Stepwise DEA;
Choosing variables for measuring technical efficiency in Norwegian electricity distribution

By
Sverre A.C. Kittelsen
SNF, Foundation for research in economics and business administration, Oslo[†]

Abstract

A production technology specification in electricity distribution leads to many product aspects that together with detailed input data gives a high dimensionality. Non-parametric methods may give meaningful results when parametric methods lack degrees of freedom, but have problems with collinear or irrelevant variables. Aggregate efficiency estimates will be little affected, but rates of transformation will be corrupted and extreme observations will be measured as efficient by default. This paper uses tests suggested in the literature to measure the significance of the change in results when disaggregating or introducing an extra variable, and lets the data aid in deciding the specification.

JEL Classification: D24, L94, C14

Keywords: Data Envelopment Analysis, Model specification, Non-parametric efficiency estimation, Electricity distribution utilities

[†] SNF-Oslo, N-0371 Oslo, Norway. Tel.:+47-22958815, Fax:+47-22958825, Email: s.a.c.kittelsen@snf.uio.no.

1 Introduction*

In the literature on the measurement of technical efficiency of production the non-parametric deterministic frontier method of characterizing production technology known as Data Envelopment Analysis (DEA) has gained popularity. Rank order tests have been used to compare efficiency in different groups or subsamples. In contrast, however, to the parametric average cost- and production function approaches there have been few attempts at constructing statistical tests of the model specification. Some authors have extended the sensitivity analysis of operations research to DEA, while others have been concerned with the theoretical conditions for one specification to be equivalent to another.

In recent developments, Banker (1993) has proven the consistency of the DEA estimators under specific assumptions and suggested statistical tests of model specification, while Korostelev, Simar and Tsybakov (1995a,b) have been concerned with the rate of convergence of non-parametric estimators. Kneip, Park and Simar (1996) extend these results to a more general model. Simar and Wilson (1995) suggests a bootstrap method for estimating the bias and confidence intervals of efficiency estimates. Even if this approach is feasible, it would be advantageous if simpler techniques were available. Kittelsen (1997) has in simulations examined the performance of some suggested hypothesis tests, and finds that these give crude but usable approximations of the true significance levels and power functions. This paper is concerned with the empirical model specification and uses the statistical methods to evaluate the estimates of technical efficiency and to judge empirically whether variable omission and aggregation is admissible, and the correct specification of the scale properties.

The idea of measuring technical efficiency by a radial measure representing the proportional input reduction possible for an observed unit while staying in the production possibility set stems from Farrell (1957) and has been extended in a series of papers by Färe, Lovell and

* I wish to thank Arne M. Torgersen, Finn R. Førsund, Harald Goldstein and participants at various presentations for helpful comments. I would also like to thank the Norwegian Water Resources and Energy Administration (NVE) for financing and the Norwegian Electricity Distribution Research Institute (EFI) for cooperation in collecting and preparing the data.

others[1]. Farrell's specification of the production possibility set as a piecewise linear frontier has also been followed up using linear programming (LP) methods by Charnes, Cooper et al.[2] The decomposition of Farrell's original measure relative to a constant returns to scale (CRS) technology into separate measures of scale efficiency and technical efficiency relative to a variable returns to scale (VRS) technology is due to Førsund & Hjalmarsson (1974) and has been implemented for a piecewise linear technology by Banker, Charnes and Cooper (1984). Their DEA formulation has served as the main model of most recent efficiency studies and is the basic model in this paper.

In parallel with the non-parametric or mathematical programming approach to efficiency measurement, considerable research has been conducted in a parametric tradition originating with Aigner and Chu (1968). Their deterministic approach was to estimate a smooth production frontier with residuals restricted to be non-negative and interpreting these residuals as a measure of inefficiency. Like in the non-parametric models, this interpretation is vulnerable to measurement errors and model mis-specification. Aigner, Lovell and Schmidt (1977) incorporated a stochastic error term in addition to the inefficiency term in a composed error model of a production frontier. To identify these terms separately in cross-section studies explicit assumptions are needed for the functional form of the distribution of each term[3]. With explicit distributions it is fairly straightforward to construct statistical tests in such models[4]. The drawbacks of the parametric approach are chiefly the structure imposed on the data by the choice of functional forms and the difficulty of modelling multiple-input multiple-output technology when no price information is available, mainly due to the necessity of a large number of observations to get enough degrees of freedom for a single equation estimation[5].

---

[1] E.g. Färe & Lovell (1978) and Färe, Grosskopf & Lovell (1985).

[2] E.g. Charnes, Cooper and Rhodes (1978) who originated the name DEA. For an overview of the literature on DEA see e.g. Seiford (1996).

[3] Panel studies often replace this with restrictions on the time pattern of inefficiency.

[4] For an overview of the econometric approach see e.g. Bauer (1990), Greene (1993).

[5] See e.g. Diewert and Mendoza (1996).

Recently some work has been done on developing stochastic non-parametric frontier models (e.g. Petersen and Olesen, 1995 or Kneip and Simar, 1995). The methods so far suggested require, however, either panel data or exogenous estimates of vital parameters such as constraint violation probabilities. The aim of the present paper is less ambitious. No account is taken of measurement errors, but tools are used that take the analysis of model mis-specification into the realm of the widely used non-parametric models.

One attraction of the non-parametric frontier methods is that the functional form is perfectly flexible without the use of numerous parameters that easily exhaust the degrees of freedom when the number of dimensions becomes large. A full set of disaggregated inputs and outputs and the inclusion of all potentially relevant variables does however create problems even in DEA. First, the DEA method will measure as efficient all units that in some senses have extreme values of the variables, including all those that have the lowest value of an input or the highest value of an output. A variable that in fact is irrelevant to the analysis could therefore destroy the efficiency measures for some units, even if the average efficiency is not much affected. Secondly, in common with the parametric methods, any two variables that are highly correlated and therefore have much of the same information value will tend to distort the rates of transformation and substitution on the frontier. Any use of these marginal properties such as returns to scale, relative shadow prices or marginal costs will therefore be affected[6]. Thirdly, the statistical efficiency of the technically efficiency estimates will be worsened by a higher dimensionality (Kittelsen, 1997). Finally, on a more practical level, a model could become unmanageable and not easy to understand if the dimensionality is very high.

The production possibilities for the Norwegian electricity distribution utilities (EDUs) is restricted not only by technology in a narrow sense, but also by the common constraints given by nature, custom work practice, government regulations, knowledge and organisational technology, including the set of incentive mechanisms available to owners and managers. One must assume that some constraints and some objectives vary across the EDUs, so that

[6] See Olesen and Petersen (1991, 1996) for a discussion of multicollinearity in the DEA model, and e.g. Koutsoyiannis (1977) in econometric models.

there is variation in the resource usage of the units. The constraints that are exogenous to the EDUs, should as far as possible be included in the model specification, since variations in resource usage due to such factor cannot reasonably be interpreted as inefficiency. The actual implementation of incentive mechanisms due e.g. to differences in the objectives or managerial skills of the decision makers, and the differences in the resource usage this gives rise to, should however, be measured as inefficiency. The distribution of the inefficiencies is therefore not truly random. If we were able to model these behavioural differences we would also have more information on how to eliminate inefficiency. Since the Industrial Organisation literature has not so far come up with models that can be tested empirically, we must instead model inefficiency *as if* it was generated randomly.

The Norwegian EDUs are local (natural) monopolies and are mainly owned by local government. Both their monopoly power and their ownership structure give rise to suppositions that the sector need not be technically or cost efficient. The ongoing liberalisation of the electricity markets in Norway has highlighted the need for regulating the EDUs. The Norwegian Water Resources and Energy Administration (NVE) as the regulatory body has therefore commissioned a study of the technical efficiency of the EDUs as a first step in setting up a regulatory regime[7]. While the inputs and the main output (electricity delivered) are easily identified, there are several potentially important climatic and geographic conditions that vary across the 173 EDUs in the sample. Additionally the data are detailed enough to support substantial disaggregation, so the potential dimensionality of the input-output space is very large. The need of a method to aid in the selection of variables is clear.

Leamer (1983) is critical to leaving the selection of variables to a computer, and it has been pointed out by e.g. Green (1993) that in a parametric model found by stepwise regression the classical inference procedures about the parameters breaks down after the first step. Their criticism emphasizes the need to use economic theory and sectorial knowledge to specify as much as possible, but cannot remove the need to use data to decide some specification issues.

[7] The main results from the project are summarised in Norwegian in Kittelsen (1994).

Various studies of the electricity industry have been published, but most deal with electricity production rather than distribution. In a parametric non-frontier setting, Salvanes and Tjøtta (1991) analyse the Norwegian electricity distribution industry, and find that results are significantly affected by introducing two outputs (energy and number of customers) rather than the usual one-output formulation. They find that the returns to scale are only locally different from one, but that there are increasing returns to density when calculated as the inverse cost elasticity of energy. With the use of a DEA model Hjalmarsson and Veiderpass (1992) analyse the Swedish electricity distribution industry and find technical efficiency measures of 0.77 to 0.86 depending on the choice of output aggregates. They compare efficiency by ownership structure and find small differences, but do not offer any statistical tests for these differences, nor for selecting the output specification.

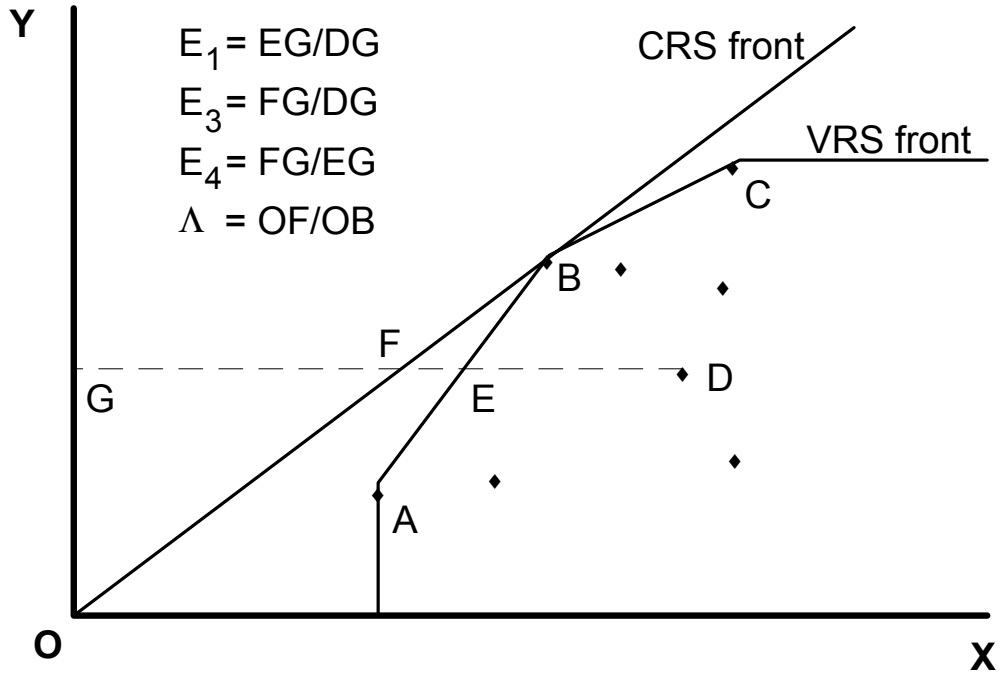


Figure 1*: DEA input saving efficiency estimates in input (X) - output (Y) space. Diamonds represent observations, and the shown measures are for unit D.*

2 The method

An EDU uses an input vector x defined over the set X of inputs and produces an output vector y defined over the set Y of outputs. The technology in the electricity distribution sector can be described by the set of all feasible input-output combinations $P=\{(y,x)|y$ can be produced from $x\}$. The analysis of efficiency uses Farrell's (1957) technical input efficiency measure $E_1(y,x,P)=\text{Min}_\theta\{\theta|(y,\theta x)\epsilon P\}$, which is a measure of the distance from a point to the frontier of this feasible set[8]. Define further the isoquant $\partial L(y)=\{x|E_1(y,x,P)=1\}$ as the set of input-combinations that cannot be reduced for a given output vector y.

The basic assumption on the data generating process follows Kneip, Park and Simar (1996), and are briefly i) that the input-output combinations of the EDUs are i.i.d. distributed on the convex and free disposal feasible technology set P, that ii) the support of the density of the output vector y is compact, that iii) the input *mix* x_j^*/x_k^* has a density conditional on the output level, and that there is a distribution of input inefficiencies γ that has a density $f(\gamma)$

[8] As noted by e.g. Färe and Lovell the Farrell input efficiency measure is the inverse of the Shephard (1970) input distance function.

conditional on output level and input mix. This implies that each EDU is observed with an output vector and input mix corresponding to a point on the isoquant $x^* \epsilon \partial L(y)$, and that the actually observed input vector is this frontier vector multiplied by an inefficiency factor $x=(1+\gamma)x^*$. Finally assume iv) that the density of inefficiencies is such that one will observe an EDU arbitrarily near the frontier when the number of observations is sufficiently large, i.e. there exists constants ϵ_1 and ϵ_2 so that for any allowable output level and input mix, for all $\gamma \epsilon [0, \epsilon_2], f(\gamma) \geq \epsilon_1$.

Given the assumptions above, the true value of Farrell's technical input efficiency is $E_1=1/(1+\gamma)$. The DEA estimate of efficiency is the input efficiency measure relative to the DEA estimate of the feasible technology set, $\hat{E}_1 = E_1(y,x,\hat{P})$, where $\hat{P}$ is given by the restrictions of the LP problem (1) below. Under these assumptions, Kneip, Park and Simar (1996) prove the consistency of the DEA estimator of technical efficiency, although the rate of convergence of this estimator is low and depends also on the curvature of the frontier[9].

The estimation is done in two stages to ensure that the reference point will be on the efficient frontier, and that the solution values to the slacks in the second stage will be meaningful[10]. The EDUs are in the set of observations $i \in N$. The efficient subset of EDUs is identified by first solving the LP-problems:

$$
\begin{aligned}
\bar{s}_i \;=\; \underset{\lambda, s^y, s^x}{Max} \Big\{ & \sum_{k \in Y} s_{ik}^y \;+\; \sum_{l \in X} s_{il}^x, \;\; s.t. \\
& \sum_{j \in N} \lambda_{ji} y_{jk} - y_{ik} = s_{ik}^y \;\; \forall \;\; k \in Y, \\
& x_{il} - \sum_{j \in N} \lambda_{ji} x_{jl} = s_{il}^x \;\; \forall \;\; l \in X, \\
& \sum_{j \in N} \lambda_{ji} = 1, \\
& s_{ik}^y, s_{il}^x, \lambda_{ji} \geq 0, \;\; \forall \;\; k \in Y, l \in X, j \in N \Big\}
\end{aligned}
\qquad \textbf{(1)}
$$

[9] Kneip, Park and Simar (1996) uses a different notation, and defines the inefficiency distribution in terms of polar coordinates and the modulus of the input vector rather than γ.

[10] The need for two stages to get meaningful slacks will become apparent in section 5 where the slacks are used to adjust the efficiency measures. This methods avoids using the non-Archimedian formulation to force a slope on the extreme facets.

which corresponds to the formulation of Charnes et al. (1985a). The constraints of this LP-formulation define the DEA variable returns to scale (VRS) feasible technology which is the minimum extrapolation set satisfying a) strong disposability of inputs and outputs, b) convexity, and c) that all observations are in the set. Define the VRS reference set as

$$V = \left\{ i \in N \mid \bar{s}_i = 0 \right\} \tag{2}$$

i.e. the set of EDUs that have no slack in the solution of (1)[11]. The input saving technical efficiency estimator $\hat{E}_1$ is then[12]:

$$\begin{aligned} \hat{E}_{1i} = \underset{\theta_i,\lambda}{Min} \Big\{ & \theta_i \ , \ s.t. \\ & \sum_{j \in V} \lambda_{ji} y_{jk} - y_{ik} = s_{ik}^{y} \ \forall \ k \in Y, \\ & \theta_i x_{il} - \sum_{j \in V} \lambda_{ji} x_{jl} = s_{il}^{x} \ \forall \ l \in X, \\ & \sum_{j \in V} \lambda_{ji} = 1, \\ & s_{ik}^{y}, s_{il}^{x}, \lambda_{ji} \geq 0, \ \forall \ k \in Y, l \in X, j \in V \Big\} \end{aligned} \tag{3}$$

This radial efficiency estimate $\hat{E}_1$ is the minimum proportion of input usage that could feasibly produce the same amount of outputs given the DEA estimate of the feasible technology set. (3) corresponds to the formulation in Banker, Charnes & Cooper (1984) with the reference set V replacing the full set of observations N. Figure 1 shows the technical efficiency measure $\hat{E}_1$ in input-output space together with the scale efficiency estimators $\hat{E}_3$ and $\hat{E}_4$, and the scale indicator Λ.

The technical productivity or overall scale efficiency estimator $\hat{E}_3$ is the proportional input

---

[11] Note that the actual value of the sum of slacks $\bar{s}_i$ in (1) is of no meaning except when zero. It is not used for any other purpose than to select the set of efficient units. There may in fact be efficient observations in the set V that can be expressed as convex combinationsof other efficient units in V. This could introduce non-uniqueness in the slacks of (3), but the reference points will still be on the efficient frontier (see Charnes, Cooper & Thrall, 1991).

[12] Following the terminology of Førsund & Hjalmarsson (1987). Details of the two-stage approach and calculation of estimators $\hat{E}_3$ and $\hat{E}_4$ can be found in Kittelsen and Førsund (1992).

usage that is necessary to produce the given output had this output been produced by a unit at the optimal scale. The input saving implied by $\hat{E}_3$ is therefore only feasible if the scale of a unit can be changed without changing input or output mix or alternatively if the underlying technology of the sector exhibits constant returns to scale (CRS). $\hat{E}_3$ is estimated by an LP-formulation in a manner parallel to (1)-(3), except that the restriction on the sum of λ's is dropped. Instead this sum $\Lambda=\Sigma_j\lambda_{ji}$ is defined as the scale indicator that, if unique, is the units proportion of optimal scale after any input inefficiency has been eliminated. The reference set for this LP-problem is named *C*.

The sensitivity of the technical efficiency estimates has been addressed with operations research methods by Charnes et al. (1985b) and Thompson, Dharmapala and Thrall (1994). Their method is suitable to assessing the change in $\hat{E}_1$ when the value of the variables for the measured unit changes, and to calculate the intervals for the variables that will measure a unit as efficient. They do not analyse the sensitivity of the average efficiency or the frontier as a whole, nor the sensitivity of results to changes in the set of inputs and outputs.

Olesen and Petersen (1991, 1996) have discussed the problems of collinearity and show how this can lead to the existence of less than full dimensional facets on the DEA frontier and locally fixed proportions (Leontieff technology). They suggest ways of identifying the presence and extent of non full dimensional facets, but offer no test of the seriousness of this phenomenon. Their preferred remedy is to aggregate, but they also suggest the alternative of excluding from the reference set those efficient units that do not support a full dimensional facet.

Färe & Lovell (1987) have looked at the theoretical conditions under which aggregation will not change efficiency estimates. They find that homothetic separability of the technology is sufficient to ensure that the estimated $\hat{E}_1$ will not change with aggregation, and that aggregation therefore is admissible. In the same volume Färe and Primont (1987) show that if an input in one model is an aggregation of inputs in another model, the measured $\hat{E}_1$ in the

disaggregated model will be greater or equal than the measured $\hat{E}_1$ in the aggregate model. This follows from the principle that the optimal value of a minimized variable can never become lower if a restriction is added to the optimisation problem. This principle also implies that a model in which a variable is included will give a measure of $\hat{E}_1$ that is as least as high as in the same model with the variable omitted, since including a variable is the same as adding an extra restriction to (1) and (3)[13]. In general therefore if model 0 is nested within model 1, then

$$\hat{E}^0_{1i} \leq \hat{E}^1_{1i} \leq 1 \qquad \textbf{(4)}$$

Testing wether a change in model specification significantly changes results can therefore be formulated as

$$H_0\colon E^0_{1i} = E^1_{1i}, \quad H_1\colon E^0_{1i} < E^1_{1i} \qquad \textbf{(5)}$$

Statistical tests have been few in the DEA literature. Valdmanis (1992) among others has used the Mann-Whitney rank-order test to compare the efficiency of public vs. not-for-profit hospitals and find the public hospitals significantly more technically efficient in seven out of ten different input-output specifications. While her approach is fruitful in assessing the performance of separate groups and demonstrates the robustness of results across specifications, her method does not give an answer to the question of which specifications is best.

If two specifications are not nested, the Wilcoxon signed-rank[14] test is applicable since this test uses the information contained in the fact that two efficiency measures are calculated for the same units. If the models are nested however, this test is meaningless as the sign of any change in $\hat{E}_1$ is the same for all units.

---

[13] Diewert and Mendoza (1996) has an informal discussion of some of these rankings, using the term Le Chatelier Principles. See the appendix of Kittelsen (1997) for a formal treatment. Thrall (1989) shows how a unit with an estimated efficiency of 1, cannot be measured as inefficient when the number of variables increases (Theorem T1), allthough the unit could be moved in or out of the set V defined in (2).

[14] See e.g. Lehmann (1975) on the method. Magnussen (1996) applies this in a DEA model.

For nested models one could calculate Pearsons correlation coefficient or Spearmans rank correlation coefficient between $\hat{E}_1^0$ and $\hat{E}_1^1$ that would reflect the extent of change in the ordering of the units by efficiency scores. There are two problems with this approach. First, under the null hypothesis of no change in the ordering, a test would need to show whether the correlation coefficient is significantly less than one. Although a Fischer transformation can be used to construct a confidence interval around a central tendency in the open interval (-1,1), to this author's knowledge no method has been suggested that can test the extreme value of one. Secondly, even if the correlation is near-perfect the efficiency measures can have changed considerably if all units change proportionately or in a way that does not change the ranking.

Banker (1993) has suggested two asymptotic test of the significance of a difference in efficiency estimates between two models that are not nested, provided that the modeller is prepared to assume a specific shape of the distribution of the underlying inefficiency. Since the DEA efficiency estimates are confined to the interval (0,1] and the common forms of one-sided distributions are usually defined in the region [0,∞), the tests are based on estimates of the inefficiency parameters γ rather than the efficiency estimates,

$$\hat{\gamma}_i = \left(\frac{1}{\hat{E}^i} - 1 \right) \qquad \textbf{(6)}$$

Banker has described tests based on two specifications of the underlying distribution of γ: The halfnormal distribution and the exponential distribution. The null hypothesis that two estimates of inefficiency $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are independently drawn from the same underlying distribution (i.i.d.) is testable by means of a statistic that is Fisher's F-distributed. If the underlying distribution is halfnormal the statistic

$$F^H = \frac{\left(\sum_{j=1}^{N_0} \hat{\gamma}_0^2 \right) / N_0}{\left(\sum_{i=1}^{N_1} \hat{\gamma}_1^2 \right) / N_1}, \; \nu_1 = N_0, \; \nu_2 = N_1 \qquad \textbf{(7)}$$

will be F-distributed with N_0 and N_1 degrees of freedom. The F-distribution is constructed as the ratio of the estimates of the variances in two normal distributions (Johnson and Kotz, 1970). If the two variances are the same under the null hypothesis the expected value of F is one. In the halfnormal distribution the mean square of the variable plays the same role as the variance of the normal distribution.

If the underlying distribution is exponential the statistic

$$F^{E} = \frac{\left(\sum_{j=1}^{N_0} \hat{\gamma}_0\right)/N_0}{\left(\sum_{i=1}^{N_1} \hat{\gamma}_1\right)/N_1}, \ \nu_1 = 2N_0, \ \nu_2 = 2N_1 \qquad \textbf{(8)}$$

will correspondingly be F-distributed with $2N_0$ and $2N_1$ degrees of freedom.

The assumption that the underlying distribution is known is a strong one. If no parametric assumptions are maintained about the inefficiency distributions, Banker (1993) further suggests using a Kolmogorov-Smirnov type of nonparametric test of the equality of two distributions. Applied to the distributions of i.i.d. efficiency estimates, and denoting the estimated cumulative distribution functions as $\hat{S}_0(E), \hat{S}_1(E)$, the statistic

$$D^{+} = \mathrm{Max}_{\mathrm{E}}\left\{\hat{S}_0(E) - \hat{S}_1(E)\right\} \qquad \textbf{(9)}$$

is asymptotically distributed with a rejection probability of

$$\Pr\left(D^{+} > \left[\frac{N_0 N_1}{N_0 + N_1}\right]^{-1/2} z\right) = e^{-2z^2}, \ z > 0 \qquad \textbf{(10)}$$

which makes it applicable for testing one-sided hypotheses (Johnson and Kotz, 1970).

The ordinary T-test for comparison of group means also avoids the problem of specifying the underlying distributional shape of the inefficiencies. The T-statistic has the form:

$$T = \frac{\text{Mean}_i(\hat{E}_i^1) - \text{Mean}_i(\hat{E}_i^0)}{\sqrt{\frac{N^1 \text{VAR}(\hat{E}_i^1) + N^0 \text{VAR}(\hat{E}_i^0)}{N^1 + N^0 - 2}\left[\frac{1}{N^1} + \frac{1}{N^0}\right]}} \quad \textbf{(11)}$$

which, if the efficiency sample means are i.i.d. normal, is T-distributed with N^0+N^1-2 degrees of freedom. By the central limit theorem the sample means will be approximately normal unless the sample size is very small[15].

This gives us four suggestions for methods of accepting or rejecting a hypothesis that two distributions of efficiency measures are drawn from the same underlying distribution. Since DEA estimates are generally not independently and identically distributed (i.i.d.), there are theoretical problems with all four tests. All tests assume independence between all observations of $\hat{E}_1$ (or $\hat{\gamma}$), which is obviously not fulfilled for *nested* models if all observations are included in the calculations under both the null and the alternative hypothesis. There will then be a strong dependence resulting from measuring efficiency for the same observations under both models. In addition, there is a weak dependence between estimated efficiencies, since they can be calculated relative to the same referencing observations. This weak dependence will diminish as the sample size increases.

Furthermore, the estimators of nested models are not identically distributed since, as is shown in (4) above, by adding an extra restriction the model specification itself makes the bias of a model greater than the bias of a model that is nested within it. As both estimators are consistent, this effect should diminish as sample size increases. Finally, unlike the Wilcoxon signed-rank test or the correlation coefficients, the tests above do not use the information that we have two (paired) efficiency measures for each and every observed unit, and will therefore not be as powerful as a test that uses this information.

---

[15] According to Bhattacharyya and Johnson (1977, p.295), the approximation is valid if the sample sizes are greater than 25 or 30.

Kittelsen (1997) finds in Monte Carlo simulations with different model specifications that although these tests are in no way perfect, there are some substantive findings that should give grounds for conclusions in empirical work. Generally, the dependence of the two estimates will tend to decrease the value of the test statistics, leading to under rejection of a true null hypothesis, and loss of power in rejection of a false null. The difference in bias between nested DEA models will however work in the opposite direction, partially or wholly offsetting the effect of dependence. Unless relative bias is very large the tests will not be improved by removing dependence, so full sample tests are better even if the models are nested.

The bias-effect seems greater for the t-test than for the F and D^+ tests. Thus, if bias is low due to e.g. large samples, low dimensionality and/or inefficiency distributions that are dense near the frontier, the T-test is quite useful. If, however, bias is somewhat higher, due to medium sample size, higher dimensionality and inefficiency distributions that have their mode away from the frontier, the F-tests and the Kolmogorov-Smirnov test seem best. There is not much difference between these. Finally, if bias is expected to be very high due to sample size less than about 100, none of the tests will be very good.

Having established a significance test for nested models one can parallel the stepwise regression approach[16] of standard econometrics and 1) Set up a basic model that contains all variables that for prior theoretical or empirical reasons must be in the model, 2) compute $\hat{E}_1^0$ for each unit, 3) compute $\hat{E}_1^1$ for each candidate variable and candidate disaggregation, 4) compute F^H , F^E , D^+ and T for each candidate. If there are one or more candidates with a significant statistic, set up a new basic model by accepting the most significant candidate, and 5) repeat from 2) until there are no more significant candidates. If the different tests give different inclusion decisions, one must examine the sample size, dimensionality and efficiency distributional forms with a view to judge the importance of relative bias.

[16] Norman & Stoker (1991) have used a stepwise approach to variable disaggregation in DEA, but use the simple Pearson correlation coefficient between variable value and efficiency measure as the criterion for choosing what disaggregation to perform. This coefficient is, however, not partial enough, since variables may well be collinear and the excluded variable can contain essentially the same information as one that is already included. There is no reason to presuppose that an excluded variable is not correlated with efficiency. Additionally these authors offer no significance test to decide when to stop disaggregating.

Table 1: *Theoretical critical values of the test statistics.*

Significance level		F^H	F^E	D^+	T
	d.f.	172,172	344,344	172	342
1%		1.428	1.286	0.164	2.337
5%		1.286	1.194	0.132	1.649

3 The data

Of the 281 EDUs registered in the electricity statistics (Statistisk sentralbyrå, 1991) there are 185 EDUs that are not also engaged in transmission. The data has been thoroughly checked with alternate sources and 13 EDUs have been excluded since reliable data were unavailable, leaving 172 EDUs in the cross-section sample with data for 1989. Table A.1 in the appendix shows how the basic 13 variables have been constructed and table A.2 gives descriptive statistics for the EDUs in the sample.

The inputs in the model must contain all known resource usage, although the aggregation level can be determined by the analysis. Data are readily available at a detailed level for lines and transformers in calculated replacement values. The power loss data are probably flawed for a single year so a truncated moving average is used instead. The data on labour hours and goods and services cannot be disaggregated. In the basic model in the next section aggregate capital I34 is used, but disaggregation into I3 transformers and I4 lines is entered as a candidate. The other input variables are in the basic model since they cannot reasonably be aggregated or excluded.

On the output side energy delivered P234 is often thought of as the only and homogeneous product of the electricity distribution sector. In principle, however, the EDUs deliver energy

separately to distinct customers and locations, each having a vector of qualitative characteristics such as voltage, a time profile (power vs energy), geographical distance from the power source and climatic differences. To capture these aspects of the aggregate production of the EDU, other variables are included as candidates. These include environmental constraints, maximum power and the number of customers. Data on voltage level are not available, but as a proxy disaggregation of energy delivered by institutional category is offered as candidates. Data on power interruption (quality) are only available for 40 EDUs, but in fact show no correlation with efficiency measures for these units.

4 The stepwise results

Since there is no prior information on the shape of the actual efficiency distribution in the electricity distribution sector, the stepwise iteration will continue until no candidate is significant at the 1% level for any of the test statistics. In the sample there are 172 observations and Table 1 shows the theoretical critical values for each statistic.

Table 2: *Model I. Variables and disaggregations that are significant at the 1%-level are marked **, and 5%-level are marked *. Variables included i the next model are shaded.*

In the model		Mean $\hat{E}_I$	S_I	Number of reference units	F^H	F^E	D^+	T
P234	Energy delivered							
I1	Labour hours							
I2	Energy loss	0.6138	0.5167	17				
I34	Capital							
I5	Goods and services							
New variables		**Change**						
R1	Distance index	0.1174	0.1144	14	**2.924	**1.769	**0.349	**5.916
R2	Corrosion index	0.0757	0.0913	10	**1.831	**1.399	**0.227	**3.655
R3	Climatic index	0.1209	0.0781	15	**2.408	**1.705	**0.314	**5.861
P1	Maximum power	0.0513	0.0458	4	*1.420	*1.238	**0.169	**2.463
P5	Number of customers	0.1320	0.1013	8	**3.174	**1.901	**0.366	**6.770
Disaggregation								
P2 P34	Energy EDUs & en.-intens. industry Energy other industry & others	0.0575	0.0773	6	**1.487	*1.270	**0.180	**2.749
P23 P4	Energy EDUs & industry Energy others	0.0986	0.0925	13	**2.062	**1.545	**0.291	**4.813
I3 I4	Transformers Lines	0.0457	0.0000	4	**1.491	*1.238	**0.180	*2.240

Since the purpose of the analysis is to provide grounds for regulating the EDUs, including a candidate whose significance is in doubt rather than to excluding it is detrimental to the electricity customers, but beneficial to the EDUs. Disaggregating or including variables leads to higher efficiency estimates, and if regulation takes account of such estimates this lessens the probability of bankruptcy due to regulations that do not take into account all relevant environmental constraints and production aspects. In addition to the statistical efficiency consequences of the model specification, there are also real economic consequences. If the interests of the EDUs are better organised than those of the consumers, there might well be reasons for choosing test and significance level to counterweigh an inclination to blow up the model dimension.

Table 3: *Model II. Variables and disaggregations that are significant at the 1%-level are marked **, and 5%-level are marked *. Variables included i the next model are shaded.*

In the model		Mean $\hat{E}_l$	S_l	Number of reference units	F^H	F^E	D^+	T
P234	Energy delivered							
P5	Number of customers							
I1	Labour hours	0.7458	0.6180	25				
I2	Energy loss							
I34	Capital							
I5	Goods and services							
New variables		**Change**						
R1	Distance index	0.0633	0.0436	15	**1.944	**1.462	**0.180	**3.670
R2	Corrosion index	0.0419	0.0291	11	**1.480	1.263	0.128	**2.356
R3	Climatic index	0.0204	0.0000	8	1.103	1.092	0.076	1.110
P1	Maximum power	0.0070	0.0000	1	1.051	1.034	0.035	0.389
Disaggregation								
P2	Energy EDUs & en.-intens. industry	0.0098	0.0070	3	1.055	1.044	0.052	0.538
P34	Energy other industry & others							
P23	Energy EDUs & industry	0.0148	0.0032	5	1.103	1.071	0.076	0.788
P4	Energy others							
I3	Transformers	0.0220	0.0000	7	*1.294	1.140	0.070	1.240
I4	Lines							

Table 2 shows the results of the first iteration (Model I)[17]. Average technical efficiency is very low at 61% and the structural efficiency[18] just over 51%, although a dimension of 5 is already larger than most DEA models. In addition only 17 units are in the reference set V and deemed efficient.

At this stage all candidates are significant at the 5% level, and most candidates at the 1% level of all four statistics, suggesting that all candidate variables and disaggregations contain relevant information and would result in a significantly different efficiency distribution and better "goodness of fit." Even so the distance index R1, the climatic index R3 and the number

[17] The computations have been carried out with the GAMS modelling language (Brooke, Kendrick and Meeraus, 1988).

[18] The efficiency of the average unit, introduced by Farrell (1957) and formalised in Førsund & Hjalmarsson (1987). The structural efficiency measures the proportion of inputs needed, had one been able to restructure the whole industry into an unchanged number of identical units. See also Kittelsen and Førsund (1992).

of customers P5 are clearly more significant than the other candidates with e.g. F^H above 2.4. P5 has the highest value for all statistics and would alone increase mean $\hat{E}_1$ by 13 percentage points. Note however that four of the other candidates give larger changes in the number of reference units. This highlights the fact that this change is *not* a good indicator of which candidate to accept; On the contrary, an increase in the number of efficient units for these candidates could well mean that more units will be deemed efficient because they are extreme in some dimension, since they do not affect the average efficiency to the same extent.

P5 is therefore included in Model II as a product in addition to energy delivered P234. With an efficiency measure of 75% and 25 efficient units, few other candidates are significant as table 3 shows. Maximum power P1 would only change $\hat{E}_1$ by less than 1%-point and has no significant *F,D* or *T*-values. This is a consequence of the high correlation between P5 and P1 that in table A.3 in the appendix can be seen to be 0.82. This result indicates that information in the number of customers P5, which primarily consists of households, in combination with the already included total energy P234 is a better description of the long-run differences between EDUs in the need for capacity, and thereby input usage, than the maximum power in the single year 1989. By the same reasoning this information also eliminates the need for disaggregating total energy P234 by institutional sector.

The climatic index R3 also loses most of its explanatory power when the number of customers P5 is included. This is mainly due to the way R3 is constructed as the "average climatic difficulties" multiplied by the number of customers to get an indicator of the "total climatic difficulties." Thus, the independent information contained in R3 is not sufficient for it to make a significant impact on the estimated level of technical efficiency. The fact that R3 had a significant impact in table 2 is a consequence of it containing much of the same information as P5 that was not included in Model I.

In table 3 the corrosion index R2 is a significant candidate by the F^H and *T* statistics, but the distance indicator R1 is clearly the strongest candidate for inclusion in the model. Again the three tests give no ambiguity on which candidate should be accepted, but although they exceed the theoretical 1% levels, they are not very high. Kittelsen (1997) reports overrejection rates when variables are not correlated, but the R1 variable is in table A3 in the

Table 4: *Model III. Variables and disaggregations that are significant at the 1%-level are marked **, and 5%-level are marked **.*

In the model		Mean $\hat{E}_I$	S_I	Number of reference units	F^H	F^E	D^+	T
R1	Distance index							
P234	Energy delivered							
P5	Number of customers							
I1	Labour hours	0.8091	0.6615	40				
I2	Energy loss							
I34	Capital							
I5	Goods and services							
New variables		**Change**						
R2	Corrosion index	0.0065	0.0003	4	1.072	1.043	0.029	0.400
R3	Climatic index	0.0204	0.0000	11	1.155	1.129	0.076	1.245
P1	Maximum power	0.0060	0.0000	1	1.061	1.038	0.035	0.369
Disaggregation								
P2	Energy EDUs & en.-intens. industry	0.0080	0.0104	2	1.068	1.049	0.041	0.491
P34	Energy other industry & others							
P23	Energy EDUs & industry	0.0114	0.0015	3	1.113	1.076	0.076	0.703
P4	Energy others							
I3	Transformers	0.0171	0.0070	7	1.261	1.132	0.064	1.069
I4	Lines							
Constant returns to scale		0.0410	0.0110	13	**1.521	*1.282	*0.134	**2.431

appendix reported as being highly correlated with the other included variables. In view of this, it is fairly clear that the distance indicator R1 should be included. The effect of including R1 is about 6 %-points on average efficiency and results in 15 new efficient units.

In Model III the distance indicator is included, giving an average estimated technical efficiency of 81% and 40 reference units. Table 4 shows that there are no candidates that are significant. When R1 is included in Model III, the significance of the corrosion index R2 disappears. Again this is a consequence of the construction of this index as an "average corrosion exposure" per km of lines multiplied by the distance to be covered represented by R1. The table shows that the independent information contained in R2 does not significantly change technical efficiency measures. This does not mean that corrosion is not a problem that demands resources for the EDUs. Corrosion implies shorter lives for lines and therefore higher rates of depreciation. In a cost efficiency analysis this would enter through a higher user cost of capital. The present results indicate that corrosion does not imply the need for

significantly higher levels of capital or other inputs although investment would be affected.

Since I3 and I4 contains mainly different information than the output variables that have been included in previous stages, this candidate's *F*, *D* and *T*-values have been fairly stable through the iterations, but is no longer significant at the 5% level of any of the tests. Note also that many of the statistics of the remaining five variables have increased from table 3 to table 4. This demonstrates the point that on the whole significance for remaining candidates go down for each candidate that is included, but not consistently. Even if no other variables had been significant in table 3, it would still have been necessary to test the remaining candidates against Model III.

Table 4 shows also the test statistics of a constant returns to scale model specification, in which case CRS is the null hypothesis nested within the VRS model. A CRS feasible set estimate would lower average estimated efficiency by 4.1%-points and make only 27 units efficient. CRS is rejected at the 5% level by all tests, and at the 1% level by two of the tests. Again, since the model has high dimensionality, the relative bias would tend to make the tests overreject, putting these results in an inconclusive range. By the criteria set up at the outset of this section, however, the rejection should stand.

The result of the stepwise DEA is thereby a VRS production specification with 7 variables instead of the 13 that were potentially to be included. A DEA analysis with all 13 variables included yields an average efficiency estimate that is 6 %-points higher than the preferred model, and this joint inclusion of 6 new variables gives statistic values of F^{H}=2.055, F^{E}=1.580, D^{+}=0.203 and T=3.953, all of which are significant at the theoretical 1% level. Since each rise in dimensionality gives rise to a further bias in efficiency estimates and a further tendency for the statistics to over-reject, and these figures are not very much higher than the theoretical critical values, it could be argued that a correct test would not reject Model III. This would mean that the information contained in the omitted variables jointly would not change results significantly. There are however equally important economic reasons why joint inclusion of these variables is not reasonable. As argued above, an inclination to blow up the number of variables could be detrimental to the consumers, and including a large number of irrelevant variables could potentially guarantee that all EDUs

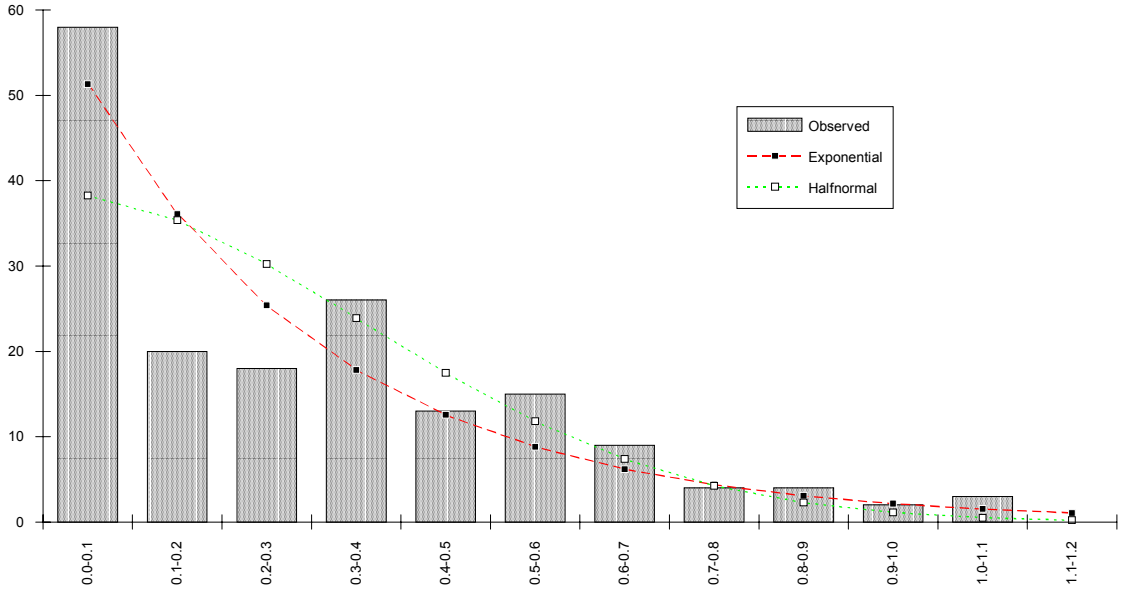


***Figure 2**: Actual distribution of γ in Model III compared to exponential and halfnormal distributions with the same expectation equal to 0.284.*

were measured as efficient.

At this stage one has information to check how well the assumptions on the underlying distributions correspond to the efficiency distribution given by Model III. Figure 2 shows the distribution of γ^{III} compared to the theoretical halfnormal and exponential distributions with the same expected value. The distributions far from coincide, and one sees that the observed distribution is even more skewed than the exponential distributions, which also is rejected by a Kolmogorov-Smirnov test[19]. As far as this carries over to the "correct" F-statistic, the exponential statistic F^E has more credibility than the halfnormal F^H-statistic, and the distribution-free D^+ and T tests are even more preferable. As one can see from the tables this would not have changed the conclusions.

5 The efficiency results

[19] The adjusted test statistic $D\sqrt{n}$ of 3.120 compares to a theoretical adjusted value of 1.308 at the 1% level for a 2-sided test.

***Table 5**: Main efficiency results. Model III.*

		Average	Standard deviation	Min.	Max.	Weighed average $\hat{I}_i$	Structural efficiency $\hat{S}_i$
$\hat{E}_1$	Technical efficiency	0.809	0.151	0.481	1.000		0.662
$\hat{E}_3$	Overall scale efficiency	0.768	0.160	0.405	1.000		0.651
$\hat{E}_4$	Pure scale efficiency	0.948	0.069	0.638	1.000		0.984
Λ	Scale indicator	0.807	0.746	0.021	4.299		0.561
Input-specific technical efficiency for:							
$\tilde{E}_{1;i1}$	Labour hours	0.791	0.164	0.327	1.000	0.810	0.662
$\tilde{E}_{1;i2}$	Energy loss	0.767	0.179	0.246	1.000	0.782	0.662
$\tilde{E}_{1;i34}$	Capital	0.776	0.170	0.352	1.000	0.789	0.662
$\tilde{E}_{1;i5}$	Goods and services	0.781	0.175	0.288	1.000	0.782	0.662

Table 5 summarises the main results for Model III in more detail. The technical efficiency estimate $\hat{E}_1$ is on average 0.809, but varies from 0.481 to 1.0. Figure 3 shows the distribution of $\hat{E}_1$ in a Salter-diagram where the inefficient units are sorted by increasing efficiency and the width of each bar is proportional to energy delivered P234. Since there are several outputs in the model there is no clear geometric interpretation of the figure, but the area above the bars as a share of the total area of the graph can be taken as an approximation of the efficiency loss in the sector. The figure shows that there is no strong correlation between size and efficiency, although the 5 largest units are also technically efficient. Even if only 40 of the 172 units are efficient, around 2/5 of the energy is delivered by efficient EDUs. Another feature of the figure is the fairly smooth fall in efficiency to the left; there is no clear mode among the inefficient units.

While the technical efficiency estimate for each unit is interpretable as an indicator of the radial or proportional input saving potential (i.e. 1-$\hat{E}_1$), the presence of slack in the LP-model

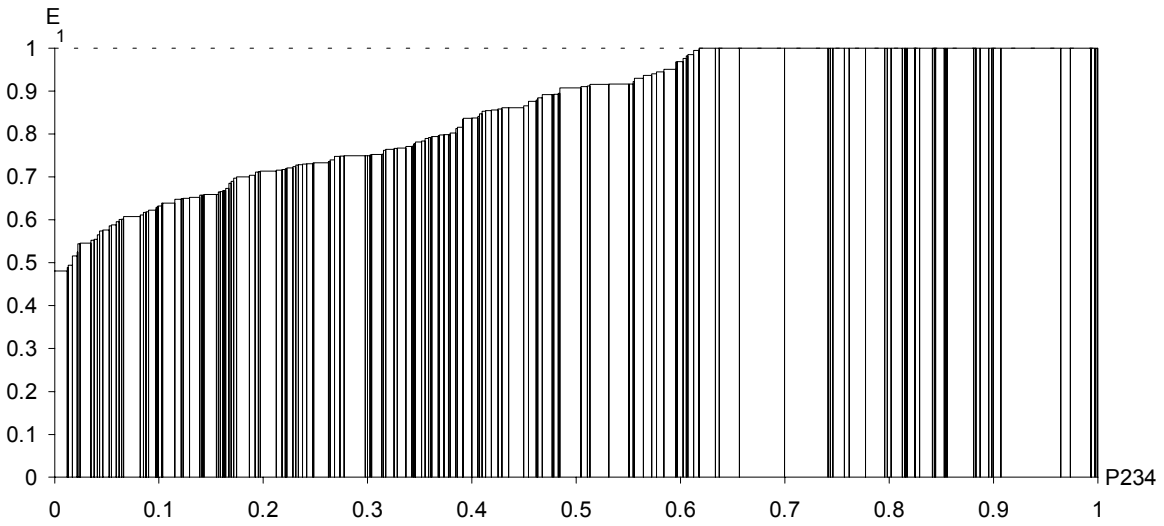


***Figure 3**: Salter diagram: The accumulated technical efficiency estimate $\hat{E}_1$ weighted by total energy delivered P234.*

(3) implies that further input reduction is possible in some dimensions. Following Torgersen et al. (1996) one can for each input define the input-specific technical efficiency as:

$$\tilde{E}_{1li} \equiv \frac{\hat{E}_{1i} x_{il} - s_{il}^{X}}{x_{il}}, \ \forall \ i \in N, \ \forall \ l \in X. \qquad \textbf{(12)}$$

These are reported in the bottom half of table 5, where the unweighed averages show that there is a degree of slack in I2 Energy loss and I5 goods and services, but less in I1 labour hours and I34 capital.

To construct weighted averages one must have a unique and meaningful variable to use as a weight. While there is no general size variable to weigh the main measures, the input-specific technical efficiency measures $\tilde{E}_{1l}$ can be weighed as in Torgersen et al. (1996) using the input in question to construct input-specific measures of industry efficiency:

$$\hat{I}_{1l} \equiv \frac{\sum_{i \in N} \tilde{E}_{1il} \ x_{il}}{\sum_{i \in N} x_{il}}, \ \forall \ l \in X. \qquad \textbf{(13)}$$

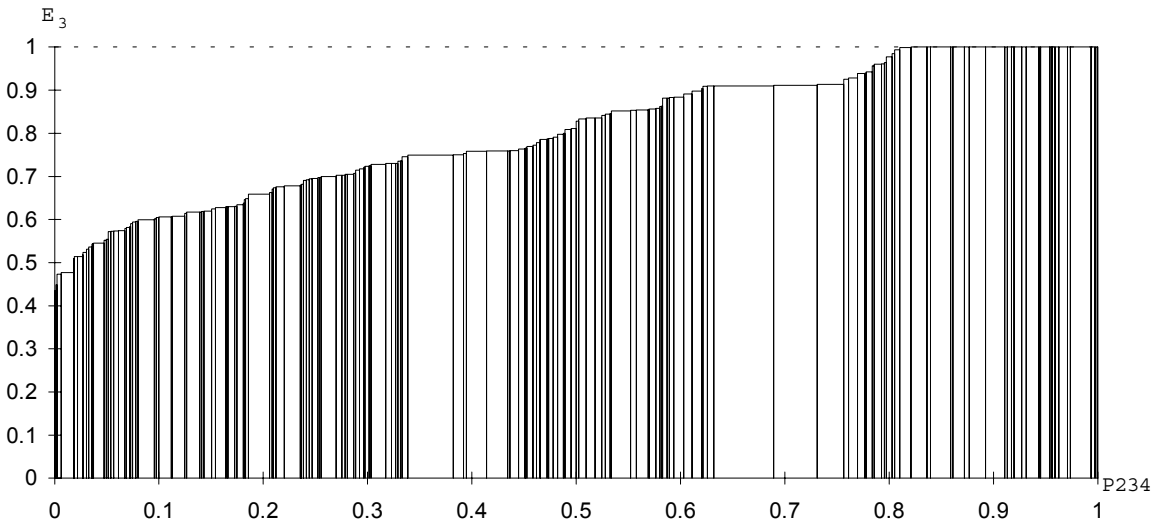


***Figure 4**: Salter diagram: The accumulated overall scale efficiency estimate $\hat{E}_3$ weighted by total energy delivered P234.*

that are also reported in table 5. (1-$\hat{I}_{1l}$) is directly interpretable as the input saving potential for each input if all units were to become efficient while output-levels are maintained. Since these weighted averages are greater than the unweighed ones for all inputs, the efficiency is slightly higher for the large EDUs than for the smaller firms.

As pointed out when testing for returns to scale in the previous section, the rejection of CRS is not very clear cut. Overall scale efficiency estimates $\hat{E}_3$ is on average 0.768 which is not very different from the technical efficiency estimates. Indeed the pure scale efficiency measure $\hat{E}_4$ is as high as 0.948, showing that the scale curvature of the production function is not strong. Figures 4 and 5 show the distribution of the scale efficiency measures. Here the largest EDUs are clearly scale inefficient and among the EDUs with the lowest pure scale measure $\hat{E}_4$. The medium size firms are spread throughout the figures, but the smallest firms are most numerous in the central part of figure 5. This means that their generally low overall scale efficiency is mainly due to the technical inefficiency of figure 3 rather than the pure scale inefficiency of figure 5.

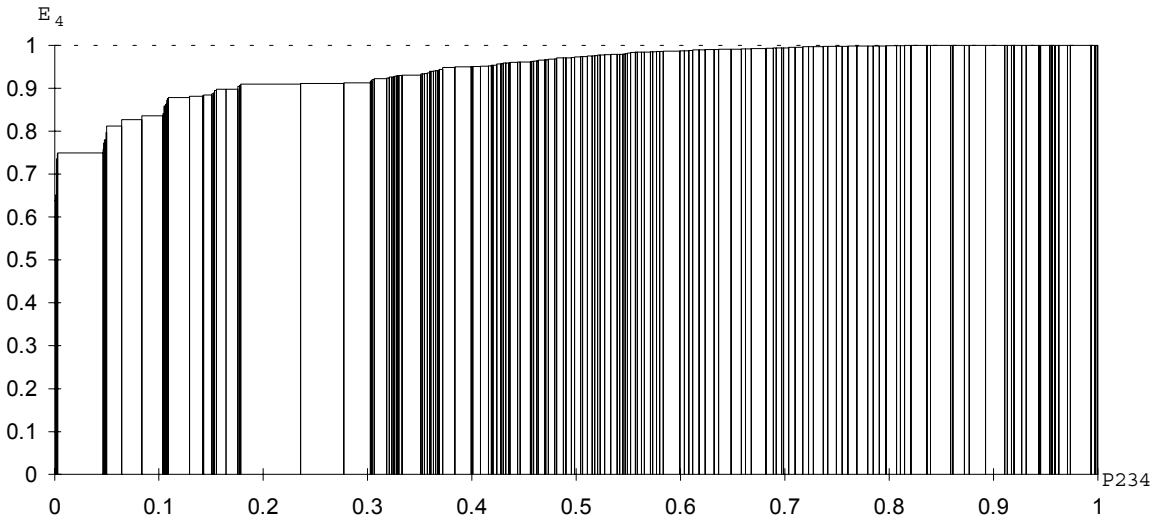


Figure 5*: Salter diagram: The accumulated scale efficiency estimate* $\hat{E}_4$ *weighted by total energy delivered P234.*

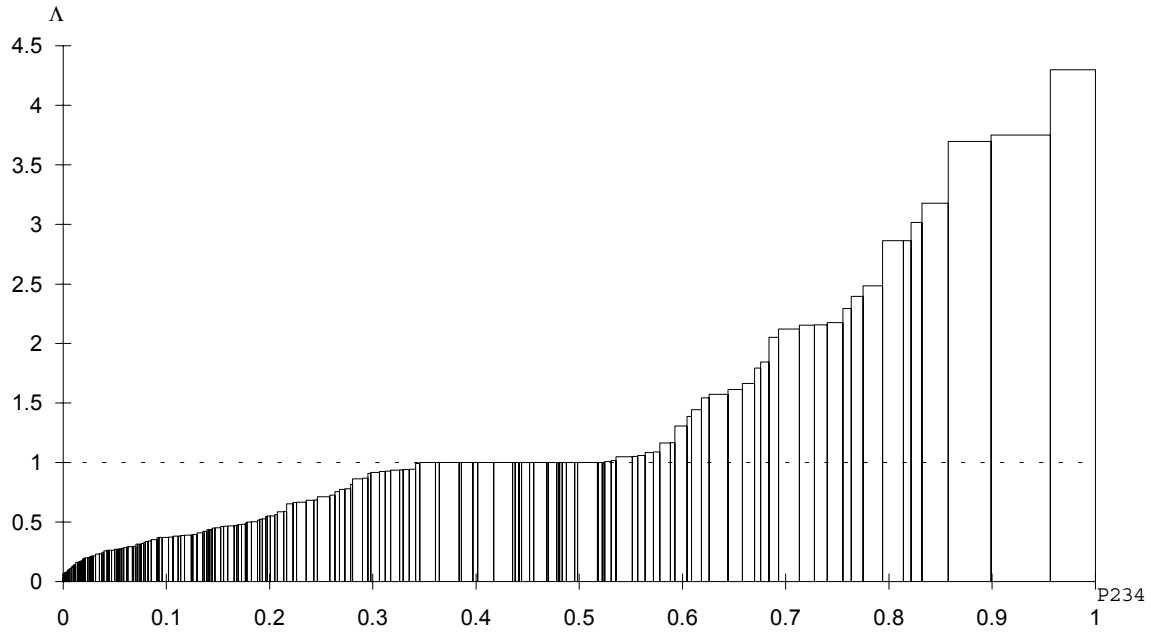


***Figure 6**: Salter diagram: The accumulated distribution of the scale indicator Λ weighted by total energy delivered P234.*

Figure 6 shows the distribution of the scale indicator Λ. Although proportion of optimal size increases with energy delivered, this is not consistently true. This is because optimal scale is direction-specific in that it depends on the input and output-mixes of each unit. The figure shows units that are as small as 1/40 of optimal size and others that are more than 4 times as large as optimal size. Though this range is large, it in no way reflects the productivity loss through wrong scale, which is captured by $\hat{E}_4$.

This point is also emphasized in figure 7 which shows the intersection of the frontier with a plane through the average EDU. Here one can see clearly how the frontier is nearly CRS in a considerable range, and only for the largest units reveals decreasing returns to scale. In table 5 the structural measures reflect the estimates for the average unit, and is a measure of the efficiency of the sector compared to a hypothetical restructuring into identical units (Farrell ,1957, Førsund & Hjalmarsson, 1987). The average EDU has a pure scale efficiency estimate of 0.984. The optimal scale in this direction is nearly twice the size of the average, but considering the linearity of the frontier in this region optimal scale should probably be interpreted as an interval. Even for the very smallest sizes the VRS frontier is very near CRS,

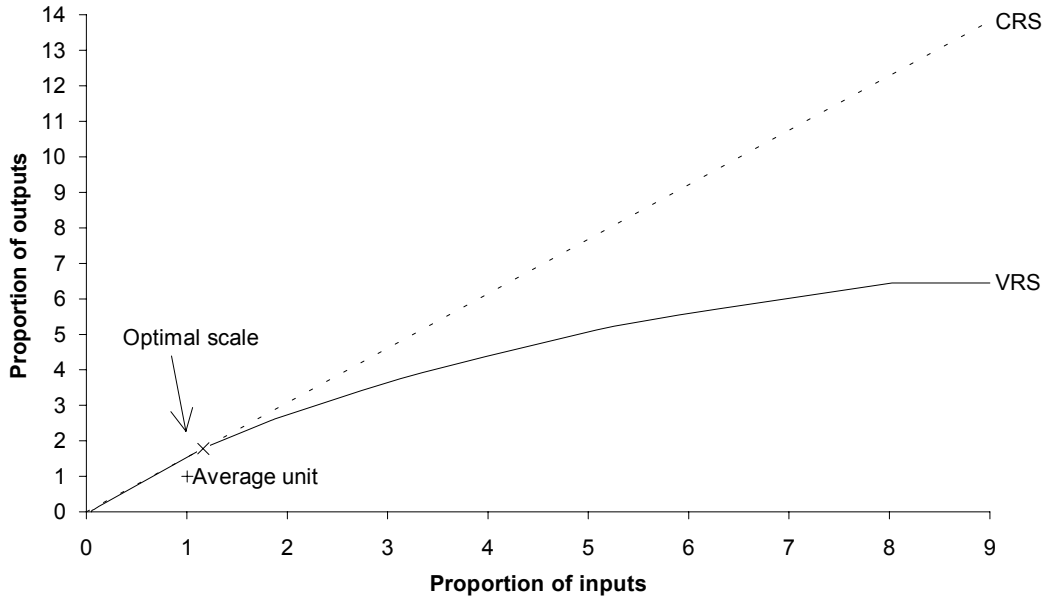


***Figure 7**: The efficient frontier in a plane defined by the average input mix and output mix.*

which is why their suboptimal size is less of a problem than their technical inefficiency.

For the largest sizes this is reversed; their excess size is the prime source of their overall inefficiency as measured while there is less to be gained from improving technical efficiency. There are clearly decreasing returns to scale above approximately 3 times the average size. For the very largest firms this could however be a quirk of the method since there are few firms of this size in the sample[20] and they are not sufficient to span out the frontier. In a sense they may be estimated as technically efficient by default, and therefore scale inefficient residually.

6 Conclusions

This paper has tried to show that there exists crude but usable methods of testing the overall goodness of fit. These tool can be used with reasonable results on nested models to decide the

[20] Several of the largest distributors in Norway are also engaged in transmission and are therefore excluded from the sample.

acceptable level of aggregation or the acceptable inclusion of variables. The tests lead to clear results on variable specification of a model of Norwegian electricity distribution. Tests of the scale properties are less conclusive, but constant returns to scale is just rejected. The EDUs have extensive technical inefficiency, with average technical efficiency estimates in the order of 80% and individual estimates ranging from about 50% to 100%. Scale inefficiency is much less of a problem.

By extension other nested specification issues should also be addressable in a similar fashion, such as the strong free disposal assumption. There are drawbacks to these methods; notably that due to the estimate biases and dependence these tests are only approximations, and the fact that the methods do not use the information contained in the pairing of efficiency measures for each unit and are therefore not very powerful. For the *F*-tests the necessity of assuming a specific shape for the efficiency distribution is also a problem. The last of these issues is avoided by the D^+ and *T*-tests but resolving the others would imply a different line of attack, perhaps building on the work of Simar and Wilson (1995) or Korostelev et al. (1995a,b) . For the comparison of non-nested models such as separate groups or competing and exclusive specifications, these tests should be used in conjunction with relevant rank-order tests.

Appendix A: Descriptive tables

Table A.1: *List of variables*.

No	Name	Calculation, source and comments
Environmental Constraints		
R1	Distance index	Travelling time from each basic district in the national census to the municipal centre (Wilhelmsen & Lilleschulstad 1990), in weighed *) average for all basic districts, multiplied by number of customers P5. Includes both distance and travelling difficulties in one variable. *) Basic districts weighed by their population. Some districts lack population figures, and these are weighed by 600.
R2	Corrosion index	Index from 1.0 to 4.0 based on (40/years of depreciation) where the years of depreciation is estimated by EFI. Multiplied by distance index R1 to get an index that covers the necessary length of lines that are exposed to corrosion.
R3	Climatic index	The absolute value of average temperature in three coldest days of each year divided by the sum of temperatures for all days of frost. (Norwegian Meteorological Institute).Multiplied by number of customers P5 to get an index of the need for peak capacity compared to average capacity.
Products		
P1	Maximum power	in kW. Electricity statistics from SSB.
P2	Energy delivered to other electricity utilities and energy-intensive industry	in MWh. Electricity statistics from SSB.
P3	Energy delivered to other industry and commerce	in MWh. Electricity statistics from SSB.
P4	Energy delivered to others (households and agriculture)	in MWh. Electricity statistics from SSB.
P5	Number of customers	Electricity statistics from SSB.
Inputs		
I1	Labour hours	Electricity statistics from SSB.
I2	Energy loss	Electricity statistics from SSB. Truncated 5-year moving average, i.e. for 1989: (loss87+2*loss88+3*loss89)/6.
I3	Transformers	i 1000 kr values. Number and capacity of local transformers from the Electricity statistics from SSB, main transformers from a database at NVE, valuation by EFI..
I4	Lines	i 1000 kr values. Voltage level and length from the Electricity statistics from SSB, , valuation by EFI.
I5	Goods and services	i 1000 kr values. Electricity statistics from SSB.

Table A.1 shows the list of variables used in the analysis at the finest level of aggregation, with the source of the data. Aggregated variables are denoted P23=P2+P3 etc. The Electricity statistics are the raw data made available from the Central Bureau of Statistics (SSB), of which summary statistics are available in Statistisk sentralbyrå (1991). EFI is the Norwegian Electricity Distribution Research Institute that has helped in collecting and transforming data.

***Table A.2**: Descriptive statistics for the variables in the analysis.*

No	Name	Units	Average	St.deviation	Min.	Max.
Environmental Constraints						
R1	Distance index	min.	87228	89599	1432	608812
	Before adjusting for size (R1/P5), i.e. average travelling time	min.	16.108	13.812	4.596	92.487
R2	Corrosion index		163300	235898	2148	1826436
	Before adjusting for size(R2/R1), i.e. 40/years of depreciation		1.672	0.805	1	4
R3	Climatic index		34.195	38.243	1.098	261.816
	B*efore adjusting for size (R3/P5), i.e. "peak/average temperature"*		0.00525	0.00114	0.00280	0.00818
Products						
P1	Maximum power	kW	34336	39328	1320	296500
P2	Energy delivered to other electricity utilities and energy-intensive industry	MWh	11222	44356	0	400008
P3	Energy delivered to other industry and commerce	MWh	66972	93577	2498	770832
P4	Energy delivered to others (households and agriculture)	MWh	88069	103560	2993	868198
P5	Number of customers	number	6418	6958	268	51510
Inputs						
I1	Labour hours	hours	62547	62479	2300	402412
I2	Energy loss	MWh	11872	14670	527	117539
I3	Transformers	1000 kr	30155	44336	1181	400376
I4	Lines	1000 kr	164048	167086	8525	1486119
I5	Goods and services	1000 kr	10856	12858	372	97249

Table A.2 gives the unit of measurement and the average, standard deviation, maximum and minimum for each basic variable.

Table A.3: *Correlation matrix; Lower left triangle shows the simple Pearson correlation, while the upper right triangle shows the partial correlation.*

	R1	*R2*	*R3*	*P1*	*P2*	*P3*	*P4*	*P5*	*I1*	*I2*	*I3*	*I4*	*I5*
R1	1	0.770	0.322	0.269	0.046	0.020	0.029	-0.241	0.103	0.169	-0.233	0.343	-0.033
R2	0.882	1	-0.570	-0.224	-0.115	0.044	-0.114	0.386	-0.223	-0.058	0.096	0.046	0.059
R3	0.734	0.518	1	-0.020	0.041	-0.025	-0.271	0.737	-0.285	-0.168	0.137	0.016	0.066
P1	0.624	0.355	0.818	1	0.091	0.071	-0.068	0.147	0.441	0.034	-0.105	-0.299	-0.165
P2	0.262	0.080	0.309	0.419	1	0.129	-0.234	-0.021	0.040	0.390	0.171	-0.019	-0.001
P3	0.793	0.648	0.935	0.769	0.320	1	0.201	0.205	-0.163	0.103	0.233	-0.077	0.109
P4	0.783	0.645	0.950	0.760	0.229	0.965	1	0.678	-0.178	-0.184	0.307	-0.040	0.137
P5	0.797	0.629	0.976	0.818	0.294	0.967	0.984	1	0.385	0.263	-0.356	0.176	0.012
I1	0.774	0.567	0.915	0.870	0.394	0.905	0.910	0.945	1	0.162	0.091	0.208	0.273
I2	0.771	0.587	0.842	0.790	0.532	0.872	0.839	0.877	0.895	1	0.287	-0.151	-0.090
I3	0.716	0.645	0.738	0.515	0.357	0.820	0.803	0.779	0.749	0.780	1	0.585	-0.005
I4	0.871	0.769	0.850	0.630	0.276	0.893	0.900	0.897	0.856	0.820	0.888	1	-0.070
I5	0.723	0.589	0.885	0.710	0.238	0.894	0.911	0.911	0.873	0.781	0.734	0.828	1

Table A.3 shows the matrix of simple Pearson correlation coefficients between each pair of variables in the lower left triangle. The upper right triangle shows the partial correlation coefficients between each pair of variables controlling for the linear effect of all other variables in the analysis. A correlation coefficient of absolute value above 0.214 is significant at the 1%-level and above 0.159 at the 5%-level.

References

Aigner, D.J. and S. Chu, 1968, On estimating the industry production function, *American Economic Review* 58, 826-839.

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

Banker, R.D., 1993, Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273.

Banker, R.D., A. Charnes and W.W. Cooper, 1984, Some models for estimating technical and scale inefficiencies, *Management Science* 30, 1078-1092.

Bauer, P.W., 1990, Recent developments in the econometric estimation of frontiers, *Journal of Econometrics* 46, 39-56.

Bhattacharyya,G.K. and R.A.Johnson, 1977, *Statistical concepts and methods*, (John Wiley & Sons, New York).

Brooke,A., D.Kendrick and A.Meeraus, 1988, *GAMS: A user's guide*, (The Scientific Press, Redwood City).

Charnes,A. and W.W.Cooper., 1985, Preface to topics in data envelopment analysis, *Annals of Operations Research*, 2, 59-94.

Charnes,A., W.W.Cooper and E.Rhodes, 1978, Measuring the efficiency of Decision Making Units, *European Journal of Operational Research* 2, 429-444.

Charnes,A., W.W.Cooper, B.Golany, L.Seiford and J.Stutz, 1985a, Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions, *Journal of Econometrics* 30, 91-107.

Charnes,A., W.W.Cooper, A.Y.Lewin, R.C.Morey, J.Rousseau, 1985b, Sensitivity and stability analysis in DEA, *Annals of Operations Research* 2, 59-94.

Charnes,A., W.W.Cooper and R.M.Thrall, 1991, A structure for classifying and characterizing efficiency and inefficiency in Data Envelopment Analysis, *Journal of Productivity Analysis* 2, 197-237.

Diewert, W.E. and M.N.F.Mendoza, 1996, The Le Chatelier Principle in Data Envelopment Analysis, Paper presented at the 2nd Georgia Productivity Workshop, University of Georgia, November 1996.

Farrell, M.J., 1957, The measurement of productive efficiency, *Journal of the Royal Statistical Society* 120, 449-460.

Färe, R., S.Grosskopf and C.A.K.Lovell, 1985, *The measurement of efficiency of production* (Kluwer-Nijhof, Boston).

Färe, R. and C.A.K.Lovell, 1978, Measuring the technical efficiency of production, *Journal of Economic Theory* 19, 150-162.

Färe, R. and C.A.K.Lovell, 1987, Aggregation and efficiency, in: Eichhorn, ed., *Measurement in Economics* (Physica-Verlag, Heidelberg).

Färe, R. and D.Primont, 1987, Efficiency measures for multiplant firms with limited data, in: Eichhorn, ed., *Measurement in Economics* (Physica-Verlag, Heidelberg).

Førsund, F.R. and L. Hjalmarsson, 1974, On the measurement of productive efficiency, *Swedish Journal of Economics* 76, 141-54.

Førsund, F.R. and L. Hjalmarsson, 1987, *Analyses of industrial structure: A putty-clay approach* (Almqvist & Wiksell International, Stockholm).

Greene, W.H, 1993, *Econometric Analysis*, 2nd edition (Prentice-Hall, London).

Hjalmarsson, L. and A.Veiderpass, 1992, Efficiency and ownership in Swedish electricity retail distribution, *Journal of Productivity Analysis* 3, 7-23.

Johnson, N.L. and S.Kotz, 1970, *Distributions in statistics: Continous univariate distributions-2* (Houghton Mifflin, Boston).

Kittelsen, S.A.C., 1994, Effektivitet og regulering i norsk elektrisitetsdistribusjon, Report No. 3/94, SNF - Foundation for research in economics and business administration, Oslo.

Kittelsen, S.A.C., 1997, Monte Carlo simulations of DEA efficiency scores and hypothesis tests, Working paper, SNF - Foundation for research in economics and business administration, Oslo.

Kittelsen, S.A.C. and F.R.Førsund, 1992, Efficiency analysis of Norwegian district courts, *Journal of Productivity Analysis*, 3, 277-306.

Kneip, A., B.U.Park and L.Simar, 1996, A note on the convergence of nonparametric DEA efficiency measures, Discussion Paper 9603, Institut de Statistique, Université Catholique de Louvain.

Kneip, A. and L.Simar, 1995, A general framework for frontier estimation with panel data, Discussion Paper 9560, Center for Operations Research and Econometrics, Université Catholique de Louvain.

Korostelev, A.P., L.Simar and A.B.Tsybakov, 1995a, Efficient estimation of monotone boundaries, *Annals of Statistics*, 23 (2), 476-489.

Korostelev, A.P., L.Simar and A.B.Tsybakov, 1995b, On estimation of monotone and convex boundaries, *Publications de l'Institut de statistique de l'Université de Paris*, 39, 3-18.

Koutsoyiannis, A., 1977, *Theory of Econometrics* (Macmillan, London).

Leamer, E.E., 1983, Model choice and specification analysis, in Z.Grilliches and M.D.Intriligator, eds., *Handbook in Econometrics*, Vol 1 (North-Holland, Amsterdam)

Lehman, E.L., 1975, *Nonparametrics: Statistical methods based on ranks* (McGraw-Hill, New York).

Lewin, A.Y., R.C.Morey and T.J.Cook, 1982, Evaluating the administrative efficiency of courts, *International Journal of Management Science* 10, 401-411.

Magnussen, J., 1996, Efficiency measurement and the operationalization of hospital production, *Health Services Research* 31, 21-37.

Norman, M. and B.Stoker, 1991, *Data Envelopment Analysis. The assessment of performance* (John Wiley & Sons, Chichester).

Olesen, O.B., and N.C.Petersen, 1991, Collinearity in Data Envelopment Analysis: An extended facet approach, Publications from Department of Management, Odense University, No. 1/1991.

Olesen, O.B., and N.C.Petersen, 1996, Indicators of ill-conditioned data sets and model misspecification in Data Envelopment Analysis: An extended facet approach, *Management Science* 42, 205-219.

Petersen, N.C. and O. Olesen, 1995, Chance constrained efficiency evaluation, *Management Science* 41, 442-457.

Salvanes, K.G. and S.Tjøtta, 1994, Productivity differences in multiple output industries: An empirical application to electricity distribution, *Journal of Productivity Analysis*, 5, 23-43.

Seiford, L.M., 1996, Data Envelopment Analysis: The evolution of the state of the art (1978-1995), *Journal of Productivity Analysis* 7 (2/3), 99-137.

Shephard, R.W., 1970, *Theory of cost and production functions*, 2nd ed. (Princeton University Press, Princeton).

Simar, L. and Wilson, P.W., 1995, Sensitivity analysis of efficiency scores: How to bootstrap

in nonparametric frontier models, Management Science, forthcoming. Available as Discussion Paper 9503, Institut de Statistique, Université Catholique de Louvain.

Statistisk Sentralbyrå , 1991, *Elektrisitetsstatistikk 1989*, Noregs offisielle statistikk B969 (Statistisk Sentralbyrå, Oslo).

Thomson, R.G., P.S.Dharmapala and R.M.Thrall, 1994, Sensitivity analysis of efficiency measures with and application to Kansas farming and Illinois coal mining, in A.Charnes, W.W.Cooper,A.Y.Lewin and L.M.Seiford (Ed.), *Data Envelopment Analysis: Theory, methodology and applications* (Kluwer, Boston).

Thrall, R.M., 1989, Classification transitions under expansion of inputs and outputs in Data Envelopment Analysis, *Managerial and Decision Economics* 10, 159-162.

Torgersen, A.M., F.R.Førsund and S.A.C.Kittelsen, 1996, Slack adjusted efficiency measures and ranking of efficient units, *Journal of Productivity Analysis* 7, 379-398.

Valdmanis, V., 1992, Sensitivity analysis for DEA models. An empirical example using public vs. NFP hospitals. *Journal of Public Economics* 48, 185-205.

Wilhelmsen, L. og G. Lilleschulstad, 1990, Personminutter - Nytt mål på befolkningens geografiske fordeling (Kommunaldepartementet, Oslo).