Disentangling inefficiency persistence from the latent firm effect: An application to the Norwegian electricity networks

Rahmat Poudineh
Oxford Institute for Energy Studies, Oxford, UK
rahmat.poudineh@oxfordenergy.org

June 8, 2016

Abstract

When measuring efficiency using panel data, the regulators of electricity networks need to make a number of critical assumptions about some factors including the nature of inefficiency (i.e., time-varying and/or time-invariant) and time-invariant effects (i.e., persistent inefficiency and/or the latent firm effect). From an economic perspective, it is conceivable that inefficiency of electricity networks is a combination of transient and persistent effects and time-invariant effects are a mixture of unobserved heterogeneity and persistent inefficiency. However, until recently panel data stochastic frontier models were not able to distinguish among these components. This paper applies the recently developed four-component random error stochastic frontier model to a panel dataset of 142 Norwegian electricity distribution networks observed between 2000-2013. The results show no trace of persistent inefficiency and thus time-invariant effects can be attributed to unobserved heterogeneity in the Norwegian electricity networks. The results also indicate that the Norwegian incentive regulation has been effective in removing rigidities from the operation of firms.

Keywords: Persistent efficiency, unobserved heterogeneity, economic regulation, Norwegian electricity networks

JEL Classification: L43, L51, L94, D21, D23, D24

1 Introduction

Over the last two decades a plethora of applied research have documented the use and outcomes of benchmarking techniques for regulation of natural monopoly infrastructure industries including electricity transmission and distribution. This period has witnessed not only the improvements in the overall practice of regulation of natural monopolies but also benchmarking techniques which evolved so as to better utilise within sector information and more accurately evaluate the relative performance of the firms. However, efficiency results are sensitive to the way they are modelled and interpreted and to the assumption underlying the model when panel data is used (Kumhakar, Lien, and Hardkar, 2014).

An important assumption is related to the nature of firms' inefficiency-i.e., whether it is time-invariant, time varying or both[1]. The early Stochastic Frontier Analysis (SFA) models were giving a one-sided distribution to the heterogeneity effect in traditional panel data models and interpret it as inefficiency (Tsionas and Kumbhakar, 2014). The underlying assumption of these models were that inefficiency is time-invariant something which may be true for short panel datasets but certainly questionable when the behaviour of firms is observed for a long time because, in the long run, firm can re-organise its production process to improve its productivity. Additionally, the persistent inefficiency is unlikely to change without some major changes in the factors that affect management of the firm such as change in the ownership.[2]. On the contrary, if inefficiency is assumed to be time-varying it can change over time without any change in the firm's operation (Badunenko and Kumbhakar, 2016). The assumption in time-varying models is that inefficiency in the current period has nothing to do with inefficiency in the previous period.

[1]In this paper the following terms are used in an interchangeable manner: time-invariant inefficiency and persistent inefficiency, time-varying inefficiency and transient inefficiency, latent firm effect and unobserved heterogeneity.

[2]A firm with persistent inefficiency can not survive in a competitive market. Furthermore, in regulated markets, if the firm can not re-organise its inputs to improve its productivity then the application of incentive regulation is, in essence, useless.

However, it is more conceivable to assume that the firms can remove parts of their inefficiency in the short run while some others sources of inefficiency will remain with the firm over time (Colombi et al., 2011). To illustrate this point, consider an electricity network company which has experienced a permanent demand reduction in its operating area (due to for example, energy efficiency measures or fuel substitution-e.g., between electricity and gas) and thus, the network has excess capacity. In the short run manager might be able to improve the firm's productivity by for example re-organising its inputs, the firm still suffers from persistent inefficiency because of the underutilised assets[3]. This implies that the estimation of both time-varying and persistent inefficiency is important as each component provides different information with different policy implication.

Another important assumption of SFA models pertains to the time-invariant effects and the question whether to interpret them as inefficiency or latent heterogeneity. The problem is that both latent firm effect and inefficiency persistence can co-exist. Thus, if any time-invariant effect is treated as inefficiency it may pick up also unobserved heterogeneity within the persistent inefficiency or even instead of persistent inefficiency. The reverse case holds when time-invariant effects are treated as unobserved heterogeneity. The models that treat time-invariant factors as inefficiency tend to create a downward bias in measuring inefficiency whereas the models that separate the latent firm-effect from time-varying inefficiency are likely to create an upward bias especially if persistent inefficiency exists.

To date, the underlying assumptions of empirical efficiency studies of electricity networks have been mainly in line with above models. However, the layout of electricity network depends on countless factors such as the distribution of generators and load, population density, geographic topography, the attributes

[3]if all firms in the sample experience demand reduction then the relative inefficiency persistence will be low.

and age of the legacy networks' components and various environmental constraints affecting sitting of new lines, transformers and substations (Joskow, 2011). More specifically, the Norwegian electricity networks are considered quite heterogeneous in terms of size, the number of customers, vertical integration, merger/split, geographical conditions, asset age and finally management style. Therefore, presence of heterogeneity and possibility of rigidity in operation of electricity networks cast doubt on the effectiveness of the aforementioned SFA models to provide an accurate measure of regulated firms' efficiency.

In recent years, new SFA models are introduced which allow for disentangling inefficiency persistence from unobserved heterogeneity (Kumhakar, Lien, and Hardkar, 2014; Colombi et al., 2014). The distinction between inefficiency persistence, latent firm effect and time varying inefficiency is important for designing effective incentive regulation models. For example, if the size of inefficiency persistence is relatively large in overall firms' inefficiency, this is an indication that the applied incentive regulation model is not able to remove rigidities in the operation of electricity networks. In this case, regulation may need to incentivise more fundamental changes than just optimising inputs, given outputs (for example, changing management or ownership through merger/split etc.).

This paper estimates inefficiency persistence, the latent firm effect and time-varying inefficiency using a four-component random error SFA model and a panel dataset of 142 Norwegian electricity distribution companies from 2000-2013. We compare the results with two other models in which either persistent inefficiency or latent firm effect is absent. The next section presents methodology whereas dataset used in analysis is discussed in Section 3. The results of estimations and discussions are presented in the section 4. Finally, Section 5 concludes.

2 Methodology

As discussed in Colombi et al. (2014), the earlier panel data SFA models either assume time-invariant inefficiency (for example, Kumbhakar 1987; Battese and Coelli, 1988) or time-varying inefficiency (for example, Battese and Coelli 1992; Lee and Schmidt 1993) without taking into count the firm effect. The later models, however, either confound the firm effect with persistence inefficiency (for example, Kumbhakar and Heshmati, 1995), or firm effect is separated from time-varying inefficiency without taking into account the possibility of persistent inefficiency (for example, Greene, 2005a and b).

The main model used in this research is a four-component random error SFA model based on the works by Kumhakar, Lien, and Hardkar (2014) and Colombi et al. (2014) (hereafter KLH(2014) model). In order to discuss the effects of latent firm effect and inefficiency persistence on overall firms' efficiency we compare the results from KLH (2014) model with Kumbhakar and Heshmati (1995) (hereafter KH (1995) model) and true random effect by Greene (2005b) (hereafter, TRE Greene (2005)).

Greene (2005 a,b) proposed two models named "true fixed effect" (TFE) and "true random effect" (TRE). These models were able to separate firm effect from technical inefficiency. If y represents the cost of firm and x_{it} is the matrix of outputs and input prices, the TRE model can be presented as in (1) (for a cost frontier)

$$y = a_0 + f(x_{it}; \beta) + \mu_i + \nu_{it} + u_{it} \tag{1}$$

where μ_i captures unobserved heterogeneity, (u_{it}) represents time-varying inefficiency and ν_{it} is a random shock with the following distributions:

$$u_{it} \sim N^+(0, \sigma^2_{it}), \quad \nu_{it} \sim N(0, \sigma^2_\nu) \quad \mu_i \sim N^+(0, \sigma^2_\nu). \tag{2}$$

In model (1) a and β are parameters that need to be estimated.

Kumbhakar and Heshmati (1995) introduced a model that separates persistent inefficiency from time-varying inefficiency as follows:

$$y = a_0 + f(x_{it}; \beta) + \nu_{it} + \eta_i + u_{it} \tag{3}$$

where $u_{it} \geq 0$ is time-varying efficiency, $\eta_i \geq 0$ captures persistent inefficiency and ν_{it} is a random shock. In this model overall technical inefficiency is $\eta_i + u_{it}$.

Finally Kumhakar, Lien, and Hardkar (2014) proposed a model that not only distinguishes between time-varying and time-invariant inefficiency but also between inefficiency persistent and the latent firm effect. As shown in equation (4), the error term is split into four components in order to take into account the various factors affecting inputs, given outputs.

$$y = a_0 + f(x_{it}; \beta) + \mu_i + \nu_{it} + \eta_i + u_{it} \tag{4}$$

where μ_i captures unobserved heterogeneity which is disentangled form persistent inefficiency (η_i) whereas u_{it} and ν_{it} are time-varying inefficiency and random shock respectively. a and β are parameters that need to be estimated.

2.1 Model specification and the estimation procedure

We present the aforementioned three models in a stochastic input distance function framework because it can handle multiple inputs and outputs and allows us to measure efficiency of firms in the absence of inputs prices. For this we define the following input requirement set:

$$L_g(q) = x : \text{input x and technology g produce q} \tag{5}$$

where x is the vector of inputs that produces the vector of outputs q using technology g which satisfies the regular axioms of closeness, boundedness, strong disposability and convexity. Given the above input requirement set we define the distance function as follows:

$$D_I(x_{it}, q_{it}, g) = \underset{\lambda}{sup}\ \lambda : x/\lambda \in L_g(q) \geq 1 \tag{6}$$

where λ shows the maximum amount that an input vector can be contracted given output. Technical efficiency then can be defined as in (7).

$$TE(x_{it}, q_{it}, g) = 1/D_I(x_{it}, q_{it}, g) \tag{7}$$

The input distance function has the following properties: homogeneous of degree one, non-decreasing concave function of inputs and non-increasing quasi-concave function of outputs. Linear homogeneity means that all inputs can be normalised with an arbitrary inputs $x_{N_{it}}$ (Galán et al., 2013):

$$1/x_{N_{it}} = D_I(x_{it}/x_{N_{it}}, q_{it}, g)exp(-\omega_{it}) \tag{8}$$

where $\omega_{it} = \ln D_I(x_{it}/x_{N_{it}}, q_{it}, g) \geq 0$ is the total technical inefficiency.

As with respect to technology we use a translog functional form to parametrise distance function. Thus we define $\epsilon_{it} = \ln D_I(x_{it}/x_{N_{it}}, q_{it}, g) - TL(x_{it}, q_{it}, g)$ where $TL(.)$ is the trasnlog function.

This means equation (8) can be written as :

$$y_{it} = TL(x_{it}/x_{N_{it}}, q_{it}, g) + \epsilon_{it} - \omega_{it} \tag{9}$$

where $y_{it} \equiv -\ln x_{N_{it}}$. If any of outputs or inputs are stochastic then ϵ_{it} is random and equation in (9) becomes the standard translog stochastic frontier (Galán et al., 2013). In order to make (9) compatible with models in previous sections we substitute $a_0 + f(x_{it}; \beta) \equiv TL(x_{it}/x_{N_{it}}, q_{it}, g)$ and also split ϵ_{it} and ω_{it} into following components:

$$\begin{aligned}\omega_{it} &= \eta_i + u_{it} \\ \epsilon_{it} &= \mu_i + \nu_{it}\end{aligned} \tag{10}$$

which produces:

$$y_{it} = a_0 + f(x_{it};\beta) + \mu_i + \nu_{it} - \eta_i - u_{it} \tag{11}$$

where, similar to equation (4), μ_i is unobserved heterogeneity, $\eta_i \geq 0$ is persistent inefficiency, $u_{it} \geq 0$ is time-varying inefficiency and finally ν_{it} is a random shock. The negative sign for η_i and u_{it} is due to distance function approach (there is also a negative sign before dependent variable).

Thus the model in (11) is equivalent to KLH (2014) model presented in (4) (but in a distance function framework). Without μ_i the model in (11) reduces to KH (1995) model and without η_i it would coincide with TRE Greene (2005).

In terms of estimation we use two approaches. Due to complexity, a maximum likelihood approach is adopted to estimate TRE Greene (2005) model by making distributional assumption on random error components as in (2) . The KH(1995) and KLH (2014) models can also be estimated in a single stage maximum likelihood procedure based on the distributional assumptions about the components of error term in each model. However, for simplicity we use the multi-step procedure proposed in Kumbhakar, Lien, and Hardkar (2014). An underlying assumption in estimation of these models is that the error components are independent of each other and also independent of the regressors. In what follows we present the multi-step approach for the later two models.

The estimation of KH(1995) model which is presented in (12) can be done in three steps.

$$y_{it} = a_0 + f(x_{it};\beta) + \nu_{it} - \eta_i - u_{it} \tag{12}$$

In order to estimate we rewrite (12) as follows:

$$y = a_0^* + f(x_{it};\beta) + \nu_{it} - u_{it}^* - \eta_i^* \tag{13}$$

where $a_0^* = a_0 - E(\eta_i) - E(u_{it})$, $u_{it}^* = u_{it} - E(u_{it})$ and $\eta_i^* = \eta_i - E(\eta_i)$.

(i) The first step is to estimate a random panel regression of (13) in order to obtain an efficient estimate of β and get predicted values of u_{it}^* and η_i^*. The estimate of η_i^* gives the best linear predictor of random individual effects.

(ii) The second step involves estimating persistent efficiency through predicted value of η_i^*. If $\hat{\eta}_i^*$ is the predicted value of η_i^* persistent technical inefficiency can be estimated from $\hat{\eta}_i = \hat{\eta}_i^* - \mathrm{Min}(\hat{\eta}_i^*)$. Persistent technical efficiency can obtained from $exp(-\hat{\eta}_i)$.

(iii) In the last step we estimate the time-varying efficiency using the residuals from the first step ($y_{it} - f(x_{it};\beta) + \eta_i = a_0 + \nu_{it} - u_{it}$). For this we adopt a standard stochastic frontier approach to estimate $r_{it} = a_0 + \nu_{it} - u_{it}$ assuming $\nu_{it} \sim N(0, \sigma_\nu^2)$ and $u_{it} \sim N^+(0, \sigma_u^2)$ where $r_{it} = y_{it} - f(x_{it};\beta) + \eta_i$. Then we use Jondrow et al.(1982) to obtain time-varying inefficiency, $\hat{u}_{it}$ conditional on estimated residuals $(\nu_{it} - u_{it})$. Finally time-varying efficiency can be calculated from $exp(-\hat{u}_{it})$ and overall technical efficiency from the product of persistent and time-varying efficiencies.

Similar to the previous case we use a three step approach to estimate KLH (2014) model. In order to do this the model in (11) is rewritten as in (14):

$$y = a_0^* + f(x_{it};\beta) + \alpha_i + \epsilon_{it}^* \tag{14}$$

where $a_0^* = a_0 - E(\eta_i) - E(u_{it})$, $\alpha_i = \mu_i - \eta_i + E(\eta_i)$ and $\epsilon_{it}^* = \nu_{it} - u_{it} + E(u_{it})$. In model (14) α_i and ϵ_{it}^* have zero mean and constant variance and can be estimated in three steps as follows:

(i) Similar to the previous case, the first step includes a standard random effect panel regression to estimate β and predict the values of α_i and ϵ_{it}^*

(ii) in the second step, the time varying technical efficiency is estimated using the the predicted value of ϵ^*_{it} from previous step by assuming $\nu_{it} \sim N(0, \sigma^2_\nu)$ and $u_{it} \sim N^+(0, \sigma^2_u)$. For this step we use the standard stochastic frontier technique and Jondrow et al. (1982) estimator to obtain the predicted value of time varying residual technical efficiency $[exp(-u_{it} \mid \epsilon^*_{it})]$.

(iii) in the last step we can estimate η_i using a similar procedure as in the previous step. For this, the best linear predictor of $\alpha_i = \mu_i - \eta_i + E(\eta_i)$ is estimated by assuming $\mu_i \sim N(0, \sigma^2_\mu)$ and $\eta_i \sim N^+(0, \sigma^2_\eta)$ and applying a standard half-normal stochastic frontier model in a cross sectional setting. The persistent technical inefficiency (η_i) is obtained through Jondrow et al.(1982) estimator and persistent technical efficiency is $exp(-\eta_i)$.

3 Data

The dataset used in this analysis comprised an unbalanced panel of 142 Norwegian electricity distribution networks observed between 2000-2013. All financial data are in real term based on 2013 prices. The electricity networks in Norway operate as institutional monopolies and are regulated by Norwegian Water Resource and Energy Directorate (NVE). The Norwegian regulatory model is an ex-ante regulatory framework applies benchmarking technique to evaluate firms' efficiency and set the allowed revenue of firms to be collected through tariff.

Following the Norwegian regulator we use three outputs and one input in our distance function model(Amundsveen, and Marit, 2015). The outputs are number of customers, length of high voltage network in Km, and number of network substations. The input is total cost to the firm including operation and maintenance costs, capital costs (depreciation and return on capital), cost of network energy losses and the value of lost load (VOLL). The summary of descriptive statistics of variables is presented in Table 1. In order to separate exogenous technical progress from technical inefficiency we have added a time trend to our model. The trend also captures those effects that we cannot measure but varies over time and has a common effect on all firms (for example, input prices). All variables are normalised with respect to sample mean.

Table 1: summary statistics of variables

Variable	**Name**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Total cost (000' NOK)	TC	8860292.3	23131099.6	2398.9	374967744
Number of customers	NC	19732.5	51229.435	12	553032
Network length (Km)	NL	738.04	1241.804	9	8744
Number of stations	NS	901.254	1711.2	1	13530

4 Results and discussion

The results of the maximum likelihood estimation for TRE Greene (2005) and the standard random-effect panel regression for the first stages of KH (1995) and KLH (2014) models are presented in Table 2[4]. As seen from the table, most coefficients are highly significant whereas all the first order outputs show the expected sign. The coefficient of first order terms can be interpreted as the elasticities of input with respect to outputs evaluated at sample mean.

The second stage of KH (1995) involves simply computing persistent efficiency using post-estimation results from the first stage. In the third step of KH (1995), however, we estimate time-varying efficiency using a stochastic frontier method and this coincides with the second step of KLH (2014) model. The results of these two estimations have been presented in Table 3. As it can be seen from the Table all parameters are significant at 0.001 significant level. The results for the third step of KLH (2014) (estimating persistent efficiency) is presented in Table 4. It shows that except for variance of latent firm effect the rest of estimated parameters are not significant.

The summary statistics of efficiencies for the three models are summarized in Table 5. As seen from the table, overall mean efficiency from TRE Greene

[4]For the ease of interpretation the coefficients in Table 2 are multiplied by -1.

Table 2: the results for TRE Greene(2005) and KH(1995)& KLH(2014)-first stage

	TRE Greene(2005) $-log(TC)$	KH(1995)& KLH(2014) $-log(TC)$
$log(NC)$	0.662***	0.715***
	(-11.45)	(-13.09)
$log(NL)$	0.213**	0.114
	(-2.77)	(-1.48)
$log(NS)$	0.139*	0.206**
	(-2.06)	(-2.87)
$0.5 * log^2(NC)$	0.150***	0.210***
	(-4.63)	(-5.71)
$0.5 * log^2(NL)$	0.174**	0.297***
	(-2.62)	(-3.83)
$0.5 * log^2(NS)$	0.0609*	0.0765**
	(-2.24)	(-2.73)
$log(NC) * log(NL)$	-0.0265	-0.158**
	(0.53)	(2.80)
$log(NC) * log(NS)$	0.0208	0.0601
	(-0.71)	(-1.78)
$log(NL) * log(NS)$	-0.104**	-0.155***
	(3.18)	(3.94)
t	0.168***	0.168***
	(-20.74)	(-20.41)
$0.5 * t^2$	-0.0164***	-0.0165***
	(13.69)	(13.77)
$Constant$	-0.622***	-0.848***
	(10.88)	(19.36)
σ_u	0.396***	0.296
	(22.23)	
σ_v	0.191***	-
	(18.14)	
σ_e	-	0.308
λ	2.074 ***	-
	(78.46)	
ρ	-	0.481

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: the results for third step of KH(1995) and second step of KLH(2014)

	KH(1995) third stage	KLH(2014) second stage
frontier		
a_0	-0.275***	-0.275***
	(-19.88)	(-19.88)
usigmas		
Constant	-2.092***	-2.092***
	(-22.22)	(-22.22)
vsigmas		
Constant	-3.170***	-3.170***
	(-37.57)	(-37.57)
Observations	1679	1679

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: the results for third stage of KLH(2014)

	KLH(2014) third stage
frontier	
$E(\eta_i)$	0.0134
	(0.18)
usigmas	
Constant	-15.28
	(-0.04)
vsigmas	
Constant	-2.598***
	(-75.26)
Observations	1679

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(2005) is significantly higher than KH (1995) (around 77 percentage point) and moderately higher than KLH (2014) model. The persistent efficiency captured by KLH (2014) and KH (1995) models are on average 0.99 and 0.13 respectively which differ sharply as opposed to their level of residual efficiencies that coincides. In order to investigate the non existence of time-invariant inefficiency in the KLH (2014) model we run a likelihood-ratio (LR) test with the null hypothesis of $\sigma_\eta = 0$. The LR which has a chi-squared distribution with one degree of freedom, has a magnitude of -8.412×10^{-7}. Using the critical values in Table 6 we can not reject the null hypothesis of zero inefficiency persistence in KLH (2014) model. However, residual inefficiency exists in this model. The LR test for the presence of residual inefficiency with the the null hypothesis of $\sigma_u = 0$ has one degree of freedom and is equal to 104.50. Thus we can strongly reject the null hypothesis (at 0.001 significance level) given the critical values in Table 6.

In order to explore the effect of three estimated models on ranking order of firms' efficiency we estimate Kendall's rank correlation coefficient. The results are presented in Table 7. As it is shown in the table, KH(1995) model has a very high ranking disagreement with TRE Greene (2005). This result is not surprising given the assumption of these two models, with respect to time-invariant effects, contradicts each other. The rank correlation coefficient for KH (1995) and KLH (2014) models is positive but low at 0.33. The KLH(2014) and TRE Greene (2005) however have moderate negative rank correlation coefficient. In a nutshell, the ranking order of technical efficiency estimated in three models varies significantly.

Table 5: Summary statistics of efficiencies

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
TRE Greene (2005)	0.906	0.076	0.313	0.993
Residual effciency	0.775	0.101	0.031	0.953
Persistent effciency	0.171	0.064	0.093	1
Total efficiency KH (1995)	0.133	0.058	0.004	0.929
Residual efficiency	0.775	0.101	0.031	0.953
Persistent efficiency	0.999	0	0.999	0.999
Total efficiency KLH (2014)	0.775	0.101	0.031	0.953

Table 6: critical values of the mixed chi-square distribution

dof	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.455	1.642	2.705	3.841	5.412	6.635	9.5

source: Table 1 in Kodde and Palm (1986, Econometrica)

Table 7: Kendall's rank correlation coefficient

	TRE Greene (2005)	KH (1995)	KLH (2014)
TRE Greene (2005)	1		
KH (1995)	-0.8046*	1	
KLH (2014)	-0.4521*	0.3330*	1

Note * p<0.05

The differences and similarities between the efficiencies estimated by above three models can be explained by the their underlying assumptions. In fact the KH (1995) model confounds inefficiency persistence with unobserved heterogeneity and thus part of inefficiency persistence captured by this model is in fact latent firm effect. On the other hand, the model by KLH (2014) shows a very low inefficiency persistence (close to 100% efficiency persistence) and this explains why the overall efficiency results from KLH (2014) is closer to TRE Greene (2005) but very dissimilar to KH (1995).

Figure 1 presents distribution of overall technical efficiency across three models. As seen, efficiencies estimated by KLH (20140) model lies between the other two models and the latent firm effect is the main cause of this. It is hard to interpret what constitute the latent firm effect in electricity networks because it can be anything. The only thing we know is that unobserved heterogeneity affects the cost of firms (positively or negatively) and controlling for it has been the main motivation behind the panel data models. Therefore, unless there is no unobserved heterogeneity among the benchmarked units, models such as KH (1995) produces biased estimates of efficiency.

Distribution of components of technical efficiency, presented in the figure two, shows that time-varying efficiency, as the main element of overall technical efficiency in the Norwegian electricity networks, varies considerably among the

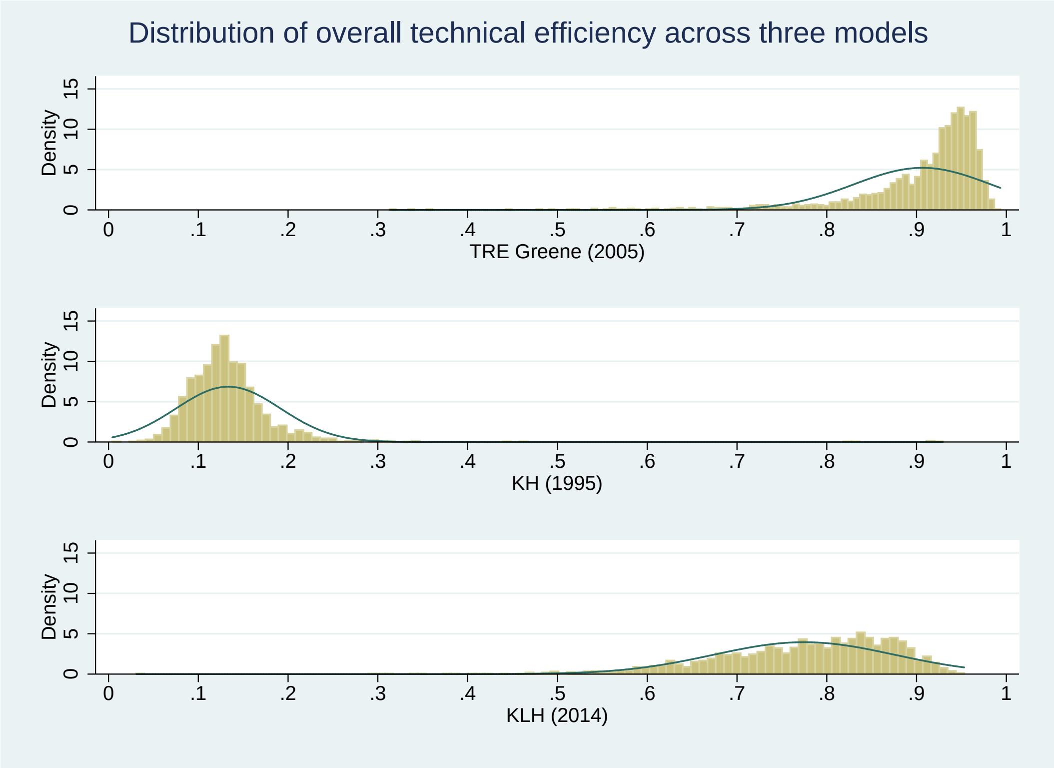

Figure 1: Distribution of overall techical efficiency across three models

firms. However, a low time-varying efficiency is not often as serious as a low long run efficiency because transient inefficiency can be caused by factors that may not be repeated in future. On the other hand, since we find no trace of inefficiency persistence, when it is disentangled from the latent firm effect in KLH(2014) model, we can argue that the Norwegian regulatory regime has been effective in removing rigidities from the operation of the electricity networks.

The fact that the time-invariant effects, in the Norwegian network utilities, are entirely unobserved heterogeneity is compatible with the characteristics and operating environment of these networks. Norway has 147 electricity distribution companies which serve a population of around 5.1 million[5] (Amundsveen and Marit, 2015). The heterogeneity among these companies is such that the Norwegian regulator excludes some of the extreme observations from its benchmarking model.

[5]For comparison, the UK has 14 electricity distribution companies which provide service for 60 million population.

According to descriptive statistics presented in Table 1, customers of electricity network companies in Norway ranges from 12 to more than 550,000, the length of electricity networks ranges from 9Km to more than 8700Km, and the number of stations are between one to around 13500. Given these huge variation among the firms, it is very likely that there are some specifications that can not be observed but have an impact on the cost and consequently productivity of the network companies.

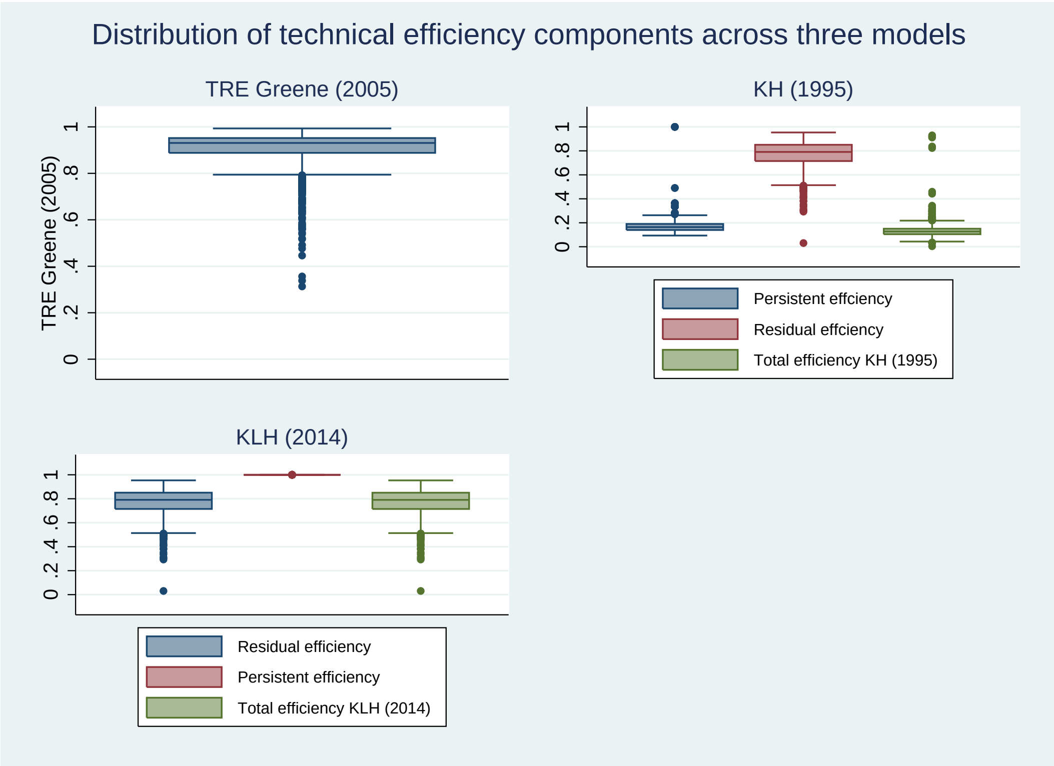

Figure 2: Distribution of technical efficiency components across three models

Figure three shows the evolution of mean efficiency and its components over the 14 observed year. The graphs show no discernible pattern for the efficiencies estimated. Time-varying efficiency in KH (1995) model coincides with total technical efficiency in KLH (2014) model which is not surprising given the non-existence of persistent inefficiency. The average efficiency in TRE Greene (2005) model shows constantly a higher value compared with KLH (2014) over the whole period.

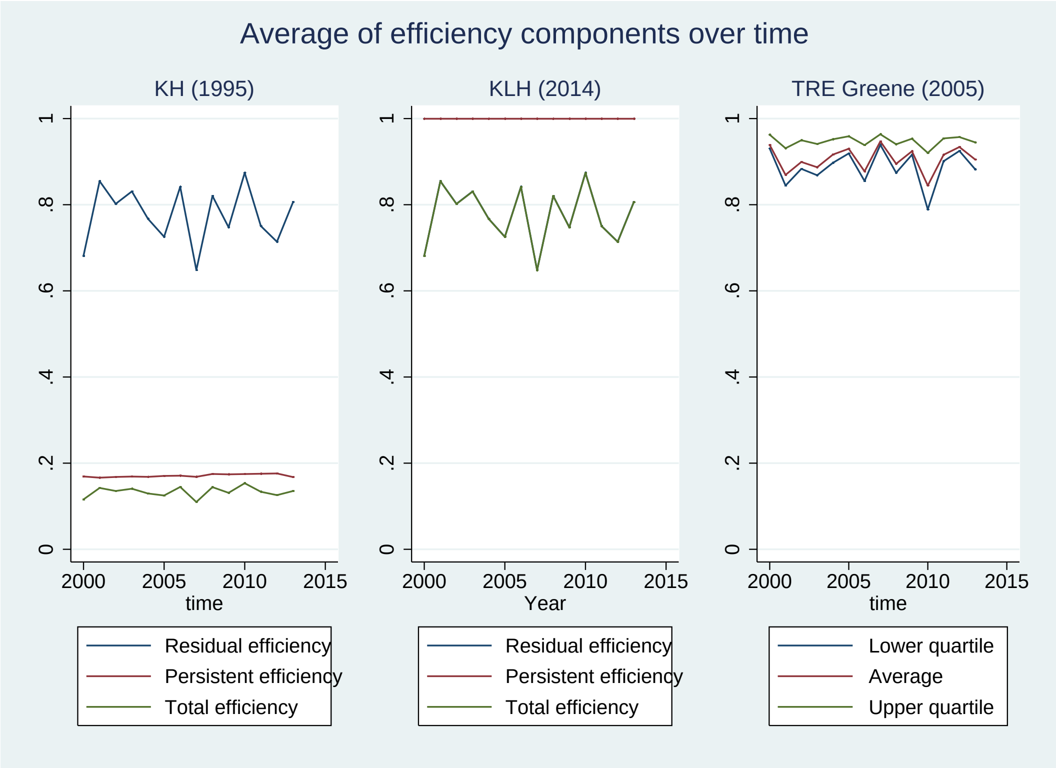

Figure 3: Variation of average efficiency over years across models

The results of our analysis in this paper provides some insights on the use of statistical benchmarking techniques for incentive regulation of electricity network companies. One of the main aim of the incentive regulation is to mimic the outcome of a competitive market condition where firms with persistent inefficiency cannot survive in the long run. In regulated environment, persistent inefficiency, where exists, is difficult to address as, unlike the short run inefficiency, it requires major regulatory and policy change (e.g, a regulation that leads to change in the ownership) beyond the conventional penalty and reward schemes . The absence of inefficiency persistence suggest that regulatory model has prevented the firms from systematic under-performance.

Furthermore, as the Norwegian incentive regulation model is based on penalty and reward scheme it is important that inefficiency persistence is disentangled from the latent firm effect because it avoids unwarranted application of incentives and penalties. This in turn highlights the importance of modelling, underlying assumptions and the interpretation of results when efficiency of electricity networks are estimated.

In summary, the following points can be concluded about disentangling the latent firm effect from persistent inefficiency.

(i) If unobserved heterogeneity constitutes a significant part of the convoluted time-invariant effect, the models that disentangle the firm effect from time-varying efficiency-such as TRE Greene (2005)-would produce results close to four-component random error in KLH (2014). In this case the model of KH (1995) underestimates efficiency.

(ii) If inefficiency persistence constitutes a significant part of the convoluted time-invariant effect, the model that disentangle inefficiency persistence from time-varying efficiency- such as KH (1995)- produce results very similar to KLH (2004). In this case the model that does not account for inefficiency persistence such as TFE Greene (2005) overestimates efficiency.

(iii) If there are moderate inefficiency persistence and unobserved heterogeneity among the firms, both KH (1995) and TFE Greene (2005) would produce biased results. Under certain statistical conditions (explained below) KLH (2014) can produce more reliable results in this case.

One important question is that whether KLH (2014) model can reliably disentangles inefficiency persistence from the latent firm effect (and also time-varying inefficiency from noise). In a recent paper by Badunenko and Kumbhakar (2016) it is shown that the reliability of transient and persistent inefficiencies estimated by four-component random error model critically depends on the following three estimated parameters: $\lambda_0 = \frac{\sigma_\eta}{\sigma_\mu}$, $\lambda_1 = \frac{\sigma_u}{\sigma_\nu}$ and $\Lambda = \frac{\sigma_\eta}{\sigma_u}$. According to them for a reliable estimation of persistent and transient efficiency λ_0 and λ_1 must be large respectively (i.e., $\lambda_0 \geq 1$ and $\lambda_1 \geq 1$). This implies that the reliability of KLH (2014) estimator is contingent upon low levels of noise and unobserved heterogeneity something which is very unlikely with real world datasets. In our model with Norwegian dataset $\lambda_0 = 0.002$, $\lambda_1 = 2.93$ and $\Lambda = 0.00$ which suggest that the estimation of transient efficiency is fairly reliable while the reliability of persistent efficiency can not be ensured.

The implication of these results is significant for benchmarking of network utilities because from the regulator's point of view it is not clear *a priori* whether there is inefficiency persistence and/or unobserved heterogeneity in the dataset. Therefore, the selection of appropriate model for estimating efficiency and interpretation of the results may not be straightforward. Although, the four-components random error model by KLH (2014) provides a comprehensive framework where most other panel data SFA models can be considered as special cases, a reliable estimation of inefficiency components can not be ensured under all conditions.

5 Conclusions

The estimation of relative efficiency of regulated energy infrastructures, using panel datasets, entails critical assumptions about the nature of inefficiency and time-invariant effects. The regulators need to form an opinion whether inefficiency is time varying and/or time-invariant and whether time-invariant firm effects are unobserved heterogeneity an/or inefficiency persistence. The panel data SFA estimators can be classified based on the their underlying assumption about aforementioned factors and until recently there was no single estimator that could distinguish among all these components of error term.

However, there is economic rationale in thinking of electricity networks' inefficiency as a combination of transient and persistent factors and also treating time-invariant effects as a tangled term of inefficiency persistence and unobserved heterogeneity. The distinction between these components provides valuable information about the effectiveness of regulatory model and also it avoids the unjustified application of penalty and reward under incentive regulation. This paper applied a four-component random error SFA model to a panel dataset of 142 Norwegian electricity distribution companies observed between 2000-2013. We compared the results with two other models: one that assumes time-invariant effects are entirely inefficiency persistence and another model which treats them as unobserved heterogeneity.

We find no traces of inefficiency persistence when it is disentangled from the latent firm effect. This indicates that the Norwegian regulatory regime has successfully removed rigidities in the operation of electricity networks. However, the results show that transient inefficiency exists and it is fairly high. The results also indicate that, due to presence of heterogeneity among the Norwegian electricity networks, the model that treats time-invariant effects as persistent inefficiency severely underestimates efficiency of the firms. On the other hand, due to absence of persistent inefficiency, the model that treated time-invariant effects as unobserved heterogeneity produced a results similar to four-component random error model although it is expected that it will overestimate efficiency where persistent inefficiency exists.

Although the four-component random error model provide a comprehensive framework where many of other panel data SFA models can be seen as special cases, its ability to produce reliable estimation of both forms of technical efficiency (transient and persistent) critically depends on the ratio of variance parameters. The results of our analysis show the ratio of variance parameter of inefficiency persistence to latent firm effect is very low. On the other hand, the ratio of variance parameter of transient inefficiency to noise is fairly high. This implies that while the estimation of transient efficiency is reasonably reliable, the reliability of persistent efficiency estimation can not be ensured.

References

[1] Amundsveen, R. and Marit, H.K (2015), The Development and Application of an Incentive Regulation Model – a Balancing Act. The ICER Chronicle, International Confederation of Energy Regulator, edition 3, March.

[2] Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38:387–399.

[3] Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis* 3:153–169.

[4] Badunenko,O. and Kumbhakar, SC. (2016) When, Where and How to Estimate Persistent and Transient Efficiency in Stochastic Frontier Panel Data Models, *European Journal of Operational Research*, doi: 10.1016/j.ejor.2016.04.049.

[5] Colombi, R. , Kumbhakar, SC., Martini, G. and Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis* 42:123–136.

[6] Colombi, R., Martini,G. and Vittadini,G. (2011) "A stochastic frontier model with short-run and longrun inefficiency random effects," Working Papers, Department of Economics and Technology Management, University of Bergamo.

[7] Galán, J., Veiga, H., Wiper, M. (2013). Bayesian analysis of dynamic effects in inefficiency: Evidence from the Colombian banking sector. Statistics and Econometrics Series 18 WP13-19, Universidad Carlos III de Madrid.

[8] Greene W (2005a) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126:269–303.

[9] Greene W (2005b) Fixed and random effects in stochastic frontier models. *Journal of Productivity Anal* 23:7–32.

[10] Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in stochastic frontier production function model. *Journal of Econometrics* 19:233–238.

[11] Joskow, P.L. (2011), Incentive regulation in theory and practice – Electricity distribution and transmission networks, NBER Chapters, in: Economic Regulation and Its Reform: What Have We Learned?, National Bureau of Economic Research, Inc.

[12] Kodde, D.A., and Palm F.C (1986) Wald criteria for jointly testing equality and inequality restrictions, *Econometrica* 54, 1243–1248.

[13] Kumbhakar SC (1987) The specification of technical and allocative inefficiency in stochastic production and profit frontiers. *Journal of Econometrics* 34:335–348.

[14] Kumbhakar SC, Heshmati A (1995) Efficiency measurement in Swedish dairy farms: an application of rotating panel data, 1976–1988. *American Journal of Agricultural Economics* 77:660–674.

[15] Kumbhakar , SC., Lien, G. and Hardaker, J.B. (2014). Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis* 41:321–337.

[16] Lee Y, Schmidt P (1993) A production frontier model with flexible temporal variation in technical inefficiency. In: Fried H, Lovell K (eds) The measurement of productive efficiency: techniques and applications, Oxford University Press, New York.

[17] Tsionas, E. G. and Kumbhakar, S. C. (2014). Firm Heterogeneity, Persistent and transient Technical Inefficiency: A generalized True Random Effects model. Journal of Applied Econometrics, 29: 110–132.