ($u_{its} = \tau_{its} \sim \text{iid}$). In the sub-company invariance inefficiency model estimated (equation (5)), where $\tau_{its} = 0$, firm inefficiency (μ_i) is assumed to be invariant over both sub-company and over time.

It should be noted, however, that it is possible to make alternative assumptions about the behaviour of both the μ_{it} and τ_{its} inefficiency terms over time. These include independence and time invariance over time as noted above, but could be extended to allow varying inefficiency over time via a deterministic scaling model (presented in the most general forms in Kumbhakar and Lovell 2000; Orea and Kumbhakar 2004). However, for the purpose of this paper, which focuses on the sub-company dimension of the panel structure, in the empirical example we retain one of the simple assumptions noted above (time invariance), and leave the development of more complex time varying models to further work (see section 5). We do show how to estimate such paths in section 4 for the case of the sub-company invariance model. Importantly sub-company data structure potentially provides a powerful way to estimate firm specific paths of inefficiency over time, since there can be many observations per firm relative to the number of time periods, vis-à-vis the use of panel data where the number of observations per firm is equal to the number of time periods to which they are observed.

4. Estimation

4.1 Dual-level inefficiency level model

We first introduce the estimation framework, which draws on the approach by Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995). In this framework we consider equation (1) rewritten as:

(7) $\ln C_{its} = \alpha_{it} + f(\mathbf{X_{its}}; \boldsymbol{\beta}) + \tau_{its} + v_{its}$

where $\alpha_{it} = \alpha + \mu_{it}$.

At this stage we have not made distributional assumptions on the two inefficiency error components, except that they are distributed independently of the random noise term v_{its} and independently of each other. We now make additional assumptions to facilitate estimation. First we make assumptions as to whether μ_{it} or correspondingly α_{it} are correlated with the regressors. If so we consider α_{it} to be a fixed effect. If not then we could consider α_{it} to be a random

effect. Second we make the assumption that τ_{its} is uncorrelated with the regressors and τ_{its} is a random effect. Treating τ_{its} as a random effect is a necessary assumption for the case of T=1.

This model could be estimated in several ways. The first two methods use maximum likelihood to estimate the model in one stage. These are variants of the 'True' fixed and random effects models proposed by Greene (2005). In both cases $\tau_{its} \sim \text{iid}\left|N\left(0,\sigma_{\tau}^{2}\right)\right|$ and $v_{its} \sim \text{iid}N\left(0,\sigma_{v}^{2}\right)$, however it is possible to relax the assumption of homoscedasticity and zero mean of the (untruncated) distributions (Greene 2005). The formulation is the same as the original formulation of the pooled stochastic frontier model proposed by Aigner et al (1977), but with effects by firm per time period[6].

In the True fixed effects case, α_{it} is treated as a fixed effect and maximum likelihood is used to estimate the model. This case allows α_{it} to be correlated with the regressors. A potential disadvantage of this estimation approach is, because of the presence of fixed effects, estimates of all parameters in the model (not just the fixed effects) may be inconsistent and biased. This is known as the incidental parameters problem (Neyman and Scott 1984; Lancaster 2000). Greene (2005) provides Monte Carlo evidence that the bias does not appear to be substantial when T=5, which is encouraging given the short nature of panels typically available for performance analysis studies.

Estimation of this model by maximum simulated likelihood yields estimates of $\alpha_{it}, \boldsymbol{\beta}, \sigma_{v}^{2}, \sigma_{\tau}^{2}$. Ignoring for now the fact that the α_{it}'s are estimates and not population values, following Schmidt and Sickles (1984),

(8) $\min\left(\alpha_{it}\right) \xrightarrow[p]{} \alpha \quad T \to \infty$

As such a consistent estimator of μ_{it} is given by

(9) $\hat{\mu}_{it} = \alpha_{it} - \max\left(\alpha_{it}\right) \xrightarrow[p]{} \mu_{it} \quad T \to \infty$

For finite T, this method of recovery of μ_{it} results in a measure of relative inefficiency (relative to the best performing firm/time observation). However this estimator cannot be constructed because the α_{it}'s have to be estimated. Thus the feasible estimator of μ_{it} is:

[6] Note that by effects by firm per time period we do not mean that this has two way effects in firm and time. Instead we mean there is one set of effects, with one effect for each year and firm. This is very general. We could replace this with an assumption that the persistent inefficiency of sub-companies in a firm is also time invariant, in which case $\alpha_{it} = \alpha_{i} = \alpha + \mu_{i}$. This is the assumption we use in our empirical example. A further assumption could be that $\alpha_{it} = \alpha_{i1} + \alpha_{i2}t + \alpha_{i3}t^{2}$, that is that the persistent inefficiency follows a Cornwell et. al. (1990) type variation over time.

(10) $\hat{\hat{\mu}}_{it} = \hat{\alpha}_{it} - \max(\hat{\alpha}_{it})$

The analytic conditional expectation estimator proposed by Jondrow et al (1982) can be used to calculate a point estimate for the residual component of inefficiency, τ_{its}:

(11) $E[\tau_{its} \mid \varepsilon_{its}] = \rho_{its*} + \sigma_* \frac{\phi(\rho_{its*}/\sigma_*)}{1-\Phi(\rho_{its*}/\sigma_*)}$

where $\rho_{its*} = \sigma_\tau^2 \varepsilon_{its} / (\sigma_\tau^2 + \sigma_v^2)$, $\sigma_*^2 = \sigma_\tau^2 \sigma_v^2 / (\sigma_\tau^2 + \sigma_v^2)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf respectively. To operationalise this, σ_τ^2 and σ_v^2 are replaced with their corresponding estimates and ε_{its} with

(12) $\hat{\varepsilon}_{its} = \ln C_{its} - \hat{\alpha}_{it} + f(\mathbf{X_{its}}; \hat{\boldsymbol{\beta}})$

In the true random effects case, α_{it} is treated as random and assumed independent of the regressors. We estimate this model by simulated maximum likelihood, rather than simple maximum likelihoods because simulation is used to integrate out the random effect α_{it} from the likelihood function. Unlike the formulation in Greene (2005), a normal distribution cannot be assumed for this effect, since this variable is truncated from below at α. Instead we assume that α_{it} comprises:

(13) $\alpha_{it} = \alpha + \mu_{it} \quad \mu_{it} \sim \text{iid}\left|N(0, \sigma_\mu^2)\right|$

The model now comprises the usual composite error term as proposed by Jondrow et al (1977) distributed independently by each sub-company and by time, but also a random parameter, the constant term, which varies independently by firm and by time period. Estimation of this model by maximum simulated likelihood yields estimates of $\alpha, \boldsymbol{\beta}, \sigma_v^2, \sigma_\tau^2, \sigma_\mu^2$. Firm and time specific estimates of α_{it}, denoted $\hat{\alpha}_{it}$, are estimated as the expectation of α_{it} conditional on the data and the estimated parameters as given in equation 32 in Greene (2005). This is a consistent estimator as $T \to \infty$ (Train 2003, p. 269). This is approximated during the simulation of the likelihood function in estimation. μ_{it} is then estimated as:

(14) $\hat{\hat{\mu}}_{it} = \hat{\alpha}_{it} - \hat{\alpha}$

Importantly note that the estimate of μ_{it} is an estimate of absolute persistent inefficiency as opposed to the relative measures which are produced by the other estimation methods discussed in this paper. This is because α is estimated through the maximum simulated likelihood process since it is the truncation point and mean of the underlying normal distribution of α_{it}. An estimate

for the residual component of sub-company inefficiency, τ_{its}, is the same as for the True fixed effects case.

An alternative estimation framework is the multistage approach outlined in Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995). In this approach the model is first estimated by either within or generalised least squares estimation, depending on whether the α_{it} s are treated as fixed or random effects respectively. Following this estimation the residuals, $\hat{\varepsilon}_{its}$, are computed and these are used to compute the fixed or random effects, $\hat{\alpha}_{it}$ (as outlined in Kumbhakar and Hjalmarsson 1995; Kumbhakar and Heshmati 1995). An estimate of μ_{it} is then recovered as

$$(15)\quad \hat{\hat{\mu}}_{it} = \hat{\alpha}_{it} - \max\left(\hat{\alpha}_{it}\right)$$

The second stage comprises the use of conditional maximum likelihood estimation[7] to estimate the parameters of the specified distributions of τ_{its} and v_{its}. Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995) utilise a half normal and normal distribution for the two error components. Adopting these distributions for τ_{its} and v_{its} the conditional log likelihood function (for each observation) is[8]:

$$(16)\quad \ell_{ist}\left(\sigma,\lambda \mid \boldsymbol{\beta},\alpha,\varpi_{its}\right) = \text{constant} - \ln\sigma + \ln\Phi\left(\varpi_{its}\lambda/\sigma\right) - \frac{1}{2}\left(\varpi_{its}/\sigma\right)^2$$

Where $\varpi_{its} = \tau_{its} + v_{its}$, $\lambda = \sigma_\tau/\sigma_v$ and $\sigma = \sqrt{\sigma_\tau^2 + \sigma_v^2}$. We replace ϖ_{its} with the consistent estimates given in the earlier stages by $\hat{\varpi}_{its} = \hat{\varepsilon}_{its} - \hat{\hat{\mu}}_{it}$.

Summing over all observations and maximising with respect to σ and λ yields consistent estimates of the parameters of the distributions of τ_{its} and v_{its}. Following this, the Jondrow et al (1982) estimator can be applied as above to yield an estimate of τ_{its} as given in equation x.

Importantly in the first stage, no distributions have been specified for any error components. As such the main parameter estimates, $\boldsymbol{\beta}$, are consistently estimated even if the resulting distributional assumptions in the second stage prove incorrect. Also if the α_{it} 's are treated as fixed effects, the multistage approach has the advantage that this model does not suffer from the incidental parameters problem since in the first stage, the incidental parameters are swept out by the within transformation. Thus it is possible to introduce correlation between the firm persistent inefficiency component and the regressors without the potential inconsistency resulting from the incidental parameters problem. The inevitable trade-off against this robustness is a loss of

[7] Conditional on the (consistent) estimates in the first stage.

[8] Note that we reverse the sign on $\varpi_{its}\lambda/\sigma$ vis-à-vis Kumbhakar and Heshmati (1995) since we are estimating a cost frontier.

estimation efficiency relative to specifying (correctly) a full maximum likelihood function to be estimated (such as using the approach by Greene (2005) above).

Since the remaining error components ($\tau_{its} + v_{its}$) are assumed to not be correlated with the regressors and μ_{it} both estimation methods are consistent[9].

4,2 Sub-company inefficiency invariance model

The case of both time invariant inefficiency and independence over time is an extension of the Pitt and Lee model with slightly different subscripts. As such we refer readers to Pitt and Lee's (1981) paper for details of the likelihood function. Likewise for the time varying models these are trivial extensions of the general time varying presented in Kumbhakar and Lovell (2000) and Orea and Kumbhakar (2004). The likelihood function for the model for standard panel data is presented in Kumbhakar and Lovell (2000) and this requires only trivial sub-script amendments to form the required likelihood functions for the variants of the model discussed in 3.2. We assume that the distribution of the inefficiency term is:

$\mu_{it} \sim N^{+}\left(\pi_i, \sigma_\mu^2\right)$ when independence over time is assumed for inefficiency and

$\mu_{it} = g_i\left(\boldsymbol{\delta}' \mathbf{Z_{it}}\right) \cdot \mu_i$ with $\mu_i \sim N^{+}\left(\pi_i, \sigma_\mu^2\right)$ when dependence of inefficiency over time is allowed for.

For all of these models, except the model which assumes independence over time of inefficiency, an estimate of firm inefficiency is given by the conditional expectation of the inefficiency component and is amended from Greene (2008) and given below:

$$(17)\quad E\left[\mu_{it} \mid \varepsilon_i\right] = g_i(\cdot) E\left[\mu_i \mid \varepsilon_i\right] = g_i(\cdot)\left[\tilde{\mu}_i + \tilde{\sigma}_i\left(\frac{\phi\left(\tilde{\mu}_i / \tilde{\sigma}_i\right)}{\Phi\left(\tilde{\mu}_i / \tilde{\sigma}_i\right)}\right)\right]$$

Where $\varepsilon_i = \varepsilon_{i11}, \ldots, \varepsilon_{i1S(i)}, \varepsilon_{i2S(i)}, \ldots, \varepsilon_{iT(i)S(i)}$, $\tilde{\mu}_i = \dfrac{(1-\gamma)\pi_i - \gamma \sum_{t=1}^{T(i)} \sum_{s=1}^{S(i)} g_i(\cdot)(d\varepsilon_{its})}{(1-\gamma) + \gamma \sum_{t=1}^{T(i)} \sum_{s=1}^{S(i)} (g_i(\cdot))^2}$,

$\tilde{\sigma}_i^2 = \dfrac{\gamma(1-\gamma)\sigma^2}{(1-\gamma) + \gamma \sum_{t=1}^{T(i)} \sum_{s=1}^{S(i)} (g_i(\cdot))^2}$, $\gamma = \sigma_\mu^2 / \sigma^2$, $\sigma^2 = \sigma_\mu^2 + \sigma_v^2$, $d = \begin{cases} 1 \text{ if production function} \\ -1 \text{ if cost fucntion} \end{cases}$

The equivalent estimator for the case of independence across time is a trivial adaptation of the estimator presented in Battese and Coelli (1988) (summation over s rather than over t) and so we do not present it here.

[9] Provided in the GLS case the regressors and μ_{it} are uncorrelated as discussed earlier.

5. Empirical application – International railway infrastructure comparisons

5.1 Context

We illustrate our approach by estimating a dual-level inefficiency model using data on five railway infrastructure managers, comprising firms from North America alongside European national infrastructure managers (IMs). A railway infrastructure manager is responsible for the management (maintenance and renewal) of the railway infrastructure (permanent way, structures, line side equipment and stations and depots). An infrastructure manager is different conceptually from a train operator who actually runs the train services. In the case of Britain, the infrastructure manager is institutionally separate from train operating companies. For the other companies, the IM also runs the train services, but importantly, separate accounts are available for the IM side and also the structure of the companies is such that the two functions can be considered divorced in terms of business organisation.

As noted earlier, this paper builds on a previous study conducted for the British Office of Rail Regulation (ORR) as part of the 2008 Periodic Review of the British infrastructure manager's efficiency performance[10]. In that work, which was exploratory in nature, and based on a smaller sample than we now have available, we estimated the simplest, single-level efficiency versions of the models presented in this paper (namely the pooled and sub-company invariant models; see section 3).

Each IM in the sample is divided into a number of regions. The number of regions per IM (S(i) using the terminology in section 3) ranges from 3 to 18. The difference in the number of regions per IM reflects both the availability of data (in respect of the number of years available for each firm) and also, importantly, the organisational structure of the IM. Thus the definition of a region for each IM is such that it is expected that there exists some management autonomy at the regional level as well as at the firm head office level. Hence, there is a need at least to consider a dual-level inefficiency model.

As noted in section 3, it is beneficial for both efficiency performance analysis and more generally cost analysis to analyse data at a level of geographical aggregation that corresponds to how firms organize their activities. This allows both for any dual-level inefficiency to be captured, but also allows for the true scale and density properties of the cost frontier to be established. Thus while the range of regions per IM may appear large, this is partly due to the overall size differences of the IMs considered. Further, we have assurance that these breakdowns have degrees of autonomy, thus making it appropriate to analyse efficiency at this level.

For some IMs our dataset is supplemented by having repeat observations over time (T(i) ranges from 1 to 5). The panel covers the period 2002 to 2007, though is unbalanced in time as noted. Overall we have a total of 89 observations on the five IMs. As discussed in section 3, an assumption about how inefficiency behaves over time is required in this case. Given the unbalanced nature of the observations over time and the generally small number of time periods

[10] See Smith et. al. (2008) and ORR (2008) for details of the work undertaken. Note that the railway companies considered are slightly different in the analysis for this paper than in the Periodic Review analysis.

for most IMs, we choose to adopt a time invariant model. Thus both the firm and sub-company inefficiency components are time invariant in our model.

The data structure enables the investigation of efficiency variation between rail systems in different countries, whilst also looking at inefficiency at the sub-company level within each system. The use of sub-company data also expands the sample size substantially without the need to collect a long panel. The utilisation of sub-company data can thus be seen as interesting and important in an international benchmarking context where cross-sections may well be small and panels short.

It should be noted that, given the sensitive nature of efficiency analysis and its implications for the companies (both from a competitive and regulatory perspective), the efficiency scores for individual companies and sub-company units are anonymised. This commitment was a formal requirement prior to obtaining the data and without which the data would not have been released for analysis. However, the results still enable us to draw conclusions about the impact of alternative methods on the firm efficiency scores, as well as the split between persistent and sub-company varying inefficiency, which is the primary focus of this paper.

5.2 Data

The data is summarised in Table 1. The dependent variable is maintenance cost, comprising all elements of railway infrastructure maintenance (e.g. permanent way, structures and signalling). Note that in railway accounts, maintenance is distinct from renewals activity, where renewals expenditure is the like-for-like replacement of assets following life expiration and maintenance expenditure is the day to day up keep of the assets to keep them in safe and operable condition. Whilst there could be definitional differences between countries which affect this variable (as is the case in any international study) as part of data collection process considerable efforts were made to harmonise definitions across countries which adds to our confidence in the data. We convert the country specific cost data into US dollars using purchasing power parity (PPP) exchange rates. We also convert the data to 2006 constant prices.

Our explanatory variables comprise tonne density, defined as gross tonne-km per track-km (TTKD) and track-km (Track) for outputs in order to account for scale and density effects. We also include the proportion of track length that is electrified (ProElect) as a proxy for the quality of the infrastructure. We do not have price indices for capital between countries, but note that the PPP exchange rate adjustment should account for some of the differences across countries. We do have wage rate data for each of the IMs. We do note that these are company-wide rather than sub-company specific and that in some cases the data is based on all staff employed by the railway, not just infrastructure maintenance. Thus the Wage variable is relatively crude and as such we discuss the sensitivity of our results to its inclusion. The data is normalised to the sample mean which implies the coefficients on the first order variables represent elasticities at the sample mean[11].

[11] Note ProElect is not normalised to the sample mean.

Table 1 Summary of data used in the study (unnormalised data)

Variable	Mean	Standard Deviation	Min	Max
Maintenance Cost	43,801,077	28,162,452	9,103,240	114,210,161
Tonne Density (Tonne-km / Track-km) (TTKD)	8,059,323	6,157,594	1,077,481	21,808,976
Track-km (Track)	928	588	252	2,988
Proportion of track-km electrified (ProElect)	0.65	0.41	0.00	1.00
Average staff cost per staff member (Wage)	57,408	9,473	39,791	84,378

Note: costs are in 2006 US $

5.3 Results

In Table 2 we present the parameter estimates from the dual-level efficiency models, estimated by assuming the α_i are fixed and random effects in turn (we use LIMDEP v9 to operationalise the multistage fixed and random effects estimation approaches, and details of the code are available from the authors on request). We also present the parameter estimates for the two special (nested) cases of the dual-level model as discussed in section 3. First, the sub-company inefficiency invariance model (fixed and random effects cases), which corresponds to the fixed / random effects models used as the first stage in the dual-level model. Second, we show the special case where inefficiency is only sub-company varying (no persistent, firm-specific effects), which we refer to as the pooled model in line with the terminology used in section 3[12].

The functional form was chosen by first estimating a Translog and then testing down. The vast majority of second order terms had very low t statistics and in addition to the squared track term, only an interaction term between wage and track was significant at any reasonable significance level. However inclusion of this term yielded a model with implausible negative wage elasticities for many observations within the sample. For this reason, we dropped this term. Importantly, the joint restriction that all of the omitted second order terms (including the wage / track interaction) were equal to zero could not be rejected at any reasonable significance level (e.g. Wald test in random effects model treatment gave a statistic value of 12.12 and an associated p value of 0.19804 (9 degrees of freedom)). As such we conclude that our specification is both a useful and intuitive economic model of the underlying cost characteristics while its parsimony is supported by the data.

Turning to the choice of fixed versus random effects, first note as discussed in section 3, this refers to the persistent, firm-specific effect in the model (α_i). The Hausman test gives a p value of 0.0861 which indicates a preference for random effects at the 5 per cent significance level. However we still report the fixed effects results for comparative purposes.

[12] We note that while the terminology "pooled model" accurately describes the pooled nature of the data over sub-companies, it should be noted that time invariance is assumed. As such the model is actually an analogue to the time invariant model first proposed by Pitt and Lee (1981).

Table 2 Parameter estimates for dual-level Inefficiency models and comparator models

	Dual Level Inefficiency Models				Comparator Models					
					Sub-company inefficiency invariance model					
	Fixed Effects Treatment of μ		Random Effects Treatment of μ		Fixed Effects[1]		Random Effects[1]		Pooled model	
Deterministic Frontier										
lnTrack	0.84514	***	0.88682	***	0.84514	***	0.88682	***	0.93453	***
lnTTKD	0.27821	***	0.30374	***	0.27821	***	0.30374	***	0.3465	***
ProElect	0.27771	**	0.18201		0.27771	**	0.18201		0.06895	
lnWage	0.00809		0.45837	**	0.00809		0.45837	**	0.61462	***
$(lnTrack)^2$	-0.23589	***	-0.19374	***	-0.23589	***	-0.19374	***	-0.15511	

*** statistically significant at the 1% level, ** statistically significant at the 5% level

[1]Note that these parameter estimates are the same as for the dual-level models due to the two stage estimation approach of the dual-level models used in this example.

Turning to the scale and density findings implied by the frontier parameter estimates, our results indicate modest returns to scale (RTS[13]) at the sample mean. RTS is defined as the inverse of the elasticity of costs resulting from a proportionate increase in track length (region size), holding traffic density (TTKD) constant. This measure implicitly therefore requires train-km to increase by the same proportion as region size and is thus analogous to returns to scale. Since our model is expressed in terms of the logs of track length (lnTrack) and traffic density (lnTTKD), RTS is computed as 1 / 0.88682 = 1.13 at the sample mean (for the random effects model) and this is significantly different from unity at the 5% level (both random and fixed effects models).The sign of the coefficient on the $(lnTrack)^2$ variable indicates that the RTS measure increases with track length and the variation within the sample is plausible (RTS between 0.7 and 2.5).

We also find much stronger returns to density. Returns to density (RTD) is defined as the inverse of the proportional change in cost resulting from a proportion change in train km holding network size constant. Given the way the variables enter our model, RTD corresponds to the inverse of the cost elasticity with respect to train density (TTKD) and is thus calculated as 1 / 0.30374 = 3.29 at the sample mean (for the random effects model), which is highly significantly different from unity in both the random and fixed effects models.

Recent evidence, based on models of rail infrastructure costs, suggests increasing returns to scale, combined with strong returns to density (see Wheat and Smith 2008; Wheat et. al. 2009; summarised in Table 3 below). Our findings of modest increasing returns to scale combined with strong economies of density are therefore in line with previous evidence[14]. Overall we would expect RTD to be much stronger than RTS for rail infrastructure, given that only a small proportion of infrastructure maintenance costs are variable with marginal increments in usage.

[13] See Caves et. al. (1981) and Caves et. al. (1984) for use of the terms returns to scale (RTS) and returns to density (RTD) in empirical applications.

[14] In interpreting these results it should be noted that the final two studies in Table 3 utilise firm-level data, whilst the other studies utilise sub-company data of varying levels of disaggregation.

Thus there is a substantial proportion of maintenance of cost that is only avoidable through line closure (see for example Wheat and Nash (2008) and AEA Technology (2005)).

Table 3 Estimates of Returns to Scale and Density from other infrastructure maintenance cost studies

Study	Country	Returns to Scale	Returns to Density
Our Study	**International study**	**1.13**	**3.29**
Munduch et al (2002)	Austria	1.449-1.621	3.70
Link (2009)	Austria	Not reported*	1.82
Wheat and Smith (2008)	Britain	2.074	4.18
Johansson and Nilsson (2004)	Finland	1.575	5.99
Tervonen and Idstrom (2004)	Finland	1.325	5.74-7.51
Gaudry and Quinet (2003)	France	Not reported*	2.70
Gaudry and Quinet (2009)	France	Not reported*	2.56
Johansson and Nilsson (2004)	Sweden	1.256	5.92
Andersson (2006)	Sweden	1.38	4.90
Andersson (2009)	Sweden	Not reported*	4.00
Marti et. al. (2009)	Switzerland	Not reported*	4.54
Smith et. al. (2008)	International study	1.11	3.25
NERA (2000)	US	1.15	2.85

Source: Amended from Wheat and Smith (2008) and Wheat et. al. (2009). RTS and RTD computed based on average elasticities or elasticities at the sample mean. * Obtaining measures of returns to scale was not the focus of the analysis and these cannot be derived from the paper given the functional form used.

Of course, there is also a much wider literature based on vertically-integrated railways, covering infrastructure and operations. Studies from the US suggest constant returns to scale, whilst the evidence is more mixed in respect of European railways, ranging from decreasing through to increasing returns to scale. The literature is, however, more conclusive on reporting increasing returns to density (see, for example, see, for example, Caves et. al., 1985; Gathon and Perelman, 1992; Andrikopolous and Loizides, 1998; and Smith, 2006). Thus our model produces estimates for RTS and RTD that are in line with both the infrastructure-only rail cost literature and the broader vertically-integrated railway literature. In this regard we also note that the RTD reported in our study, and the range of studies in Table 3, are typically greater than those reported for the vertically integrated railway literature, which is to be expected since stronger returns to density are anticipated in respect of infrastructure than operations (see Nash, 1985).

The implication of our findings on scale and density are that for this sample there would be cost savings from making maintenance regions bigger (increasing returns to scale). The policy prescription may therefore be for regulatory bodies to press for internal re-organisation, though that decision would need to be assessed against the loss of yardstick information (loss of a region), and should also consider other evidence[15]. Further, as is commonly reported in railway studies, our study indicates that there are substantial unit cost savings from utilising networks more intensively (increasing returns to density). The policy implications of the latter are probably

[15] Note that we find the degree of RTS to increase with track length, which if interpreted literally and simply extrapolated, would imply a single region within each company. Of course we are more confident in the findings of our model at the sample mean than at the extremes of the sample (or even out of sample).

limited to the extent that most network duplication has been eliminated, and the network structure is largely determined by political considerations.

The *a priori* sign of ProElect is ambiguous given the extent to which the variable is a proxy for track quality (that is, higher quality track might be expected to have lower maintenance costs). On the other hand, electrification means that there are more assets to maintain, makes access to the infrastructure more complex and may also be associated with higher speed services which increases cost. Thus the positive coefficient on ProElect (only significant in the fixed effects model) is neither in line nor at odds with prior expectations. The literal interpretation of the coefficient in the random effects model, given that ProElect is a proportion variable, is that electrifying the network (from 0% to 100% track-km electrified) increases maintenance costs by exp(0.18201)-1=20%.

The coefficient on the wage variable is statistically significant in the random effects model. We believe that the wage coefficient is insignificant in the fixed effects model since this variable is invariant for each IM at a given point in time. Thus it is likely there is some correlation between this and the fixed effects (note however that the Hausman test still prefers random effects). However, in both models the null hypothesis that the coefficient is different from the average labour cost share (65%)[16] fails to be rejected even at the 10% level. Dropping the wage variable does not seem to affect the estimates of the deterministic cost frontier.

Overall we find that the parameter estimates are in line with expectations and previous evidence, thus giving us confidence in the resulting efficiency findings, to which we now turn.

First we consider the statistical significance for each of the inefficiency components within our model[17]. The persistent, firm-specific inefficiency effects are modelled as either fixed or random effects, the latter being estimated by generalised least squares in the two-stage approach that we adopt here. As such we do not undertake LR tests for whether the variance parameters are zero as these are not estimated in this estimation framework. Instead we undertake an F test for the joint significance of the fixed effects and an LM test for the appropriateness of a model without effects. The F test has a value of 5.34 which yields a p value of 0.00073. As such we find evidence that the fixed effects are jointly statistically significant.

For the LM test we adopt the Moulton/Randolph standardised form (SLM, Moulton and Randolph, 1989) which is appropriate for unbalanced panels and is a one sided test (the variance of the random effect can only be non-negative). Thus we would expect the test to have greater power than the more standard Breusch and Pagan (1980) test. The value of the SLM statistic is 4.59 and is distributed standard normal under the null of zero random effect variance. Thus we can reject a model with no effects at any reasonable significance level. Thus all of the tests provide evidence of significant persistent firm-specific effects. These are then transformed into persistent efficiency scores via a Schmidt and Sickles (1984) transformation as described in section 4.

[16] Owing to lack of data, this is an estimate based solely on Network Rail data.

[17] As noted in section 3 and section 5.1, given the unbalanced nature of the observations over time and the generally small number of time periods for most IMs, both the firm-specific and sub-company inefficiency components are time invariant in our model.

Turning now to the statistical significance of the sub-company varying inefficiency term. In the two stage approach adopted for this example, we estimate this term by maximum likelihood. As such we undertake LR tests for the significance of the variance parameter of the inefficiency distribution. For the dual-level random effects model, the LR statistic is 18.15 and for the dual-level fixed effects model the LR statistic is 33.23. As described in Coelli et. al. (2005), this statistic has a non-standard mixed chi square distribution (1 degree of freedom). The large statistic values mean that in both cases the null hypothesis of zero variance is rejected at any reasonable significance level. As such we conclude that we have evidence that the data set exhibits dual-level inefficiency.

Table 4 shows overall firm efficiency scores for each infrastructure manager. It also decomposes the efficiency scores into the two components; persistent and sub-company varying. As explained in section 3, these two components can be interpreted as the degree of external and internal inefficiency respectively. In this example the average persistent efficiency scores for the dual-level models are 0.849 and 0.835 (random and fixed effects formulations respectively), and 0.851 to 0.690 for the sub-company varying component (random and fixed effects formulations respectively). Thus the random effects formulation points to roughly equal external and internal components, while the fixed effects formulation points to more internal than external inefficiency. As discussed earlier we prefer the random effects results due to the result of the Hausman test. Overall firm efficiency is the product of the two components, and is higher, on average, for the random effects dual level model (0.724) than for the fixed effects alternative (0.564). The overall efficiency scores for the preferred random effects dual level model are within plausible ranges.

Recall the comparator models are the pooled model and sub-company inefficiency invariance model. The former assumes that there is no persistent inefficiency within firms, and the inefficiency of each sub-company is assumed to be identically and independently distributed across all sub-company units irrespective of the firm to which they belong. The second comparator model comprises persistent, firm-specific effects only, representing the case where there is no variation in efficiency performance between sub-company units within the same firm (the model parameters for these models are simply those for the dual-level models reported in Table 2). Note that the average overall firm efficiency is considerably lower using the dual-level model as compared to all three of the comparator models. This is because the comparator models are constrained models and only consider one source of inefficiency. As discussed above, both restrictions are rejected for this dataset so we prefer the dual-level models.

In summary, this empirical example has demonstrated the possibility of separating firm inefficiency into a persistent and a sub-company varying component. It also shows that the failure to account for the dual-level nature of inefficiency, for example, by estimating one of the three, simpler comparator models, may cause overall firm inefficiency to be underestimated. We consider that this latter result holds in general, in the sense that the inefficiency estimated from a dual-level model will always be at least as large as that estimated from a single level model, at least asymptotically for the random effects case. Both of these findings are important in the sphere of economic regulation.

Table 4 Summary of efficiency results

	Dual Level Inefficiency Models			Comparator Models		
				Sub-company inefficiency invariance model		
Firm	Fixed Effects Treatment of μ	Random Effects Treatment of μ		Fixed Effects	Random Effects	Pooled model
Persistent Efficiency Score - External Efficiency						
1	1.000	1.000		1.000	1.000	1.000
2	0.770	0.880		0.770	0.880	1.000
3	0.925	0.840		0.925	0.840	1.000
4	0.617	0.687		0.617	0.687	1.000
5	0.862	0.839		0.862	0.839	1.000
Average	0.835	0.849		0.835	0.849	1.000
Sub-company Varying Efficiency Score - Internal Efficiency						
1	0.621	0.881		1.000	1.000	0.916
2	0.734	0.857		1.000	1.000	0.879
3	0.593	0.819		1.000	1.000	0.779
4	0.853	0.849		1.000	1.000	0.761
5	0.649	0.850		1.000	1.000	0.830
Average	0.690	0.851		1.000	1.000	0.833
Overall Efficiency Score						
1	0.621	0.881		1.000	1.000	0.916
2	0.565	0.754		0.770	0.880	0.879
3	0.549	0.688		0.925	0.840	0.779
4	0.527	0.583		0.617	0.687	0.761
5	0.560	0.713		0.862	0.839	0.830
Average	0.564	0.724		0.835	0.849	0.833

6. Conclusions

This paper has outlined a dual-level inefficiency model that supports the analysis not only of firm inefficiency, but also separates out internal inefficiency from the wider external inefficiency measure. The distinction between external and internal inefficiency is important in any efficiency context, but particularly in the regulatory environment. Economic regulators are interested not only in ensuring that firms match the best practice achieved elsewhere, but also that they consistently apply that best practice across all parts of the organisation.

The models proposed for dealing with sub-company data are re-interpretations and extensions of existing panel data inefficiency models. We have demonstrated via an international dataset of rail infrastructure providers that it is possible to obtain estimates of both components of inefficiency – internal and external inefficiency – and that the dual-level inefficiency model is preferred over the simpler single level alternatives. Indeed, our example shows that failing to account for the dual-level nature of inefficiency may cause overall firm inefficiency to be underestimated; a

result that we consider applies generally (at least asymptotically), and not just in the example presented in this paper.

The use of sub-company data also two important, wider benefits. It substantially increases the number of observations for analysis which addresses a common problem in economic regulation (small N). It would not have been possible to attempt econometric estimation based on just five firms, given the short panel. It is also beneficial for both efficiency performance analysis and more generally cost analysis to analyse data at a level of geographical aggregation that corresponds to how firms organise their activities. This allows both for any dual-level inefficiency to be captured, but also allows for the true scale and density properties of the cost frontier to be established. In this respect we note that the frontier parameter estimates of our model indicate small economies of scale and much stronger economies of density, in line with previous studies utilising disaggregate data. The finding of plausible frontier parameter estimates also gives us confidence in the resulting efficiency findings. The implication of our findings on scale and density are that for this sample there would be cost savings from making maintenance regions bigger (increasing returns to scale). The policy prescription may therefore be for regulatory bodies to press for internal re-organisation, though that decision would need to be assessed against the loss of yardstick information (loss of a region), and should also consider other evidence.

We therefore consider that the approach demonstrated in this paper has wide application in a range of regulatory and other contexts. Most large, regulated companies have some kind of sub-company structure, often based on geographical disaggregation, with some degree of management autonomy at the sub-company level. To our knowledge, the benefits and modelling issues associated with expanding datasets to include sub-company data – including the similarities to and differences from the standard panel case – have not been discussed in the literature.

We note two issues however. Firstly, whilst sub-company data offers some interesting efficiency analysis possibilities, one concern might be that disaggregation increases the degree of noise in the data. Secondly, an additional challenge might be that regulated firms can influence what costs are recorded in each sub-company, although it may be possible for economic regulators to set strict reporting requirements and standards to address this problem. Further empirical analysis of sub-company datasets in a regulatory context, using the models set out in this paper, would therefore be valuable in shedding light on these issues.

Acknowledgements

This work was funded partly by the British Office of Rail Regulation and partly by a part-time PhD scholarship provided by the UK Engineering and Physical Sciences Research Council. We also gratefully acknowledge the contributions of the individual infrastructure managers who provided data and commented on this work, as well as comments on the analysis and assistance with data collection from the British Office of Rail Regulation. Finally, we acknowledge the comments of two anonymous referees. All remaining errors are the responsibility of the authors.

References

AEA Technology (2005). *Recovery of Fixed Costs – Final Report.* A report for the Office of Rail Regulation. Available at http://www.rail-reg.gov.uk/upload/pdf/aea_enviro_rep.pdf [accessed 17/01/2011]

Aigner, D.J., Lovell, C.A.K, and Schmidt, P. (1977), 'Formulation and Estimation of Stochastic Frontier Production Function Models', *Journal of Econometrics*, vol. 6 (1), pp 21-37.

Andersson, M. (2006), 'Marginal railway infrastructure cost estimates in the presence of unobserved effects', *Case study 1.2D I Annex to Deliverable D 3 Marginal cost case studies for road and rail transport, Information Requirements for Monitoring Implementation of Social Marginal Cost Pricing*, EU Sixth Framework Project GRACE (Generalisation of Research on Accounts and Cost Estimation).

Andersson, M. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8 Annex 1A - Rail Cost Allocation for Europe:* Marginal Cost of Railway Infrastructure Wear and Tear for Freight and Passenger Trains in Sweden. Funded by Sixth Framework Programme. VTI, Stockholm.

Battese, G.E. and Coelli, T.J. (1988), 'Prediction of Firm-Level Technical Efficiencies With a Generalised Frontier Production Function and Panel Data', *Journal of Econometrics*, vol. 38, pp 387-399.

Breusch, T. S. and Pagan, A. R. (1980). 'The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics', *Review of Economic Studies* 47, pp. 239-253.

Burns, P. and Weyman-Jones, T.G. (1994), 'Regulatory Incentives, Privatisation and Productivity Growth in UK Electricity Distribution', *CRI Technical Paper,* no. 1, London, CIPFA.
Coelli, T., Rao, D.S.P, O'Donnell, C.J. and Battese, G.E., 2005. An Introduction to Efficiency and Productivity Analysis, 2nd edition. New York, Springer.

Caves, D.W., Christensen, L.R. and Tretheway, M.W. (1984), 'Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ', *The RAND Journal of Economics,* 15 (4) 471-489.

Caves, D.W., Christensen, L.R. and Swanson, J.A. (1981), 'Productivity Growth, Scale Economies, and Capacity Utilisation in U.S. Railroads, 1955-74, *American Economic Review*, vol. 71, issue 5, pp. 994-1002.

Coelli, T., Rao, D.S.P, O'Donnell, C.J. and Battese, G.E., 2005. An Introduction to Efficiency and Productivity Analysis, 2nd edition. New York, Springer.

Cornwell, C., Schmidt, P. and Sickles, R.C. (1990), 'Production Frontiers With Cross-Sectional And Time-Series Variation in Efficiency Levels', *Journal of Econometrics*, vol. 46, pp 185-200.

Gaudry, M., and Quinet, E (2003), *Rail track wear-and-tear costs by traffic class in France,* Universite de Montreal, AJD-66.

Gaudry, M., and Quinet, E. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8* Annex 1Di – Track Maintenance Costs in France. Funded by Sixth Framework Programme. VTI, Stockholm.

Greene, W.H. (2005).'Reconsidering Heterogeneity in Panel data Estimators of the Stochastic Frontier Model,' *Journal of Econometrics* 126, 269-303.

Greene, W. H. (2008). 'The Econometric Approach to Efficiency Analysis.' In Fried, H. O., Lovell, C. A. K., Schmidt, S. S. Eds. *The Measurement of Productive Efficiency and Productivity Growth.* Oxford University Press, New York.

Heshmati, A. and S. Kumbhakar, (1994), 'Farm Heterogeneity and Technical Efficiency: Some Results from Swedish Dairy Farms,' Journal of Productivity Analysis, 5, pp. 45-61.

Johansson, P. and Nilsson, J. (2004), 'An economic analysis of track maintenance costs', *Transport Policy*, 11(3), pp. 277-286.

Jondrow, J., Lovell, C.A.K., Materov, I.S. and Schmidt, P. (1982), 'On Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model', *Journal of Econometrics*, vol. 19, pp. 233-238.

Kennedy, J. and Smith, A.S.J (2004), 'Assessing the Efficient Cost of Sustaining Britain's Rail Network: Perspectives Based on Zonal Comparisons', *Journal of Transport Economics and Policy,* vol. 38 (2), pp. 157-190.

Kumbhakar, S., (1991), "Estimation of Technical Inefficiency in Panel Data Models with Firm- and Time Specific Effects," Economics Letters, 36, pp. 43-48.

Kumbhakar, S., and A. Heshmati, (1995), "Efficiency Measurement in Swedish Dairy Farms 1976-1988 Using Rotating Panel Data," American Journal of Agricultural Economics, 77, pp. 660-674

Kumbhakar, S., and L. Hjalmarsson, (1995), "Labor Use Efficiency in Swedish Social Insurance Offices," Journal of Applied Econometrics, 10, pp. 33-47

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

Lancaster, T. 2000. The Incidental Parameters Problem Since 1948. *Journal of Econometrics*, **95**: 391-414.

LEK (2003), *Regional Benchmarking: Report to Network Rail, ORR and SRA*, London.

Link, H. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8* Annex 1C - Marginal costs of rail maintenance and renewals in Austria. Funded by Sixth Framework Programme. VTI, Stockholm.

Marti, M., Neuenschwander, R. and Walker, P. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8 Annex 1B - Rail Cost Allocation for Europe:* Track maintenance and renewal costs in Switzerland. Funded by Sixth Framework Programme. VTI, Stockholm.

Moulton, B. R. and Randolph, W. C. (1989), 'Alternative tests of the error components model', *Econometrica,* 57, pp. 685-693.

Munduch, G., Pfister, A., Sogner, L. and Stiassny, A. (2002), 'Estimating Marginal Costs for the Austrian Railway System', *Vienna University of Economics Working Paper Series*, no. 78.

Nash, C.A. (1985), 'European Railway Comparisons – What Can we Learn?', in K.J Button and D.E. Pitfield eds*., International Railway Economics*, Aldershot, Gower, pp. 237-269.

NERA (2000), *Review of Overseas Railway Efficiency: A Draft Final Report for the Office of the Rail Regulator*, London.

Neyman, J. and E. Scott. 1948. 'Consistent Estimates Based on Partially Consistent Observations'. *Econometrica* **16**: 1-32.

OFGEM (2003), C*hanges to the Regulation of Gas Distribution to Better Protect Customers*, London.

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

OFWAT (1994), '*Modelling Sewerage Costs 1992-93 - Research into the Impact of Operating Conditions on the Costs of the Sewerage Network: Tables and Figures*'. Report prepared for OFWAT by Professor Mark Stuart, University of Warwick.

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

Orea, C. and S. Kumbhakar, 2004, "Efficiency Measurement Using a Latent Class Stochastic Frontier Model," Empirical Economics, 29, pp. 169-184.

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

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

Smith, A.S.J (2006), 'Are Britain's Railways Costing Too Much? Perspectives Based on TFP Comparisons with British Rail; 1963-2002, *Journal of Transport Economics and Policy*, vol. 40 (1), pp. 1-45.

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

Tervonen, J. and Idstrom, T. (2004), *Marginal Rail Infrastructure Costs in Finland 1997-2002,* Report by the Finnish Rail Administration. Available at www.rhk.fi [accessed 20/07/2005].

Theil, H. (1954). *Linear Aggregation of Economic Relations,* North Holland Publishing Company, Amsterdam.

Train, K., 2003, *Discrete Choice Methods with Simulation*, Cambridge, Cambridge University Press.

Wheat, P. and Nash, C. (2008), *Peer review of Network Rail's indicative charges proposal made as part of its Strategic Business Plan*. Report for the Office of Rail regulation. Available at http://www.rail-reg.gov.uk/upload/pdf/cnslt-ITS_rev-NR_charg-props.pdf [accessed 17/01/2011]

Wheat, P. and Smith, A. (2008), Assessing the Marginal Infrastructure Maintenance Wear and Tear Costs for Britain's Railway Network*, Journal of Transport Economics and Policy, 42(2), 189-224.*

Wheat, P., Smith, A. and C. Nash (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), *Deliverable 8 - Rail Cost Allocation for Europe*. Funded by Sixth Framework Programme. VTI, Stockholm.