What is the Best Practice for Benchmark Regulation of Electricity Distribution? Comparison of DEA, SFA and StoNED Methods

Timo Kuosmanen[1], Antti Saastamoinen[1], Timo Sipiläinen[2]

1) School of Economics, Aalto University, 00100 Helsinki, Finland. E-mail. Firstname.Lastname@aalto.fi.
2) Department of Economics and Management, University of Helsinki, Finland. E-mail. Firstname.Lastname@helsinki.fi.

ABSTRACT

Electricity distribution is a natural local monopoly. In many countries, the regulators of this sector apply frontier methods such as data envelopment analysis (DEA) or stochastic frontier analysis (SFA) to estimate efficient cost of operation. In Finland, a new StoNED method was adopted in 2012. This paper compares DEA, SFA and StoNED in the context of regulating electricity distribution. Using data from Finland, we compare the impacts of methodological choices on cost efficiency estimates and acceptable cost. While the efficiency estimates are highly correlated, the cost targets reveal major differences. In addition, we examine performance of the methods by Monte Carlo simulations. We calibrate the data generation process (DGP) to closely match the empirical data and the model specification of the regulator. We find that the StoNED estimator yields the root mean squared error (RMSE) of 4% with the sample size of 100. Precision improves as the sample size increases. The DEA estimator yields the RMSE of approximately 10%, but performance deteriorates as the sample size increases. The SFA estimator has RMSE of 144%. Poor performance of SFA is due to the wrong functional form and multicollinearity.

Key words: *Frontier estimation; nonparametric production analysis; productive efficiency*

1. Introduction

Electricity distribution firms typically enjoy a natural local monopoly. This creates a need to regulate the distribution sector. In the theory of regulation, it is well-known that the 'cost-of-service' type of pricing does not provide incentives for the electricity distribution firms to minimize the cost (see, e.g., Laffont and Tirole, 1993). To determine a more objective yardstick for the acceptable cost level, Shleifer (1985) suggested comparing the observed cost of a firm with that of its competitors. However, as Pollit (2005) points out, it is often difficult to find exactly identical or even sufficiently similar competitor that could serve as an appropriate yardstick. Instead of using a discrete set of benchmark firms, one could apply frontier estimation methods to estimate a continuous frontier cost function that represents the best practice benchmark. Today, benchmark regulation is applied as an integral part of the regulatory framework in many countries (see, e.g., Jamasb and Pollit, 2001). According to the recent study by Bogetoft and Otto (2011, Ch. 10), at least 9 European regulators currently apply the axiomatic DEA (*Data Envelopment Analysis,* Charnes et al., 1978; Farrell, 1957) and the econometric SFA (*Stochastic Frontier Analysis*, Aigner et al., 1977), or some combination thereof.

Ever since the DEA and SFA approaches have been introduced to regulation, there has been lively debate about the suitability of these methods for the purposes of regulation (e.g., Dassler et al., 2006; Irastorza, 2003). There is a large and growing academic literature on the application of DEA and SFA in the electricity distribution industry (e.g., Agrell et al., 2005; Cullmann, 2009; Forsund and Kittelsen, 1998; Hjalmarsson and Veiderpass, 1992; Iglesias et al. 2010; Jamasb and Pollit, 2003; Kopsakangas-Savolainen and Svento, 2008; Korhonen and Syrjänen, 2003; and Weyman-Jones, 1991). As yet, however, there is no clear conclusion on which method is superior. The inconclusive results have raised concerns about the suitability of any single method for the purposes of benchmark regulation. Thus, many regulators have recently opted to use a combination of both DEA and SFA (see also Azadeh et al., 2009). In Germany, for example, the regulator estimates efficiency of each firm using both DEA and SFA, and then chooses the larger of the two estimates (e.g. Agrell and Bogetoft, 2007) According to Bogetoft and Otto (2011; Ch. 10), at least four European regulators apply some combination of both DEA and SFA.

The Finnish Energy Market Authority (*Energiamarkkinavirasto*, EMV) is one of the pioneers in the practical implementation of benchmark regulation. EMV has used frontier methods as an integral part of the regulatory model since 2005, starting with DEA (Korhonen and Syrjänen, 2003), adopting SFA in 2008 (Syrjänen et al., 2006).[1] In 2010, EMV commissioned several studies to address the critique of DEA and SFA presented by the distribution firms and the energy industry. After a rigorous evaluation process, EMV considered the report by Kuosmanen et al. (2010, *in Finnish*) as the most promising attempt to overcome the pitfalls of DEA and SFA. Following the recommendation of that report, in 2012 EMV replaced DEA and SFA by the new StoNED method (*Stochastic Semi-Nonparametric Envelopment of Data*; Kuosmanen and Kortelainen, 2012). The main appeal of StoNED is its ability to accommodate the main advantages of both DEA and SFA: StoNED combines the non-parametric, piece-wise linear DEA-style frontier with the stochastic SFA-style treatment of inefficiency and noise. In fact, both DEA and SFA can be obtained as constrained special cases of the more general StoNED-model (Kuosmanen and Johnson, 2010; Kuosmanen and Kortelainen, 2012). The less restrictive assumptions of the StoNED method imply a wider range of applicability, making StoNED more robust to both model misspecification and noise.[2]

The recent reform of the Finnish regulatory model of electricity distribution has attracted a lot of interest in other countries and in other network industries (e.g., gas,

[1] Further information about the Finnish regulatory model can be found on the EMV website: www.emvi.fi. Kinnunen (2006) reviews the EMV model from the perspective of investment incentives.

[2] Previous published applications of the StoNED method are in the areas of agriculture (Kuosmanen and Kuosmanen, 2009), electricity generation (Mekaroonreung and Johnson, 2012) and electricity distribution (Kuosmanen, 2012).

utilities, telecommunication), both among academic scholars and practitioners. Kuosmanen (2012) describes the semi-nonparametric cost frontier model adopted by EMV, its estimation by the StoNED method, the econometric specification tests that have been considered, and the main empirical results. The purpose of this paper is to complement Kuosmanen (2012) by presenting a systematic comparison of the DEA, SFA, and StoNED-model in the context of regulation. Focusing on the model specifications actually employed by EMV, we compare the efficiency estimates produced by these three different methods and the average of DEA and SFA, and examine factors that explain the observed differences. More importantly, we also compare the implications of the methodological choices on the monetary cost targets. While the efficiency scores obtained with different methods are usually highly correlated, the economic implications in terms of the cost targets are substantial.

The empirical comparisons show that the choice of the benchmarking method matters in practice. However, empirical comparisons do not allow us to conclude that one method is better than another. Therefore, we also compare the precision of the estimators in the controlled environment of Monte Carlo simulations. A novel feature of our simulations is that we calibrate the data generation process of the simulations to match the essential characteristics of the EMV data as closely as possible to ensure the relevance of the simulation evidence for the real-world regulation. We draw random samples of pseudo-firms from the piece-wise linear cost frontier that the EMV is currently using as a part of its regulatory model. Further, we take into account the heavily skewed distribution of the firm size and the high positive correlation between the output variables in our random sampling procedure. The customized data generation process of the simulations enables us to measure performance of the alternative estimators in the specific context of the EMV's regulatory framework. Our simulation evidence shows that the StoNED method is superior to the conventional DEA, SFA, and their average at all sample sizes considered.

The rest of the paper is organized as follows. Section 2 describes general cost frontier estimation framework and describes the empirical data. Section 3 introduces the benchmarking methods considered in this study, and presents some empirical results obtained with each method. The main focus of this study is on the comparative assessment, to be presented in Sections 4 and 5. Section 4 presents an empirical comparison of the methods using the data and model specifications of EMV. Section 5 presents a systematic comparison of the methods in the controlled environment of Monte Carlo simulations. In the Section 6 we briefly comment the implementation of the methods in the EMV regulatory model. Section 7 concludes.

2. Cost frontier model and data

The cost regulation of EMV is based on the following generic model of cost frontier (see Kuosmanen, 2012, for a more detailed discussion)

$$\ln x = \ln C(y_1, y_2, y_3) + \delta z + u + v \qquad (1)$$

where

x is the observed total cost (TOTEX) (1,000 €)
C is the frontier cost function
y_1 is the energy transmission (GWh)
y_2 is the total length of the network (km)
y_3 is the number of customers
z is the proportion of underground cables
δ is the coefficient of the z variable
u is the random variable representing inefficiency
v is the random variable representing stochastic noise

In this study the cost variable x refers to the total expenditure (TOTEX), which consists of three components: controllable operational costs (OPEX), capital expenditures (CAPEX) and the external supply interruption costs for customers (INT).[3] The last component can also be viewed as a quality component, as the lack of supply interventions can be interpreted as an indicator of good service. Since the outputs are almost time-invariant throughout the period, all variables are defined as the yearly averages over the period 2005-2008 (see Kuosmanen, 2012, for further discussion). Before averaging, the total costs are deflated to the prices of 2005. In this specification, inefficiency u represents the average inefficiency over the evaluation period. Averaging of data also reduces the variance of the noise term v.

The output variables are the weighted amount of energy transmitted through the network (y_1, GWh of 0.4 kV equivalents), the total length of the network (y_2, km), and the total number of customers connected to the network (y_3, number). In y_1, the transmission of electricity at different voltage levels is weighted according to the average cost of transmission such that the high-voltage transmission gets a lower weight than the low-voltage transmission. We stress that y_1 depends on the observed demand for electricity, whereas outputs y_2 and y_3 capture the potential or latent demand (see Kuosmanen 2012, for a more detailed discussion). In essence, outputs y_2 and y_3 capture the fixed cost of maintaining a sufficient capacity to provide service for the given network area irrespective of the actual consumption of electricity.

In addition to the three outputs, the latest EMV specification introduced a contextual variable z, defined here as the proportion of underground cables in the total length of the network. The z-variable is not an input or output as such; it can be interpreted as a control for the heterogeneity of the firms and their operating environments. Note that the contextual variable enters model (1) in a parametric form, analogous to the standard regression analysis, while the output variables are modeled using a nonparametric

[3] Our empirical comparison is based on the original data and the model specification recommended in Kuosmanen et al. (2010) and Kuosmanen (2012). EMV has made some subsequent modifications to the model and the data. The distribution firms have challenged the EMV model, and the final model specification is subject to the ruling of the Finnish Market Court.

specification of the cost function C. Since we are interested in both the parametric and nonparametric parts of the model, it is appropriate to characterize model (1) as a semi-nonparametric, partially linear model of cost frontier. Note that the parametric formulation of the contextual variable is analogous to the conventional two-stage DEA approach of regressing DEA efficiency scores on contextual variables z (e.g., Simar and Wilson, 2007; Banker and Natarajan, 2008; Johnson and Kuosmanen, 2011, 2012). The parametric model of contextual variables allows us to capture the average effect of underground cabling on cost (represented by the coefficient δ), without increasing the number of explanatory variables included in the nonparametric part (subject to the curse of dimensionality).

Our data consists of 89 Finnish electricity distribution companies, whose networks cover practically all regions of Finland. Table 2.1 presents the descriptive statistics for total costs, three outputs, and the underground cabling variable, which describes the operational conditions of a company (see Section 3.4 for details).

Table 2.1: Descriptive statistics of variables

Variable	Mean	St. Dev.	Min.	Max.
x = Total cost (1,000 €)	8 418.91	18 047.78	267.81	117 554.10
y_1 =Energy transmission (GWh)	480.39	971.51	14.81	6 599.71
y_2 =Length of network (km)	4 135.27	10 223.27	50.80	67 611.05
y_3 =No. customers	35 448.68	71 870.65	24.25	420 473.00
z = Proportion of underground cables	0.33	0.26	0.01	1.00

Table 2.1 reveals that the industry consists of a very heterogeneous set of firms. For example, the size of companies measured by the amount of transmitted energy varies from 15 to 6,600 GWh per year. The operating environments of companies also substantially differ. On average the proportion of underground cabling is 33% but the range is practically from 1% to 100%. The proportion of underground cabling is highest in the dense urban areas. Note that the data also includes some industrial network operators, which transmit a large amount of energy to a small number of industrial customers.

3. Frontier estimators

This section introduces the estimation approaches for the cost frontier and cost efficiency of electricity distribution units. We start with the most general StoNED estimator. The conventional DEA and SFA estimators are then introduced as special cases of StoNED. For a more detailed introduction to DEA and SFA, see e.g. Fried et al. (2008). Lastly, the naïve model averaging of DEA and SFA is discussed. In Sections 3.1 - 3.3 we present some characteristics of the estimated cost frontiers in terms of the marginal costs, while the detailed comparison of efficiency estimates will be provided in Section 4. The purpose of presenting some empirical results already in this section is to shed some empirical light to the methodological discussion: the marginal cost analysis partly helps us to identify the underlying factors behind the differences between the methods.

3.1. StoNED

The StoNED estimator combines the axiomatic, non-parametric frontier (the DEA aspect) with a stochastic noise term (the SFA aspect). Thus it can be seen as a more general estimation framework. StoNED takes stochastic noise into account similar to SFA. However, StoNED does not require any *a priori* assumptions about the functional form of the cost frontier. Like in DEA, the StoNED model is based on some general axioms (or regularity conditions) concerning the benchmark technology.[4] The set of axioms imposed in the EMV model is the following:

1) *C* is monotonic increasing in all outputs
2) *C* is globally convex in outputs
3) *C* exhibits constant returns to scale (CRS)

[4] The term benchmark technology refers to the frontier used as a reference for the observed firms in productivity and efficiency comparisons. The axioms of the benchmark technology represent our ex ante requirements for efficient performance (e.g., monotonicity stems from the definition of technical efficiency by Koopmans, 1951). The underlying production technology does not necessarily satisfy these axioms.

The first two conditions are standard properties in DEA. The third axiom could easily be relaxed. However, the CRS axiom could not be rejected in the empirical specification test reported by Kuosmanen (2012). More importantly, the CRS axiom is preferable from the regulatory point of view, as the benchmark technology exhibits the same level of total factor productivity irrespective of the firm size. For example, suppose firms enjoy economies of scale in reality. The CRS axiom of the regulatory model then provides an incentive for firms to seek productivity improvement through mergers. Such an incentive would be lost if the CRS axiom were relaxed and variable returns to scale (VRS) were imposed. Indeed, the use of the VRS benchmark may give wrong incentives for firms to split or merge for strategic reasons to game the regulator.

The cost frontier model (1) can be estimated with convex nonparametric least squares (CNLS: Johnson and Kuosmanen, 2011, 2012; Kuosmanen, 2008). Denoting the composite error term by $\varepsilon_i = u_i + v_i$, the CNLS problem is stated as

$$\min_{\phi,\beta,\delta,\varepsilon} \sum_{i=1}^{n} \varepsilon_i^2$$

$$\text{s.t.} \tag{2}$$

$$\begin{aligned}
&\ln x_i = \ln \gamma_i + \delta z_i + \varepsilon_i \quad \forall i \\
&\gamma_i = \beta_{1i} y_{1i} + \beta_{2i} y_{2i} + \beta_{3i} y_{3i} \quad \forall i \\
&\gamma_i \geq \beta_{1h} y_{1i} + \beta_{2h} y_{2i} + \beta_{3h} y_{3i} \quad \forall h, i \\
&\beta_{ki} \geq 0 \;\; \forall k = 1,2,3; \forall i
\end{aligned}$$

The beta coefficients represent the marginal costs of outputs (shadow prices). Alternatively, these coefficients can be interpreted as the slopes of the tangent hyperplanes to the piece-wise linear cost frontier. These coefficients are directly analogous to the multiplier weights in DEA. Note that the coefficients $(\beta_{1i}, \beta_{2i}, \beta_{3i})$ are firm-specific. This allows for greater heterogeneity of distribution networks than the usual parametric approaches (cf., e.g., Cullmann, 2012). For example, urban distribution networks with lots of customers are assigned a higher marginal cost for output y_3 than rural networks, for which the network length (output y_2) is the main cost driver. The contextual variable z also captures heterogeneity of firms. In contrast to the two-stage DEA estimation (Simar and Wilson, 2007), the contextual variable enters directly the first constraint (the regression equation) of the CNLS problem (2) (see Johnson and Kuosmanen, 2011, 2012).

In the second stage we impose some distributional assumptions on the terms u_i and v_i in order to distinguish inefficiency from noise and to estimate μ based on the CNLS residuals $\hat{\varepsilon}_i$. We maintain the usual assumptions of the SFA literature: we assume u_i has a half-normal distribution such that $u_i \geq 0$, and v_i has a normal distribution with zero-mean and a finite constant variance. Using the method of moments (see Kuosmanen and Kortelainen, 2012, for details), the parameter estimates $\hat{\sigma}_u, \hat{\sigma}_v$ (i.e., the standard deviations of inefficiency and noise, respectively) can be obtained based on the sample variance and skewness of the CNLS residuals $\hat{\varepsilon}_i$. Given the half-normal distribution of u, the expected inefficiency is $E(u_i) = \mu = \hat{\sigma}_u \sqrt{2/\pi}$ (assumed constant across firms).

The optimal γ_i from problem (2) is a consistent estimator of the total cost x_i, conditional on outputs (y_{1i}, y_{2i}, y_{3i}), that is,

$$E(x_i \,|\, y_{1i}, y_{2i}, y_{3i}) = C(y_{1i}, y_{2i}, y_{3i}) \times \exp(\mu), \tag{3}$$

To estimate the frontier cost function, we must adjust the estimated γ_i with μ. Thus, the StoNED cost frontier is obtained by adjusting the estimated γ_i downward according to

$$\hat{C}^{StoNED}(y_{1i}, y_{2i}, y_{3i}) = \gamma_i \times \exp(-\hat{\sigma}_u \sqrt{2/\pi}). \tag{4}$$

Finally, we can utilize the Jondrow et al. (1982) decomposition to obtain firm-specific inefficiency estimates $\hat{u}_i$. For comparability with the DEA efficiency scores, we convert the inefficiency estimates as cost efficiency measures as follows

$$CE_i = 100\% \times \exp(-\hat{u}_i) \tag{5}$$

In practice, the CNLS problem (2) can be solved by mathematical programming solvers for convex problems. In this study we use GAMS (*General Algebraic Modeling System*) and its MINOS solver as this solver is suitable for solving nonlinear programming problems. Problem (2) is nonlinear due the logarithmic transformations applied to the observed costs and the estimable frontier costs. Since there is a large number of constraints and parameters, problem (2) is computationally more burdensome than for example the OLS. With the present hardware and software capacity, however, problem (2) is solvable in tolerable time by standard PC, provided that the sample size is not too large (see Lee et al., 2011, for discussion). The accuracy of results depends on the solver and the optimization routines applied therein, as well as the precision of data and the way it is presented to the solver. Further, the optimal $(\beta_{1i}, \beta_{2i}, \beta_{3i})$ are not necessarily unique (analogous to the multiplier weights in DEA), and thus the use of a different solver or different ordering of the data points may result as an alternate optimum.

Parameter estimates of cost functions are usually presented in a form of tables reporting the coefficients and their standard errors (and/or *t*-statistics and *p*-values). In the case of the nonparametric StoNED estimator, the beta coefficients are firm-specific, potentially non-unique, and the standard errors are not readily available. To summarize the firm-specific estimates, Table 3.1 presents the average marginal costs for 10 groups of firms, grouped according to the estimated beta coefficients.[5] The groups have been sorted in a descending order according to the marginal cost on energy transmission. These marginal costs are the most favorable ones for each company: no company could increase its efficiency by deviating from the marginal costs implied by StoNED even if the regulator allowed firms to freely choose their marginal costs.

The average marginal costs are reported on the bottom row of Table 3.1. The estimated marginal costs (0.48 c/kWh for electricity transmission, 930 €/km for network

---

[5] See Kuosmanen (2012) for 3-dimensional graphical illustrations of the estimated StoNED frontier.

length, and 13 €/user) appear reasonable based on our experience of this sector (for comparison, the estimated DEA, SFA, and OLS coefficients will be reported in Tables 3.2, 3.3, and 3.4 below). However, firm-specific coefficients can differ substantially from these average values. For example, the marginal cost per user is lowest in Group 1, which consists of firms operating in rural areas, whereas the marginal cost per user is highest in Group 7, consisting of city firms. The last column of Table 3.1 reports the average cost efficiency (CE) of firms within each group. While there are differences in marginal costs, the differences in the average cost efficiency levels are relatively small. In our interpretation, this suggests that the method does not systematically favor some firms due to their operational environment.

Table 3.1: StoNED marginal costs and average efficiencies by firm groups (CRS)

Group	No. of firms	Energy transmission (€ cents/kWh)	Network length (€/km)	No. customers (€/customer)	Average efficiency (%)
1	11	0.6043	876.74	0.87	92 %
2	36	0.5597	984.94	1.23	92 %
3	3	0.4434	908.77	22.25	94 %
4	10	0.4566	1038.81	1.86	93 %
5	3	0.4200	970.69	21.00	92 %
6	4	0.3662	964.71	27.86	95 %
7	3	0.2929	232.21	60.11	92 %
8	7	0.3493	930.93	33.43	91 %
9	6	0.3324	983.05	29.61	90 %
Others	6				96 %
Average		0.4773	930.09	12.94	92 %

3.2. DEA

DEA is an axiomatic, nonparametric approach, similar to StoNED. Kuosmanen and Johnson (2010) have shown that DEA can be obtained as a restricted special case of the CNLS problem (2). If we restrict the residuals ε_i to take only positive values and exclude the contextual variable z, the CNLS problem (2) is equivalent to the input-oriented DEA under CRS (see Kuosmanen and Johnson, 2010, for details). Thus, DEA maintains the same assumptions concerning the shape of the frontier as StoNED.

The nonparametric orientation of DEA is generally seen as its main advantage over the parametric alternatives. However, the main shortcoming of DEA is that it assumes away the stochastic noise term *v*. Hence, all deviations from the estimated frontier are attributed to the inefficiency term *u*. DEA is also sensitive to outliers, as only a few observations determine the frontier.

If we assume away noise, the DEA estimator is consistent, but biased in the small samples (Banker, 1993). In the case of the cost frontier, DEA overestimates the true unobserved cost function in the small samples but it converges to the true frontier as the sample size tends towards infinity. Statistical inference on DEA can be conducted by using the bootstrap methods (e.g., Simar and Wilson, 2008). However, if the stochastic noise term is included in the model, the DEA estimator can be biased in both directions. In this case the bootstrap inferences are invalid. Indeed, it seems a common misunderstanding to assume that the bootstrap method (or robust frontiers) would make DEA more robust to noise. We must emphasize that the probabilistic treatment of sampling error does not address stochastic noise at all.

In the present context, the sensitivity of DEA to noise becomes evident if we examine the evolution of DEA efficiency scores over time. Figure 3.1 plots the DEA efficiency scores for years 2005 – 2008 separately. The firms have been ordered with respect to the efficiency score in year 2005. The DEA efficiency scores plotted in Figure 3.1 are extremely volatile from one year to another. For example, on the right end of the figure there is a company which has fallen from the relatively good performance level (~90%) to an abysmal level of 30% only in a period of two years. Figure 3.1 would suggest that firms can easily improve their efficiency by 10 to 20 percentage points in the period of one year. This sounds unrealistic. In our interpretation, Figure 3.1 illustrates that DEA efficiency scores are very sensitive to stochastic variations due to random noise.

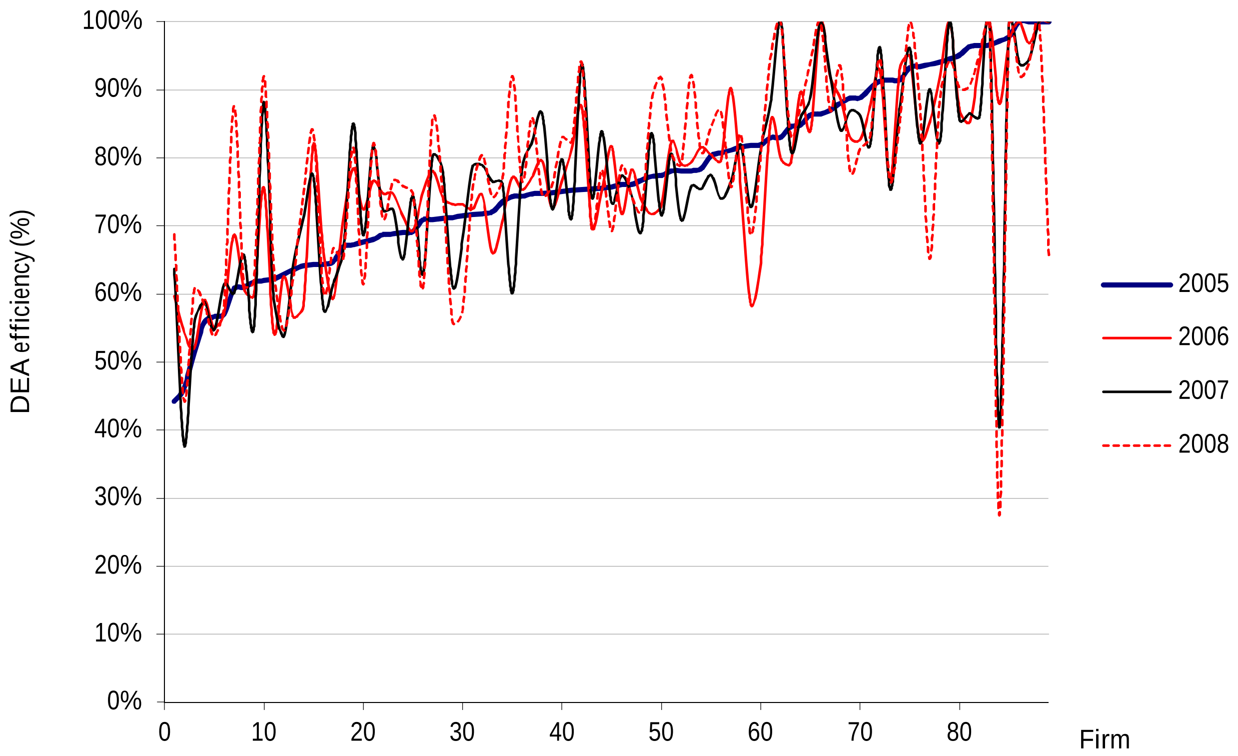

Figure 3.1: The yearly DEA efficiency scores (CRS)

The summary statistics of the marginal costs (shadow prices) of outputs estimated by DEA are presented in Table 3.2 (analogous to Table 3.1). Firms have been classified into 13 groups in a descending order with respect to the marginal cost on energy transmission. The figures are the average marginal cost in each of the groups. Note that for many groups the marginal cost equals zero. In particular, the estimated marginal cost of energy transmission is zero for 5 groups (29 firms). This can partly explain why the average of the DEA estimates for the marginal cost of energy transmission (0.35 c/kWh) is lower than the corresponding StoNED estimate (0.48 c/kWh). Recall that the DEA frontier envelops all observations, attributing all deviations from frontier to inefficiency, whereas the StoNED frontier takes the noise explicitly into account. Therefore, we can expect that the DEA estimates of firm-specific marginal costs are generally lower than the corresponding StoNED estimates. The average DEA shadow price is indeed lower than the StoNED estimate for the network length (DEA: 762 €/km; StoNED: 930 €/km). As for the marginal cost per user, the DEA estimate is notably higher than the StoNED estimate (DEA: 46 €/user; StoNED: 13 €/user). This implies that the shapes of the estimated DEA and StoNED cost frontiers differ considerably, particularly for the output profile of the urban networks that assign a high shadow price for the number of customers (Groups 9 – 12 in Table 3.2).

Table 3.2: DEA marginal costs by firm groups (CRS)

Group	Number of firms	Energy transmission (€ cents/kWh)	Length of network (€/km)	No. customers (€/customer)
1	2	0.5972	0	54.46
2	8	0.5910	489.33	41.28
3	23	0.5866	846.33	15.42
4	3	0.5857	930.62	0
5	12	0.5494	958.45	0
6	2	0.3604	494.77	71.72
7	7	0.3504	863.33	45.77
8	3	0.1491	1142.03	0
9	3	0	0	133.71
10	2	0	182.59	128.49
11	5	0	606.91	111.63
12	16	0	820.00	95.85
13	3	0	1069.20	34.67
Average		0.3526	762.47	46.20

The EMV specification of the DEA model applied in the previous regulation period 2008 – 2011 did not include any contextual variables *z*. The conventional approach to modeling *z*-variables in DEA is to resort to a two-stage approach, where efficiency is first estimated using DEA, and then the DEA efficiency scores are regressed on *z*-variables, using OLS, probit, tobit, or truncated regression. Simar and Wilson (2007) present heavy critique of this approach. Recently, Johnson and Kuosmanen (2012) have shown that one-stage estimation of *z*-variables is possible in DEA. However, in this paper we follow the EMV specification: the empirical DEA results presented above and in Section 4 refer to the model specification where the z-variable has been omitted.

3.3. SFA

The econometric SFA approach requires some parametric assumptions concerning the functional form of the cost frontier. The Cobb-Douglas and translog are the most common functional forms applied in the SFA literature. The SFA model can be obtained as a special case of the generic cost frontier model (1), obtained by imposing some specific functional form for the cost function *C*. Note that if *C* is assumed to be linear (as in Syrjänen et al., 2006, specification, implemented by EMV in 2008 – 2011), then the SFA estimator is obtained as a special case of the StoNED estimator discussed in Section 3.1. Specifically, if we restrict the marginal costs beta to be same for every firm (i.e., $\beta_{ki} = \beta_{kh} \; \forall i,h,k$), then the CNLS problem (2) reduces to a nonlinear least squares problem (note: nonlinearity is due to the logarithmic transformation applied in (2); excluding the log-transformation, problem (2) reduces to the standard OLS problem). The second-stage method of moments estimator discussed in Section 3.1 is commonly applied in the SFA literature, referred to as the Modified OLS (MOLS) (Aigner et al., 1977; Olson et al., 1980).[6] Indeed, StoNED can be seen as an axiomatic, nonparametric variant of the classic MOLS; the conditional expected value $E(x_i \mid y_{1i}, y_{2i}, y_{3i})$ is estimated by CNLS instead of OLS, but otherwise the StoNED estimator follows the standard MOLS procedure.

The main shortcoming of SFA is that the functional form assumptions are somewhat arbitrary and difficult to justify. In the present context, we note that many

[6] Sometimes MOLS is referred to as corrected OLS (COLS) (see, e.g., Azadeh et al., 2009). We prefer to use MOLS for the probabilistic estimator that takes into account noise, and reserve the term COLS for the deterministic estimator that envelopes all observations.

commonly used functional forms fail to capture the economies of scope in joint production (e.g., Syrjänen et al., 2006). For example, the standard Cobb-Douglas function is quasi-concave at all parameter values, which implies the Cobb-Douglas cost function exhibits economies of specialization rather than economies of scope. Thus, the use of the Cobb-Douglas cost function as a benchmark in regulation could give wrong incentives to specialize in provision of just one output instead of a balanced portfolio of outputs. Note that in the regulatory model of EMV the network length and the number of customers are treated as output variables to capture the capacity of the network and the potential demand for electricity. Thus, the Cobb-Douglas functional form is considered problematic in this setting (Syrjänen et al., 2006). The flexible functional forms such as translog are subject to the same problem, and the larger number of parameters would likely cause additional problems with multicollinearity. This is why EMV chose to use the linear functional form in their SFA specification. However, the linear functional form assumes that outputs are perfect substitutes. Thus, the linear cost frontier tends to favor the "average firm" over the firms operating with an atypical output profile (e.g., industrial networks) or firms operating in an atypical environment (e.g., largest cities or remote rural areas).

As a partial adjustment to the heterogeneity of firms and their operating environments, the total network length y_2 was divided in two parts in the SFA model EMV applied in the previous regulation period 2008 – 2011), specifically,

$$y_2 = y_{2A} + y_{2B}, \tag{6}$$

where

y_{2A} = length of underground cabled urban network (km)

y_{2B} = length of other network (km)

Treating y_{2A} and y_{2B} as separate outputs in the SFA model, the marginal cost of the underground cabled urban network is allowed to be higher than that of the other network. This can partly alleviate the restrictive linear specification of the cost frontier. However, the use of different sets of output variables in the DEA and SFA models is problematic for the parallel use of both methods as a part of the regulatory model. We discuss this issue in more detail in Section 3.4. Finally, we must emphasize that the SFA literature offers abundant number of ways for modeling contextual variables *z* (e.g., Kumbhakar and Lovell, 2000, Ch. 7, and references therein). However, in this study we restrict to the EMV specification discussed above. In the following we examine the effect of dividing the network length to two separate output variables on the parameter estimates of the slope coefficients. In the empirical comparison of the efficiency estimates presented in Section 4, we apply the EMV specification of the SFA model involving four output variables.

The SFA estimates of the marginal costs of outputs are presented in Table 3.3. For completeness, we report the estimates for the three-output model where the total network length (y_2, Model A) is used as an output and for the four-output model where the network length is separated in two components (y_{2A} and y_{2B}, Model B). The SFA model is estimated by maximum likelihood assuming CRS. In Sections 3 and 4, we assume the truncated normal distribution for the inefficiency distribution, as this is the specification that EMV used in the previous regulation period, following Syrjänen et al. (2006).

Table 3.3: Marginal costs of outputs estimated by SFA; in Model A the total network length is used; in Model B the urban network (y_{2A}) and other network (y_{2B}) are treated as separate outputs

	Model A	*Model B*
y_1 : Energy trans. (€ cents/kWh)	0.61**	0.60**
	(0.000)	(0.000)
y_2 : Network length (€/km)	896.74**	–
	(0.000)	
y_{2A} : Urban network (€/km)	–	1115.94**
		(0.001)
y_{2B} : Other network (€/km)	–	904.06**
		(0.000)
y_3 : No. customers (€/customer)	25.32	20.12
	(0.114)	(0.264)

p-values in parenthesis, significance indicated with an asterisk; 1%**, 5%*

Comparing the results of Table 3.3 with the marginal costs reported in Table 3.1 (StoNED) and 3.2 (DEA), we find that the marginal costs suggested by SFA differ from the average marginal costs estimated by DEA or StoNED. For energy transmission, for example, the marginal cost estimates obtained by SFA are notably larger than the average of the StoNED estimates (only for Group 1 in Table 3.1, the marginal cost is close to the SFA estimates), and almost twice as large as the average of DEA estimates (Groups 1 – 4 in Table 3.2 yield marginal costs nearly as high as the SFA estimates). Interestingly, the average StoNED marginal cost for energy transmission (0.4773) is approximately the average of the average DEA marginal costs (0.3526) and the SFA marginal cost (0.60) from the second model. This illustrates the unifying nature of the StoNED approach.

In Model B, the estimated marginal cost of underground cabled urban network is higher than that of the other network, as expected. Interestingly, the marginal cost of the total network length in Model A is lower than the marginal cost of the other network in Model B. Division of the network length on two parts has little effect on the marginal cost of the energy transmission, but does have a notable impact on the marginal cost per user. Clearly, taking the heterogeneity of firms into account influences the marginal cost estimates. Note that the nonparametric DEA and StoNED methods allow for firm-specific marginal costs, which provides greater flexibility in terms of the heterogeneity of firms and their operating environments, as discussed at the end of Section 3.1.

The SFA estimate for the marginal cost per user is relatively small and insignificant at the conventional significance levels. The StoNED estimates for the marginal cost per user are larger for some groups (particularly firms operating in large cities), but the average of StoNED estimates falls below the SFA estimate. The DEA estimates are notably larger, for three groups the marginal cost estimate exceeds 100€ per user. For firms operating in rural areas, the number of customers is not the main cost driver; majority of Finnish distribution networks operate in rural areas. This explains why the SFA estimate and the averages of DEA and StoNED estimates of the marginal cost per user are rather low.

The SFA results reported in Table 3.3 have been estimated using the heteroskedasticity correction suggested by Syrjänen et al. (2006). Specifically, it is assumed that the variances of the inefficiency and noise terms are proportional to the amount of transmitted energy (y_1). In econometrics, the textbook treatment of such heteroskedasticity is to normalize all variables by y_1. However, the assumed form of heteroskedasticity appears completely arbitrary: one could equally well assume that heteroskedasticity is driven by the network length, the number of customers, or perhaps some combination of all output variables. To examine the effect of heteroskedasticity correction in more detail, we have estimated the SFA model again using each output variable as the normalizing criterion, and without any normalization. The modified OLS (MOLS) parameter estimates of the models with alternative normalizations are reported in Table 3.4, both under CRS (the top part) and variable returns to scale (VRS, the bottom part) [in the CRS case the constant is restricted to zero].[7]

Table 3.4 shows that the choice of the normalization has a major impact on the parameter estimates (the marginal costs of outputs). One of the output variables has a negative marginal cost in seven out of the eight specifications considered: only in the VRS model normalized by energy transmission all coefficients are positive as expected. The normalization also influences the skewness of the residuals. If no normalization is applied, or the normalization is based on the number of customers, then the skewness of the OLS residuals has a wrong sign, and hence the stochastic frontier reduces to the OLS curve. In Table 3.4, these cases are indicated by # on the row “Expected efficiency”. On the other hand, if the normalization is based on the network length, the skewness is so large that the

---

[7] The maximum likelihood estimator of the SFA model fails due to wrong skewness in six out of the eight specifications considered. For comparison, we report the Modified OLS (MOLS) estimates throughout all eight specifications considered in Table 3.4.

estimate of $\hat{\sigma}_v$ becomes negative. These cases are indicated by §. Thus, we find that the normalization by energy transmission is not only important for heteroskedasticity correction: it is the only specification in Table 3.4 that yields meaningful efficiency estimates as well as positive marginal costs in the VRS case. We suspect the parameter estimates are sensitive to the choice of normalization due to multicollinearity of output variables. Indeed, the output variables are highly correlated. For example, the correlation coefficient between the energy transmission and the number of customers is 0.985. This explains why the normalization by one variable has such a dramatic impact on the OLS coefficients.

Table 3.4: The impact of normalization on SFA (MOLS estimates)

CRS model	*Normalization*			
	None	*By energy*	*By other network*	*By customers*
Energy (€ cents/kWh)	0.09	1.00**	0.49**	0.57**
U. cabling (€/km)	-462.17	1464.38**	-2445.20**	16340.92**
Other network (€/km)	1044.65**	916.59**	1367.96**	248.66
Users (€/user)	113.03**	-0.67	56.00**	-62.62
Expected efficiency	#	72%	§	#
R^2	0.998	0.876	0.949	0.997
VRS model	*Normalization*			
	None	*By energy*	*By other network*	*By customers*
Constant (t€)	108.85	111.82**	-183.51**	713.27**
Energy (€ cents/kWh)	0.08	0.90**	0.56**	0.44**
U. cabling (€/km)	-534.79	1067.59**	-994.90	-279.20
Other network (€/km)	1053.56**	854.50**	1412.23**	578.94**
Users (€/user)	113.82**	10.50	50.39**	40.49
Expected efficiency	#	81%	§	#
R^2	0.997	0.899	0.961	1.000

*Significance indicated with an asterisk; 1%**, 5%**

 # indicates negative skewness (negative $\hat{\sigma}_u$)

§ indicates too large skewness (negative $\hat{\sigma}_v$)

3.4. Naïve model averaging (NMA)

Given the relative strengths and limitations of DEA and SFA, it might be tempting to try alleviate the risk of model misspecification by taking the average of the two estimators. In Finland, EMV applied the average of DEA and SFA estimators in the previous regulation period 2008 – 2011.[8] We refer to this simplistic approach as naïve model averaging (NMA). This section provides a brief but critical examination of the shortcomings of NMA.

Let us first examine the statistical properties of NMA based on the known properties of SFA and DEA. If the parametric assumptions of the SFA estimator hold, the MOLS and the maximum likelihood estimators of the cost frontier C are unbiased and consistent (Greene, 2008). The firm specific inefficiency term u_i can be estimated by using the conditional expected value of Jondrow et al. (1982). This estimator is unbiased, but inconsistent. In the cross-sectional setting, the inconsistency of the firm-specific inefficiency estimator is due to the fact that inefficiency is estimated based on the residuals and there is only one observation available for each firm. While an increase in the sample size improves the fit of the cost frontier, it does not improve precision of the firm-specific efficiency estimates. Thus, if we are interested in firm-specific efficiency scores, then inconsistency of the SFA estimator directly implies the NMA estimator is inconsistent even if the assumptions of the SFA model hold.

To obtain a consistent estimator of firm-specific efficiency, we must assume away noise. In this case, the DEA estimator is consistent under the stated axioms. The SFA estimator remains inconsistent even if the functional form is correctly specified, so there is little benefit to introduce SFA: the DEA estimator is consistent, whereas NMA is not. By assuming away the noise, we lose the most desirable property of SFA.

As for the estimation of the cost frontier C, the statistical consistency of the NMA estimator requires that the assumptions of both DEA and SFA hold simultaneously. That

[8] Similar practice of combining DEA and SFA estimators has been used or considered for use in other countries as well, see, e.g., Pollit (2005), Azadeh et al. (2009), and Bogetoft and Otto (2011), Ch. 10.

is, the NMA estimator is consistent only if the frontier is linear with respect to outputs, inefficiency *u* has a truncated normal distribution, and there is no noise *v*. In this situation SFA estimator is unbiased and consistent. The DEA estimator is consistent but biased. Thus, the NMA estimator is consistent but biased. We conclude that under the assumptions required for the statistical consistency of the NMA cost frontier estimator, the SFA estimator is both unbiased and more efficient than the NMA estimator: introducing the DEA estimator does not provide any real benefit in this situation.

The problems of NMA are further intensified by the fact that EMV applied different sets of output variables in DEA and SFA. In DEA the total network length was used as an output, whereas in SFA the network length was divided in two output variables, the urban underground cabled network and other network, as discussed in Section 3.3. This creates a profound misspecification problem. If the two models are differently specified with respect to the output variables, then one of the models (if not both) has to be misspecified. If one of the models is misspecified, then so is the NMA estimator. There is no reason to expect that averaging wrongly specified estimators would be beneficial. The Monte Carlo simulations presented in Section 5 demonstrate this point. But first, we proceed to an empirical comparison of the efficiency estimates obtained with StoNED, DEA, SFA and NMA.

4. Comparison of efficiency estimates

The previous section presented some selected empirical evidence of the cost frontier obtained with different methods. In this section we compare the empirical estimates of cost efficiency (CE). [9] Our focus on the CE scores is motivated by the fact that the Finnish legislation mandates the use the efficiency improvement targets as the regulatory instrument of EMV.

---

[9] For all methods, we follow the model specifications applied by EMV. For comparability, CRS is imposed throughout all estimation methods considered.

Let us first examine the correlations between the CE scores estimated by the four methods. Table 4.1 reports the correlation matrices of the Pearson product moment correlation coefficients (the left side), and the Spearman rank correlation coefficients (the right side). There is a high positive correlation in every pair of CE estimates. Based on the correlation analysis alone, one might be tempted to conclude the choice of the estimation method has little effect on the efficiency estimates. However, this conclusion proves wrong in a closer inspection of the levels of CE estimates.

Table 4.1: The correlation analysis of efficiency scores

Pearson correlation					Spearman rank-correlation			
	StoNED	DEA	SFA	NMA	StoNED	DEA	SFA	NMA
StoNED	1	0.9089	0.8956	0.9367	1	0.9338	0.8788	0.9498
DEA		1	0.8568	0.9726		1	0.8456	0.9732
SFA			1	0.9523			1	0.9329
NMA				1				1

Table 4.2 reports some descriptive statistics of the CE scores obtained by different methods. There are notable differences in the levels of efficiency scores. In particular, we find that StoNED yields considerably higher efficiency scores than any other method, both in terms of the mean and the minimum. This is due to the fact that StoNED takes the noise term explicitly into account and captures heterogeneity of firms and their operating environments through the use of the contextual variable z, which is omitted in other methods.

Table 4.2: The descriptive statistic of efficiency scores

	Mean	St. Dev.	Median	Min	Max
StoNED	0.924	0.069	0.940	0.764	1.000
DEA	0.802	0.119	0.807	0.466	1.000
SFA	0.862	0.092	0.892	0.545	0.981
NMA	0.832	0.102	0.848	0.505	0.990

The summary statistics of Table 4.2 facilitate the comparisons of an average or a median firm. To shed further light on efficiency of individual firms, we have plotted the StoNED efficiency scores against the NMA estimates in Figure 4.1. Points in this diagram represent the pair of efficiency estimates obtained by the average of DEA and SFA (NMA, the horizontal axis) and StoNED (the vertical axis). The broken line in the middle of diagram indicates the 45 degree line: for points above this line the StoNED efficiency estimate is greater than that of NMA.

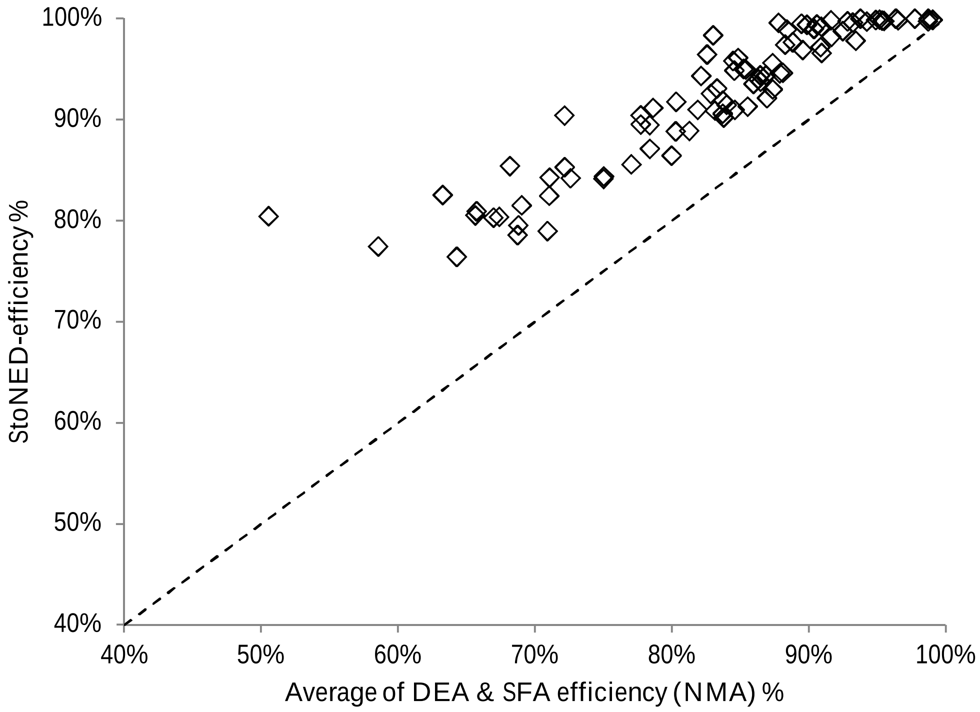

Figure 4.1: Comparison of StoNED and NMA efficiency scores

Figure 4.1 illustrates that the StoNED estimator is more favorable for each individual firm than the average value of DEA and SFA; the StoNED efficiency scores are higher than the corresponding NMA values. For some companies the use of NMA value would yield efficiency improvement targets around 35% to 50% (efficiency of 50% to

65%). Improvements of this magnitude seem highly unrealistic. There are also a number of firms which are rather close to the efficient StoNED frontier. This is in contrast especially with DEA, where only a few observations define the efficient frontier.

The main objective of the efficiency estimates is to provide cost targets that EMV imposes to the distribution firms. To examine the impacts of the methodological choice on the bottom line, we have converted the firm-specific efficiency estimates to monetary cost reduction targets, calculated as $(1-\exp(-\hat{u}_i))x_i$, where $\hat{u}_i$ is the firm-specific estimate of inefficiency (the JLMS estimator used for SFA and StoNED) and x_i is the observed total cost. The total cost reduction target for the whole industry is reported in the first column of Table 4.3 (all figures in €1,000 at prices of the year 2008). The remaining columns provide summary statistics of the firm-specific cost reduction targets.

Table 4.3: Monetary cost reduction targets (Thousand € in prices of 2008)

	Industry	Mean	St. Dev	Min	Max
StoNED	47 508	534	1 326	0.000	11 113
DEA	141 382	1 589	3 888	0.000	27 654
SFA	93 023	1 045	2 185	0.024	13 599
NMA	117 205	1 317	2 947	0.017	20 627

Comparison of the cost reduction targets reveals substantial differences between the methods considered. The calculated total cost reduction target of the industry based on the StoNED estimates is somewhat lower than 50 Million €. The SFA estimate is approximately 50 Million € larger than the corresponding StoNED figure. Further, the DEA estimate is approximately 50 Million € larger than the SFA estimate. Although the efficiency scores obtained with the different methods are highly correlated, the monetary figures presented in Table 4.3 illustrate that the choice of the estimation method does have a significant economic impact within the regulatory framework.

Based on Table 4.3, one might conclude that StoNED is most favorable to the regulated firms, whereas DEA is the best method from the perspective of consumers. However, the implementation of the cost reduction targets in the regulatory framework may change this perception. In reality, more modest targets may be easier to enforce than targets that appear unrealistic to achieve. In the previous regulation period 2008 – 2011, the NMA efficiency improvement targets were not imposed directly on the total cost TOTEX as in Table 4.3. Rather, the NMA efficiency estimates were first adjusted upwards by a so-called "error margin correction". Subsequently, the efficiency improvement targets, calculated based on the total cost, were attributed to the operational cost OPEX, based on the argument that firms can influence the OPEX while the capital cost CAPEX is fixed in the short run.[10] The estimated cost efficiency (*CE)* estimates based on TOTEX, could be converted to OPEX by multiplying with the ratio OPEX/TOTEX, but EMV used the inverse TOTEX/OPEX. As a result of these peculiar adjustments, the NMA efficiency improvement targets were watered down from the original 29 percent of the total cost to mere 5 percent. In other words, the estimated average efficiency of 71 percent was adjusted upward to 95 percent at the implementation stage.

In the current regulation period EMV intends to enforce the StoNED targets more vigorously. As a result, a large consortium of distribution firms sued EMV at the beginning of 2012, demanding that the cost targets are attributed to OPEX only, the cost of interruption (INT) is excluded, and that an "error margin correction" is implemented similar to the previous regulation period, among other things. The Finnish Market Court will give the final ruling regarding the implementation by the end of 2012.

One important lesson for the implementation is worth noting. As we emphasized in Section 3, consistent estimation of firm-specific inefficiency is impossible in the stochastic

[10] If efficiency improvements are attributed only to OPEX, this could lead to unwarranted overinvestment in CAPEX where there are no efficiency targets. This gives firms more room to '*game*' the regulator by showing artificial improvement of efficiency (see e.g. Jamasb et al., 2003, 2004, for further discussion).

setting. By the Finnish law, however, EMV needs to use the efficiency improvement targets as its regulatory instrument. To circumvent this dilemma, EMV has replaced the cost efficiency measure (5) (which applies the JLMS estimator of inefficiency *u*) by the cost frontier based measure

$$CE' = C(y_1, y_2, y_3) \times \exp(\delta z) / x \,. \qquad (7)$$

In this measure, the nominator can be consistently estimated. The denominator (observed total cost) contains both inefficiency and noise, but the presence of noise is not a problem if the main objective is to specify efficiency improvement targets such that firms reach the efficient cost level $C(y_1, y_2, y_3) \times \exp(\delta z)$.

5. Monte Carlo simulations

The empirical comparison presented in the previous section shows that the choice of the frontier estimation method does matter in the regulation. We next examine performance of alternative methods in a simulated setting where the true cost frontier and the firm-specific inefficiencies are known beforehand. The advantage of the Monte Carlo (MC) comparison is that it allows us to quantify the performance of each method in terms of standard criteria such as the bias and root mean squared error (to be defined below).

A critical step in the MC analysis is the specification of the data generating process (DGP) that produces the simulated data. For the empirical relevance of the MC analysis, it is desirable to specify the DGP to imitate both the characteristics of the regulatory model the observed patterns of empirical data. In this study, we have calibrated the DGP to reflect both these aspects.

5.1 Data generating process (DGP)

The generic cost frontier model (1) forms the basis of our DGP. To ensure comparability, in the MC comparisons we apply exactly the same model specification across all methods.

Thus, we assume a three output case and omit the contextual variable *z*. In contrast, the empirical comparisons presented in Section 4 were conducted using the EMV specifications of each method (e.g., three outputs in DEA and StoNED, four outputs in SFA, the contextual variable omitted in DEA and SFA).

We first generate random data for the three output variables using the formulas presented in Table 5.1. The DGP for output variables has been specified to mimic the observed data as closely as possible. The empirical distribution of the logarithms of outputs is approximately uniformly distributed within the range [3,11]. The simulated data of the first output (energy transmission) are generated by drawing random numbers from the uniform distribution, and applying the exponential transformation. The other two outputs are subsequently generated conditional on the first output, applying the empirical correlations between the output variables. For example, to simulate data of the network length, we draw a new random variable from the uniform distribution, apply the exponential transformation, and take a weighted average of thus obtained new variable and the previously generated energy output, with weights $\sqrt{(1\text{-}0.87^2)}$ and 0.87, respectively. The weight 0.87 is the empirical correlation coefficient between the network length and the transmitted energy. Thus, the simulated output data exhibit similar correlations as the observed output variables in our empirical data.

Table 5.1: The DGP for the output variables

Output	DGP
Energy	$y_{1,i} = \exp(\text{Uni}[3,11])$
Network length	$y_{2,i} = \sqrt{(1\text{-}0.87^2)} \times \exp(\text{Uni}[3,11]) + 0.87 \times y_{1,i}$
Customers	$y_{3,i} = \sqrt{(1\text{-}0.98^2)} \times \exp(\text{Uni}[3,11]) + 0.98 \times y_{1,i}$

Given the simulated output data, the next step is to generate the total cost. This requires a specification of the cost function. Recall that the commonly used functional forms such as the Cobb-Douglas and translog are inappropriate in the present context. To calibrate our DGP to the current regulatory practice of EMV as closely as possible, we apply the piece-wise linear cost frontier applied by EMV in the regulation period 2012 – 2015. Given the output vector ($y_{1,i}$, $y_{2,i}$, $y_{3,i}$), the value of the cost frontier is calculated as

$$C_i = \max_h(\beta_{1h} y_{1i} + \beta_{2h} y_{2i} + \beta_{3h} y_{3i}) \tag{8}$$

where $(\beta_{1h}, \beta_{2h}, \beta_{3h})$, $h = 1,\ldots,H$ are the slope coefficients (marginal costs) of the H different hyperplane segments of the piece-wise linear cost frontier implemented by EMV (compare with problem (2) and the shadow prices reported in Table 3.1). Note that the max operator in (8) selects the most favorable output prices for each simulated data point.

Having calculated the values of the frontier cost function (which represents the efficient cost level) for each simulated point, the observed total cost are generated using

$$x_i = C_i \times \exp(u_i + v_i) \ , \tag{9}$$

where the inefficiency u and for the noise v are distributed as: $u_i \sim \left|N(0, 0.17^2)\right|$ and $v_i \sim N(0, 0.09^2)$. The parameter values of the standard deviations of the inefficiency and noise terms are calibrated based on the empirical estimates obtained by applying the method of moments estimator to the CNLS residuals in the StoNED procedure.

Before proceeding to the results, it is worth to discuss whether and to what extent the DGP provides an unfair advantage to any of the methods considered. First, the DGP does not violate any of the assumptions of the StoNED method. The piece-wise linear functional form of the true cost function used in the simulations is compatible with the form of the StoNED frontier, but the same is true for DEA. The fact that the coefficients $(\beta_{1h}, \beta_{2h}, \beta_{3h})$ and the parameters (σ_u, σ_v) have been ex ante estimated by the StoNED method does not give any particular advantage to this or that method: the purpose of the ex ante estimation is to match the DGP with the current regulatory practice of EMV. As for DEA, the presence of the noise term *v* violates the deterministic nature of this method. However, empirical data are always subject to some noise, and some authors explicitly suggest that DEA is robust enough to tolerate some noise (see e.g., Gstach, 1998; Banker and Natarajan, 2008). In fact, the noise term can help to alleviate the small sample bias of the DEA estimator, as we note below. Regarding SFA, the piece-wise linear functional form violates the maintained assumption of the linear cost function. In all other respects, the SFA estimator is correctly specified: we assume the half-normal distribution of the inefficiency term (in contrast to the EMV specification of truncated normal inefficiency used in the previous sections). For comparability of SFA and StoNED, we apply the MOLS estimation strategy for SFA and the method of moments estimator in StoNED. As the rigid functional form of SFA and the deterministic orientation of DEA are the well known characteristics of these methods, and the NMA approach is supposed to resolve these issues, we do not find that the DGP described above would give any unfair advantage to any method considered.

5.2 Performance measures

Recall from Section 3 that the SFA and StoNED estimators of the cost frontier *C* are consistent, whereas the JLMS estimator of firm-specific inefficiency is inconsistent. Since no consistent estimator of firm-specific inefficiency is available in the stochastic setting involving noise, we compare performance of the methods in terms of their precision in estimating the cost frontier *C*. Given the simulated values C_i (calculated using (9)) and the corresponding estimates $\hat{C}_i$ (obtained with StoNED, DEA, SFA, and NMA), the performance of the method is measured using the root mean squared error (RMSE) and bias, defined as

$$RMSE = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{C}_i - C_i}{C_i}\right)^2} \qquad (10)$$

$$BIAS = \frac{1}{M}\frac{1}{n}\sum_{m=1}^{M}\sum_{i=1}^{n}\frac{\hat{C}_i - C_i}{C_i} \qquad (11)$$

where M denotes the number of replications in the simulation. Note that the RMSE is always greater than or equal to zero, with zero indicating perfect precision. In contrast, the bias can be positive or negative, the positive values indicating overestimation and the negative values underestimation of the cost function. For both performance statistics, values close to zero are desirable. Both RMSE and bias have been normalized such that the performance statistics have an interpretation as an average dispersion or bias. For example, RMSE = 0.05 indicates that the estimates $\hat{C}_i$ deviate from the true C_i value by 5 percent on average.

5.3 Simulation results

The MC simulations were conducted using the GAMS software and the MINOS solver run on a standard desktop PC (the GAMS code for the simulations is available as online annex). We consider four different scenarios with sample sizes n = 25, 50, 100, and 200. The sample sizes are chosen to be relatively small to reflect the usual number of firms in this kind of sector (in the EMV data, n = 89). Each scenario has been replicated M = 1,000 times. The results of the MC analysis are reported in Table 5.2.

Table 5.2: Simulation results

	RMSE				BIAS			
	n=25	n =50	n =100	n =200	n =25	n =50	n =100	n =200
StoNED	0.072	0.057	0.044	0.027	0.030	0.022	0.014	0.009
DEA	0.088	0.091	0.107	0.129	-0.025	-0.060	-0.091	-0.118
SFA	0.469	0.886	1.439	1.923	-0.253	-0.666	-1.192	-1.661
NMA	0.254	0.464	0.750	1.003	-0.139	-0.363	-0.641	-0.890

Consider first the RMSE statistics reported on the left panel of Table 5.2. The StoNED estimator has a lower RMSE than other methods at all sample sizes. The average dispersion of approximately 5 percent from the true value is a very good result in the stochastic setting involving noise. Note that the precision of the StoNED estimator improves (RMSE decreases) as the sample size increases, as expected. The DEA estimator yields a relatively good precision of RMSE less than 10 percent at small sample sizes. However, the RMSE increases together with the sample size. This is due to the fact that DEA ignores the noise term. In small samples, the noise term and the small sample bias offset each other, but as the sample size increases, the bias due to the noise term starts to dominate. The SFA estimator yields catastrophic results in this comparison, with average deviations of the magnitude of 50 – 200 percent. Recall that the linear functional form is severely wrongly specified in these simulations; most reported MC simulations assume the correct (or almost correct) functional form for SFA. It is not surprising to find that the linear functional form fails to capture the piece-wise linear cost function (8) used in our simulations. Further, the high correlation between the output variables makes SFA vulnerable to multicollinearity. Moreover, note that the RMSE of SFA increases alarmingly as the sample size increases. Finally, the MC simulations illustrate the weakness of the NMA approach: the poor performance of SFA carries over to the NMA estimator. In this case, the use of DEA alone is clearly superior to NMA.

The bias statistics are reported on the right panel of Table 5.2. The bias of the StoNED estimator is small, and decreasing as the sample size increases. In contrast to DEA and SFA, the bias of the StoNED estimator is positive, which means that StoNED tends to overestimate the true cost level in this setting. In the context of regulation, modest overestimation is generally preferred to underestimation. The conventional wisdom of DEA suggests that the DEA estimator is systematically biased towards overestimation of cost. However, this idea stems from the deterministic setting, whereas in the present MC simulations the DGP contains noise. The results of Table 5.2 aptly illustrate that the DEA estimator is downward biased under noise. In very small samples, the noise term can offset the small sample bias, as we noted above.

Finally, we must emphasize that the previous MC comparison has been calibrated to mimic the regulatory model of EMV and the empirical data of the Finnish electricity distribution firms as closely as possible. The purpose of such tailored simulations is to ensure the relevance of the MC evidence in the specific context of the Finnish regulatory model. We stress that the results of this section do not necessarily apply to other sectors or in other countries. As MC simulations are nowadays relatively inexpensive, we suggest that investigating the internal consistency the benchmarking methods through MC simulations calibrated to the specific regulatory context should be routinely conducted.

6. Conclusions

In this paper we have compared the frontier estimation techniques applied in the benchmark regulation of electricity distribution firms. The comparison was conducted both in terms of the empirical data from Finland and in the controlled environment of Monte Carlo simulations. Our empirical comparison demonstrated that the choice of benchmarking method has significant economic effects on the regulatory outcomes, even when the efficiency estimates from different methods are highly correlated. Although the frontier estimation methods are often used for assessing relative efficiency and ranking of firms, in the context of regulation, also the level of efficiency matters.

A unique feature of our Monte Carlo simulations concerns the specification of the data generating process. We calibrated the simulation model and its parameters to capture as closely as possible the key characteristics of the distribution sector and the regulatory system in Finland. This allows us to estimate the potential bias and dispersion of the frontier estimates obtained with different frontier estimation methods in the setting that mimics the empirical reality of this sector.

We find there are at least five important lessons to be learned from this study:

1) *Heterogeneity:* a large proportion of the observed dispersion in the cost per kWh across firms can be explained and attributed to the heterogeneity of firms and their operating environments. The benchmarking model should be flexible enough to take into account the different circumstances of small firms and large corporations, firms operating in rural area or in a large city, and firms that supply power to households or heavy industry.

2) *Noise:* the cost data are subject to random variation from various sources. For example, the capital expenditures depend on somewhat arbitrary accounting rules and depreciation rates. Random weather events such as storms cause interruptions, which influence the operational costs. In these circumstances, stochastic frontier models that explicitly recognize a random noise term are preferable to deterministic benchmarks that attribute all deviations from the frontier to inefficiency.

3) *Use the frontier:* Third, it is important to recognize that the estimation of the frontier cost function (or production function) rests on a much sounder statistical foundation than the estimation of firm-specific efficiency scores. Therefore, it is generally recommended to set the efficiency improvement targets based on the frontier cost or production function, rather than the firm-specific efficiency estimates.

4) *Tailored simulations:* in this paper we have shown that it is possible to calibrate the simulation model to mimic the characteristics of the regulated industry as well as the regulatory model. Conducting tailored simulations is an inexpensive way to compare the performance of alternative benchmarking tools in the specific context of application. We would recommend the use of calibrated Monte Carlo simulations as a test for the internal consistency of the chosen benchmarking model.

5) *Implementation:* Finally, it is worth to recognize that there is little benefit of the use of high-tech econometric tools for the estimation of benchmarks if the regulatory outcomes are watered down or nullified at the implementation stage. Development of a benchmarking model should not be viewed as an isolated exercise, but rather as an integral part of designing the regulatory framework as a whole.

References

Agrell, P., Bogetoft, P., 2007. Development of benchmarking models for German electricity and gas distribution, Sumicsid AB, Project GERNER/AS6, Final Report. http://www.bundesnetzagentur.de/cae/servlet/contentblob/88060/publicationFile/1932/GutachtnSUMICSID-Id9598.pdf (accessed 12 Oct 2011).

Agrell P., Bogetoft, P., Tind, J., 2005. DEA and dynamic yardstick competition in Scandinavian electricity distribution. Journal of Productivity Analysis 23, 173–201.

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier models. Journal of Econometrics, 6, 21-37.

Azadeha, A., Ghaderia, S.F., Omranib, H., Eivazya, H. 2009 An integrated DEA–COLS–SFA algorithm for optimization and policy making of electricity distribution units. Energy Policy 37 (7), 2605-2618.

Banker, R.D., 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science, 39 (10), 1265-1273.

Banker, R.D., Natarajan, R., 2008. Evaluating contextual variables affecting productivity using data envelopment analysis. Operations Research, 56 (1), 48-58.

Bogetoft, P., Otto, L., 2011. Benchmarking with DEA, SFA, and R. International Series in Operations Research & Management Science, 157, Springer. DOI 10.1007/978-1-4419-7961-2_10.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429-44.

Cullmann, A., 2009. Parametric and nonparametric efficiency analysis in electricity distribution - A European perspective. Ph.D. Dissertation, Technische Universitat Berlin. http://opus.kobv.de/tuberlin/volltexte/2009/2192/pdf/cullmann_astrid.pdf (accessed 24 May 2012).

Cullmann, A., 2012. Benchmarking and firm heterogeneity: A latent class analysis for German electricity distribution companies, Empirical Economics, in press.

Dassler, T., Parker, D., Saal, D.S., 2006. Methods and trends of performance benchmarking in UK utility regulation. Utilies Policy, 14, 166-174.

Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society A 120: 253-81.

Forsund, F.R., Kittelsen, S.A.C., 1998. Productivity development of Norwegian electricity distribution Utilities. Resources and Economics 20, 207-224.

Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (eds.), 2008. Measurement of Productive Efficiency and Productivity Growth, second ed., Oxford University Press, New York.

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.

Gstach, D., 1998. Another Approach to Data Envelopment Analysis in Noisy Environments: DEA+. Journal of Productivity Analysis, 9, 161-176.

Hjalmarsson, L., Veiderpass, A., 1992. Productivity in Swedish electricity retail distribution. Scandinavian Journal of Economics, 94, 193-205.

Iglesias, G., Castellanos, P., Seijas, A., 2010. Measurement of productive efficiency with frontier methods: A case study for wind farms. Energy Economics 32: 1199-1208.

Irastorza, V., 2003. Benchmarking for distribution utilies: A problematic approach to defining efficiency. Electricity Journal, 16 (10), 30-38.

Jamasb, T., Pollit, M., 2001. Benchmarking and regulation: international electricity experience. Utilies Policy, 9, 107-130.

Jamasb, T., Pollit, M., 2003. International benchmarking and regulation: An application to European electricity distribution utilizes. Energy Policy, 31, 1609-1622.

Jamasb, T., Nillesen, P., Pollitt, M. 2003. Gaming the regulator: A survey. Electricity Journal 16 (10), 68-80.

Jamasb, T., Nillesen, P., Pollitt, M., 2004. Strategic behaviour under regulatory benchmarking. Energy Economics, 26, 825-843.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19 (2-3), 233-238.

Johnson, A.L., Kuosmanen, T., 2011. One-stage estimation of the effects of operational conditions and practices on productive performance: Asymptotically normal and efficient, root-n consistent StoNED method. Journal of Productivity Analysis, 36, 219-230.

Johnson, A.L., Kuosmanen, T., 2012. One-stage and two-stage DEA estimation of the effects of contextual variables. European Journal of Operational Research, in press.

Kinnunen, K.. 2006. Investment incentives: regulation of the Finnish electricity distribution. Energy Policy 34, 853–862.

Koopmans, T.C., 1951. An Analysis of Production as an Efficient Combination of Activities, in: Koopmans, T.C., (Ed.), Activity Analysis of Production and Allocation. Cowles Commission for Research Monograph No. 13, New York.

Kopsakangas-Savolainen, M., Svento, R. 2008. Estimation of cost-effectiveness of the Finnish electricity distribution utilities. Energy Economics, 30 (2), 212-229.

Korhonen, P., Syrjänen, M., 2003. Evaluation of cost efficiency in Finnish electricity distribution. Annals of Operations Research, 121, 105-122.

Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic frontier analysis. Cambridge University Press, New York.

Kuosmanen, T., 2008. Representation theorem for convex nonparametric least squares. Econometrics Journal, 11, 308-325.

Kuosmanen, T., 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics, forthcoming.

Kuosmanen, T., Johnson, A.L., 2010. Data envelopment analysis as nonparametric least squares regression. Operations Research, 58 (1), 149-160.

Kuosmanen, T., Kortelainen, M., Kultti, K., Pursiainen, H., Saastamoinen, A., Sipiläinen, T., 2010. Sähköverkkotoiminnan kustannustehokkuuden estimointi StoNED-menetelmällä: Ehdotus tehostamistavoitteiden ja kohtuullisten kustannusten arviointiperusteiden kehittämiseksi kolmannella valvontajaksolla 2012-2015. In Finnish, Available at: www.energiamarkkinavirasto.fi (accessed 13 December 2011)

Kuosmanen, T., Kortelainen, M., 2012 Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, in press.

Kuosmanen, T., Kuosmanen, N., 2009. Role of benchmark technology in sustainable value analysis: An application to Finnish dairy farms. Agricultural and Food Science, 18 (3-4), 302-316.

Laffont, J-J., Tirole, J., 1993. A theory of incentives in procurement and regulation. The MIT Press, Cambridge, MA.

Lee, C.Y., Johnson, A., Moreno-Centeno, E., Kuosmanen, T., 2011. A more efficient algorithm for convex nonparametric least squares, SSRN working paper, available at: http://dx.doi.org/10.2139/ssrn.1981260 (accessed 15.5.2012).

Mekaroonreung, M., Johnson, A.L., 2012. Estimating the shadow prices of SO2 and NOx for U.S. coal power plants: A convex nonparametric least squares approach. Energy Economics, in press.

Pollit, M., 2005. The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks. Utilies Policy, 13, 279-288.

Shleifer, A., 1985. A theory of yardstick competition. Rand Journal of Economics, 16 (3), 319-327.

Simar, L., Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136 (1), 31-64.

Simar, L., Wilson, P.W., 2008. Statistical inference in nonparametric frontier models: recent developments and perspectives. In Fried HO, Lovell CAK, and Schmidt SS (eds.) Measurement of productive efficiency and productivity growth, 2nd edition, Oxford University Press.

Syrjänen, M., Bogetoft, P., Agrell, P., 2006. Analogous efficiency measurement model based on stochastic frontier analysis. Gaia Consulting Oy 11.12.2006 (Available at: www.energiamarkkinavirasto.fi).

Weyman-Jones, T.G., 1991. Productive efficiency in a regulated industry: The area electricity boards of England and Wales. Energy Economics 13 (2), 116-122.