

This item was submitted to Loughborough's Research Repository by the author. Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Stochastic non-parametric efficiency measurement and yardstick competition in electricity regulation

PLEASE CITE THE PUBLISHED VERSION

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Weyman-Jones, Thomas G. 2019. "Stochastic Non-parametric Efficiency Measurement and Yardstick Competition in Electricity Regulation". figshare. https://hdl.handle.net/2134/418.



This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.

COMMONS DEED
Attribution-NonCommercial-NoDerivs 2.5
You are free:
 to copy, distribute, display, and perform the work
Under the following conditions:
Attribution . You must attribute the work in the manner specified by the author or licensor.
Noncommercial. You may not use this work for commercial purposes.
No Derivative Works. You may not alter, transform, or build upon this work.
 For any reuse or distribution, you must make clear to others the license terms of this work
 Any of these conditions can be waived if you get permission from the copyright holder.
Your fair use and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full license).
Disclaimer 🖵

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/

Stochastic Non-parametric Efficiency Measurement and Yardstick Competition in Electricity Regulation

Thomas G. Weyman-Jones

Department of Economics, Loughborough University, Leicestershire, LE11 3TU, UK t.g.weyman-jones@lboro.ac.uk

Abstract

Stochastic non-parametric efficiency measurement constructs production or cost frontiers that incorporate both inefficiency and stochastic error. This results in a closer envelopment of the mean performance of the companies in the sample and diminishes the effect of extreme outliers. This paper uses the Land, Lovell and Thore (1993) model incorporating information on the covariance structure of inputs and outputs to study efficiency across a panel of 14 electricity distribution companies in the UK during the 1990s. The purpose is to revisit the 1999 distribution price control review carried out by the UK regulator. The regulator's benchmarking is contrasted with the stochastic non-parametric efficiency results and with other comparative efficiency models offering close envelopment of the data. Some conclusions are offered about the possible regulated price effects in the UK case¹.

Keywords: yardstick competition, price cap regulation, stochastic DEA, electricity distribution

JEL numbers: C14, L33, L51

¹ Comments from seminar participants at Aston, Warwick and Cambridge are acknowledged. Usual disclaimer applies. This paper reports work in progress.

Introduction.

Yardstick competition, or more loosely benchmarking, is a principal component of price cap regulation for groups of network monopolies. Considerable theoretical and applied research by economists has been directed to this, and several regulatory benchmarking exercises have been carried out, for example in the UK, Netherlands, Norway and Australia. There is a developing interest in the subject by many European electricity and gas regulators with the possibility of cross border comparative efficiency studies. The single electricity and gas market directives of the European Union have also discussed the possibility of forms of yardstick competition.

Three major issues arise in applying this principle:

- Data comparability and coverage
- Translations of the yardstick competition results into price caps
- The nature of the models used to implement comparative efficiency studies.

These issues have been central to the exchanges between regulators and regulated companies in different countries. While all three issues are discussed in the paper, the major emphasis is on the last of these concerning the different models of efficiency and productivity analysis used by regulators. The major candidate models in use are non-parametric efficiency measurement or data envelopment analysis, DEA, stochastic frontier analysis, SFA, and other regression based but deterministic models such as corrected ordinary least squares, COLS. The trade off between DEA and SFA is well known. DEA models, which include the non-convex free disposal hull FDH, do not impose the assumptions: (a) that all companies share the same production function and (b) that the distribution function of inefficiencies is known. SFA models, which do

impose these assumptions, do not assume that all of the variation amongst companies is due to inefficiency and are able to accommodate stochastic errors. Nevertheless it is true that SFA is rarely used by regulators, perhaps due to lack of data, and several regulatory judgements have made use of DEA (Netherlands, Norway, Australia) or some version of COLS (UK, but an idiosyncratic version).

The choice of model clearly affects the way in which companies participate in the regulatory process. Participation or individual rationality constraints are important to the game theoretic justification of regulation by yardstick competition and consequently the choice of model must affect the outcome of the regulatory game. One possible resolution of the DEA-SFA trade off lies in the use of stochastic non-parametric efficiency measurement, i.e. stochastic data envelopment analysis, SDEA. This model incorporates a fuzzy frontier which allows for both inefficiency and stochastic error to determine the relative positions of the companies. Only a limited amount of research has been done in applying SDEA in a regulatory framework or in contrasting its results with other models, and consequently this paper is written to compare the use of DEA and SDEA in the context of a particular regulatory experience. The regulation in question is the 1999 UK electricity distribution review carried out by the Office of Gas and Electricity Markets, OFGEM (previously OFFER). That review was based partly on a form of COLS and led to some controversial price cap changes.

The paper will revisit the 1999 review of electricity company operating expenditures, OPEX, by applying DEA and SDEA models in order to gauge the effect of allowing for stochastic error in regulatory benchmarking. In particular it will be interesting to compare the OFGEM benchmarks with the DEA and SDEA results to

determine whether company rankings are greatly affected by the choice of model. These comparisons can also be applied to the price caps actually implemented since it may be the case that a particular price cap is not sensitive to the choice of OPEX benchmarking model. The advantage of revisiting the 1999 UK review is that well documented datasets and outcome sets have been published. The object is not of suggest new price caps but to investigate the sensitivity of price cap outcomes to the choice of model particularly where the model incorporates stochastic errors.

The structure of the paper is as follows. Section 2 sets the context for yardstick competition using an argument due to Bogetoft (1997) about the choice of model in a game theoretic framework. Section 3 develops the idea of the closest envelopment of the data implicit in the choice of model. Section 4 describes the SDEA model suggested by Land, Lovell and Thore (1993) and explains how it is implemented. Section 5 describes the data while Section 6 presents the results of the exercise. Section 7 offers conclusions about the regulatory implications of the analysis.

Incentives and Yardsticks

Although Shleifer (1985) is the classic theoretical basis for yardstick competition there have been further developments and among the most notable of these has been a series of papers by Peter Bogetoft and others, see especially Bogetoft (1997) and Agrell, Bogetoft and Tind (2000). In these papers the authors explore the relationship between yardstick competition and the use of data envelopment analysis, DEA which has been one of the most widely used methods of comparative efficiency measurement amongst regulated utilities. This section presents a very simplified summary of the Bogetoft

models rather than a detailed analysis. The aim is to do no more than capture Bogetoft's insight that yardstick competition requires the regulator to think about the choice of benchmark model in terms of the way it will envelop the data². The timeline for Bogetoft's regulatory principal agent game, illustrated in figure 1, is essentially similar to that of Shleifer but a different range of possibilities is explored. Among *n* different utilities the typical firm is observed to have input expenditures of *wx* where *x* is a vector of inputs and *w* is a vector of input prices, a vector of outputs *y* and possibly a vector of non-controllable inputs *z*. This data is verifiable in the sense that the regulator can measure and check the data on outputs, input expenditures and non-controllable inputs. The regulator contracts with the firms at the beginning of the game to pay a revenue cap *b* according to a formula that depends on the observed costs and outputs of the firms.

FIGURE 1 HERE

The utility (but not the regulator) knows the minimal cost of using current technological possibilities to produce the outputs given the inputs, input prices and non-controllable inputs:

 $C(y|z,w) = \min_{x} \{wx : x \text{ and } z \text{ can make } y\}$

The firm (or the managers) can choose a degree of slack, *s*, which is also unknown to the regulator, so that the actual cost experienced by the firm is:

C(y|z,w)+s

The regulated firm's ex-ante utility is assumed to depend on the difference between (i) its allowed revenue cap b and its actual verified input expenditure wx, plus

 $^{^{2}}$ The description here is an abuse of notation in combining the ideas from two different papers, Bogetoft (1997) and Agrell Bogetoft and Tind (2000)

(ii) a fraction r of the difference between the expenditure on inputs and the cost (including slack) of producing its output target:

$$U = b - wx + \mathbf{r}(wx - C(y|w, z) - s).$$

where the strict inequalities 0 < r < 1 are satisfied. The slack is consumed by the firm in converting inputs into outputs using the available technological possibilities. The restrictions on the marginal utility of the slack, *r*, ensure that at the margin the firm prefers to increase profit rather than to consume slack although both yield positive utility. The regulator is unaware of the minimal cost function, but knows or estimates the firm's marginal utility of slack, and endeavours to minimise the informational rent paid to each regulated firm through the revenue cap.

Bogetoft et al show that generally an optimal (individually rational and incentive compatible) revenue cap contract which will minimise the amount of informational rent to be paid to the firms takes the following form for each firm:

$$b = wx + \mathbf{r}[c * -wx]$$

i.e.
$$b = \mathbf{r}c * + (1 - \mathbf{r})wx$$

where c^* is a "best practice cost norm" or "minimal extrapolation cost standard" set to act as a benchmark for the firm in question. In other words the firm is paid its observed input cost plus a proportion of the difference (positive or negative) between a benchmark of the cost of meeting the firm's observed output level and its observed input cost.. This best practice cost norm is "the maximal cost of producing the firm's outputs that is consistent with the a priori assumptions about possible cost structures and the realised production plans [costs and outputs] of the other firms" (Bogetoft (1997) p.285). The term *maximal cost* is used because the role of the benchmark is to provide an upper bound on the informational rent paid to the firm so it is essential that it at least exceeds the minimal technological cost of production:

 $c^* \ge C(y|w,z)$

To ensure that the informational rent required to encourage participation and correct revelation of costs is minimised the benchmark should reflect the cost that would be observed if some other frontier efficient firm was to supply the output of the firm in question. Clearly too low a benchmark c^* will discourage the firm from participating in the regulatory game. On the other hand too high a benchmark will lead to inefficiently high payments to the firm. Consequently it is required that c^* is the *least upper bound* of the possible values of the cost of production. Without knowing the minimal technological cost of production the regulator has to find the least upper bound of the set which contains this unknown function. The observed input expenditures, outputs and non-controllable inputs of the firms that are subject to the yardstick competition can provide information about this least upper bound. In particular, under assumptions of disposability and convexity of the production possibility set, the DEA efficient cost under constant returns to scale, $C^{DEA-CRS}$ could be a candidate for c^* :

 $C^{DEA-CRS} = \min wx$ s.t. $x \ge \sum_{j=1}^{j=n} \mathbf{I}_j x_j$ $y \le \sum_{j=1}^{j=n} \mathbf{I}_j y_j$ $\mathbf{I}_j \ge 0, j = 1 K n$

However, for incentive reasons the benchmark should exclude the cost and output of the firm in question from the reference set for which the frontier is calculated. In this respect his suggestion replicates the DEA model of Andersen and Petersen (1993) which was

initially suggested as a way to rank firms all of which are efficient according to the standard DEA model. In general however, these arguments provide both a model of yardstick competition and an analytical justification for using the DEA frontier efficiency measure.

Bogetoft argues that, in the presence of significant uncertainty about the technology on the part of the regulator, the DEA based cost norm has several advantages:

- 1. it requires very little a priori technological information
- 2. it allows flexible non-parametric modelling of multiple output and multiple input production processes
- 3. it is essentially conservative is determining the informational rents.

Closest Envelopment Models and Nested Costs

The essence of Bogetoft's argument is that to encourage participation in the regulatory game the agent must be offered a contract based on an upper bound of the possible costs he or she could incur, but to minimise information rents the principal will seek a least upper bound to the set of possible costs. Within the range of comparative efficiency models it is possible to discover some that are nested and some that are not. This allows the researcher to examine different least upper bounds for the cost set under different assumptions. Begin with the basic distinction between parametric and non-parametric models. Parametric, i.e. regression based models assume that each company uses the same underlying technology represented by the production function³. Within this group stochastic frontier models allow some of the variation in cost to be random while

corrected ordinary least squares models attribute all of the variation to inefficiency. Consequently, in this case the relationship between the minimal cost extrapolation or least upper bound of the of the relative frontier cost levels is:

 $C^{COLS} \leq C^{SFA}$

This means that the efficient benchmark for cost performance in the corrected ordinary least squares model is at a lower total cost level than the efficient benchmark for the stochastic frontier cost model.

However since non-parametric models allow each company to have a different technology it is clear that parametric and non-parametric models are not nested even when one is deterministic and the other is stochastic:

C^{DEA} , C^{SFA} cannot be unambiguously ranked before calculation.

Within DEA based models variable returns to scale possibilities allow some of the variation in cost to be attributable to inefficient scale rather than pure technical inefficiency so that nesting is possible on the basis of scale assumptions. This means that the efficient frontier cost benchmark is at a lower overall level for constant returns to scale models than for variable returns to scale models:

$C^{DEA-CRS} \leq C^{DEA-VRS}$

It is however possible to obtain a larger least upper bound to the frontier costs by using as an alternative to the conventional DEA methodology the free disposal hull (FDH) method - see Desprins, Simar and Tulkens (1984), and Tulkens (1993). This method imposes a further constraint on each of the elements of the intensity vector, namely that I_j is either zero or unity: $I_j \in \{0,1\}$ $j = 1 \text{K} \ n$. This turns the conventional

³ The assumption may be relaxed in random parameters models but the nesting effect is not clear.

LP problem of DEA into a mixed integer programming problem (MIP). Actual calculation of the FDH frontier can proceed by MIP or more easily by Tulkens' (1993) complete enumeration algorithm based on vector dominance. The economic impact of the change is quite profound, and can be best illustrated by a diagram.

FIGURE 2 HERE

In figure 2 we see observations represented by the points a - f. The DEA isoquant is the piecewise linear frontier connecting a, b, d, and f, with c and e as inefficient points. Reducing inputs radially at points c and e brings those input-output observations back onto the frontier on the segments connecting a to d and d to f respectively. However this means points c and e are being compared with hypothetical but potentially efficient combinations of the actual observations at b, d and f. Supporters of the FDH methods do not recognise the validity of this comparison with hypothetical input-output combinations, and seek to compare efficiency scores of an observed firm only with other observed firms. The FDH isoquant is the stepped line connecting a, b, c, d, e, and f. Each of these points is then regarded as efficient. Only observed point g in the diagram now counts as actually inefficient since unlike any of the others it is dominated by the actual observations c and d. Desprins, Simar and Tulkens (1984, p.264) put the case for this approach on two grounds: it rests on the weakest assumptions regarding the production set and identification of dominating observations reveals an information set of direct use for managers

The assumptions required are simply input and output disposability (i.e. the firm can reduce slack or unused inputs or use them to expand outputs without using up other additional resources). No assumption is made regarding the nature of the returns to scale. Nevertheless the method is controversial in the context of identifying potential efficiency gains. Two contrasting views illustrate the issue. The first can be stated thus: we should seek the frontier which shows the firm in the best light: "evaluation of a given unit under the most favourable conditions has been claimed as one of the advantages of DEA" -Petersen (1990). This suggests the tightest envelopment surface is to be preferred, and this is the FDH frontier. The second argument is that we should seek potential efficiency gains for firms: "at the heart of the method [DEA] is the assumption that we may interpolate between any number of units within the comparator set to construct efficient units which could have existed in principle even if not observed in practice. This way we create an efficient boundary of units, some observed some not, and then the distance of a unit from the boundary gives us a measure of its efficiency" - Thanassoulis (1999). This suggests that convexity is an important comparator property for identifying potential efficiency gains, hence the DEA frontier is preferred from this point of view.

It is possible to read into actual regulatory decisions the views of the players in the process. For example the Netherlands regulator specifically resisted the use of the FDH model using arguments based on the importance of allowing for potential competition, DTE (2000). The key argument was that ignoring potential comparators would cause the regulator to lose credibility with consumer groups. In contrast, US regulatory commissions have indicated a reluctance to compare a regulated company with anything but other actual companies.

By ignoring potential combinations of efficient companies the FDH frontier implies market entry or performance is not feasible at that combination. However, in the context of a group of regulated monopolies, customer groups are unlikely to believe this, and rival companies certainly will not.

We can rank the DEA technical efficiency scores (q) that arise from relaxing assumptions as follows:

 $0 \leq \boldsymbol{q}^{DEA-CRS} \leq \boldsymbol{q}^{DEA-VRS} \leq \boldsymbol{q}^{FDH} \leq 1$

and consequently obtain the nested frontier costs:

$$0 \le C^{DEA-CRS} \le C^{DEA-VRS} \le C^{FDH}$$

so that FDH will always show companies in at least as good a light and usually better than DEA under variable returns to scale, and this in turn is at least as good as and usually better than DEA under constant returns to scale.

When stochastic models are considered the SDEA approach will always permit a closer envelopment of the data than the deterministic model so that:

$$C^{DEA-CRS} \leq C^{SDEA-CRS}$$

However it is not possible to argue which relaxation will have the greater effect compared to deterministic constant returns to scale DEA: stochastic DEA or non-convex DEA so that C^{FDH} , $C^{SDEA-CRS}$ for example are not nested costs. As well as the choice of technique, other factors such as sample size and model specification also affect the minimal extrapolation costs. These are summarised in table 1.

Efficiency Study context	Companies loosely enveloped = low mean efficiency	Companies closely enveloped = high mean efficiency	Regulatory credibility	Comment
Scale	CRS	$VRS \rightarrow FDH$	VRS yes, FDH no?	Potential comparators
Sample Size	Large	Small	Want large comparator set	Data comparable?
Number of variables	Low	High	Put a limit on the dimensions of comparison	Companies say their operations are all different
Stochastic	No	Yes	Data availability limited	Some limits needed on efficiency differences

 Table 1 factors affecting the data envelopment

Stochastic Data Envelopment Analysis

The procedure for DEA measurement of input based technical efficiency is well known. We take each firm in turn and compare it with the reference set of the whole industry. This is represented by the input requirements set for a given level of outputs, which is bounded below by the isoquant. The object here is to find the largest reduction in the firm's actual input usage which will allow it to remain in the input requirements set, i.e. achieve a position on the efficient frontier isoquant determined by the observations on the industry as a whole.

Doing this for each firm in turn we identify the firm's q value. This is the firm's Farrell efficiency: $0 \le q \le 1$. Values of q = 1 indicate that the firm is already one of those

which defines the frontier and is 100 per cent efficient. The firm's inefficiency is $(1-q) \times 100\%$. In what follows it is necessary to examine particular output and input constraints which can be written in terms of *s* outputs: y_{ij} , r = 1K *s*, j = 1K *n* and *m* inputs: x_{ij} , i = 1K *m*, j = 1K *n* for the *n* different producing units. The input requirement set is defined by the following inequalities for each producing unit in turn. The producing unit under observation is subscripted '0' to distinguish it from all of the producing units together: j = 1K *n*

rth typical output constraint:

$$\mathbf{y}_{r}'\mathbf{l} - y_{r0} \ge 0$$
 i.e. $\sum_{j=1}^{j=n} y_{rj}\mathbf{l}_{j} - y_{r0} \ge 0$ $r = 1, K$, s

ith typical input constraint:

$$\mathbf{x}_{i}'\mathbf{l} - x_{i0}\mathbf{q} \le 0$$
 i.e. $\sum_{j=1}^{j=n} x_{ij}\mathbf{l}_{j} - x_{i0}\mathbf{q} \le 0$ $i = 1, K$, m

We measure the producing unit's technical efficiency by calculating the following linear programme for the firm in question (now subscripted 0):

$$\min \mathbf{q} \quad s.t$$

$$\mathbf{y}_{r} \mathbf{l} - \mathbf{y}_{r0} \ge 0$$

$$\mathbf{x}_{i} \mathbf{l} - \mathbf{x}_{i0} \mathbf{q} \le 0$$

Now consider the chance constrained DEA problem described by Land, Lovell, and Thore (1993). This allows the constraints to hold with probability level $a \in (0,1)$ i.e. with less than certainty:

$$\min \boldsymbol{q}$$

s.t.
$$\Pr\left(\mathbf{y}_{r}'\boldsymbol{l} - y_{ro} \ge 0\right) \ge \boldsymbol{a} \qquad r = 1K \ s$$

$$\Pr\left(\mathbf{x}_{i}'\boldsymbol{l} - x_{io}\boldsymbol{q} \le 0\right) \ge \boldsymbol{a} \qquad i = 1K \ m$$

The argument for this formulation is as follows. The deterministic DEA problem allows firms to lie on or inside the production frontier. The constraints in the deterministic problem can be thought of as holding with probability one. Land, Lovell and Thore (LLT) allow a small number of firms to be super-efficient, i.e. to lie beyond the production frontier. For these firms the output and input constraints in the envelopment DEA model will be violated. In the general statement of the problem the constraints will hold with probability less than one. The implication of this is that the frontier is not defined by these outlier firms but lies closer to the observations of the mass of firms in the sample. In a sense the frontier is defined in a more fuzzy manner. After developing the statement of the model, an alternative interpretation of chance constrained DEA due to Olesen and Petersen (1995) is presented which reinforces this idea of stochastic variation in the constraints.

The basic probability statement to be used is the conventional result for the normal distribution:

$$\Pr(z \le z_0) = \int_{-\infty}^{z_0} \boldsymbol{f}(z) dz = \Phi(z_0)$$

where z is the standard normal deviate with probability density function:

$$\boldsymbol{f}(z) = \frac{1}{\sqrt{2\boldsymbol{p}}} e^{-\frac{1}{2}z^2} \, .$$

and $\Phi(z)$ is the cumulative distribution function of the normally distributed variable. Since the cumulative distribution function is non-decreasing over the support $[-\infty,\infty]$, it has the inverse function:

 $z = \Phi^{-1}(a)$ for given a.

Though not amenable to analytical evaluation, the inverse is well known from the tables of the standard normal distribution, e.g.:

$$\Phi^{-1}(0.95) = 1.645$$
 and $\Phi^{-1}(0.975) = 1.96$

The symmetry of the distribution around zero provides the two additional properties (*a*) and (*b*) below which are used in the construction of the tables for *z*. Symmetry implies: f(z) = f(-z) and this together with integration by change of variable easily establishes property (*a*) and it is this property which is used in the Charnes Cooper derivation.

(a) $\Phi(-z_0) = 1 - \Phi(z_0)$ i.e.: $\Pr(z \le -z_0) = 1 - \Pr(z \le z_0) = \Pr(z \ge z_0)$ and: $\Pr(z \ge -z_0) = \Pr(z \le z_0) = 1 - \Pr(z \le -z_0)$

(b)
$$\Phi^{-1}(1-a) = -\Phi^{-1}(a)$$
 e.g. $\Phi^{-1}(0.05) = -1.645 = -\Phi^{-1}(0.95)$

We now return to the chance constrained DEA problem:

$$\min \boldsymbol{q}$$

s.t.
$$\Pr\left(\mathbf{y}_{r}'\boldsymbol{l} - y_{ro} \ge 0\right) \ge \boldsymbol{a} \qquad r = 1K \quad s$$

$$\Pr\left(\mathbf{x}_{i}'\boldsymbol{l} - x_{io}\boldsymbol{q} \le 0\right) \ge \boldsymbol{a} \qquad i = 1K \quad m$$

Charnes and Cooper (1963) show how to use the idea of a modified certainty equivalent to transform this stochastic linear programming problem into a deterministic non-linear programming problem. As we noted earlier, the difference between the firm's output and the reference weighted outputs of all the firms is treated as a random variable. The difference between the firm's input adjusted for its efficiency and the reference weighted inputs of all the firms in the industry is also treated as a random variable.

We begin with the constraints relating to the outputs, and re-write them as below. In these steps we assume that the random variable has a finite positive variance so that the

standard deviation: $\left(\operatorname{var} \left(\mathbf{y}_{r}' \mathbf{l} - \mathbf{y}_{ro} \right) \right)^{\frac{1}{2}}$ can be used as a divisor.

If
$$\Pr\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro} \ge 0\right) \ge \mathbf{a}$$

then

$$\Pr\left(E\left(\mathbf{y}_{r}'\mathbf{l}-\mathbf{y}_{ro}\right)-E\left(\mathbf{y}_{r}'\mathbf{l}-\mathbf{y}_{ro}\right)+\left(\mathbf{y}_{r}'\mathbf{l}-\mathbf{y}_{ro}\right)\geq 0\right)\geq \mathbf{a}$$

and so re-arranging and dividing by the standard deviation

and so re - arranging and dividing by the standard deviation

$$\Pr\left[\frac{\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)-E\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)\right)^{\frac{1}{2}}} \ge -\frac{E\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)\right)^{\frac{1}{2}}}\right] \ge \mathbf{a}$$

Now assume the random variable representing the output shortfall is normally

distributed:

$$\begin{pmatrix} \mathbf{y}_{r}' \mathbf{l} - y_{ro} \end{pmatrix} \sim N \left[E \left(\mathbf{y}_{r}' \mathbf{l} - y_{ro} \right) \operatorname{var} \left(\mathbf{y}_{r}' \mathbf{l} - y_{ro} \right) \right]$$
then
$$z = \left[\left\{ \left(\mathbf{y}_{r}' \mathbf{l} - y_{ro} \right) - E \left(\mathbf{y}_{r}' \mathbf{l} - y_{ro} \right) \right\} / \left(\operatorname{var} \left(\mathbf{y}_{r}' \mathbf{l} - y_{ro} \right) \right)^{\frac{1}{2}} \right] \sim N(0,1)$$

We use z to replace the first expression in the preceding probability statement to obtain:

$$\Pr\left[z \ge -\frac{E\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro}\right)\right)^{\frac{1}{2}}}\right] \ge \mathbf{a}$$

i.e.
$$\Pr\left[z \le \frac{E\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro}\right)\right)^{\frac{1}{2}}}\right] \ge \mathbf{a}$$

This step uses the symmetry property (a) described earlier. Now we use the definition of the probability statement in terms of the distribution function to write:

$$\Phi\left(\frac{E\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)\right)^{\frac{1}{2}}}\right) \geq \mathbf{a}$$
and so
$$\frac{E\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)}{\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l}-y_{ro}\right)\right)^{\frac{1}{2}}} \geq \Phi^{-1}(\mathbf{a})$$

giving:

$$E\left(\mathbf{y}_{r}'\mathbf{l}-\mathbf{y}_{ro}\right)\geq\Phi^{-1}\left(\mathbf{a}\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l}-\mathbf{y}_{ro}\right)\right)^{\frac{1}{2}}\right)$$

This completes the transformation of the probabilistic version of the linear output constraint into a deterministic non-linear form using what Charnes and Cooper (1963) refer to as a modified certainty equivalent. It is useful to write it in a slightly more general form as follows.

$$\mathbf{y}_{r}'\mathbf{l} + (\mathbf{E}\mathbf{y}_{r} - \mathbf{y}_{r})'\mathbf{l} - \Phi^{-1}(\mathbf{a})\left(\operatorname{var}\left(\mathbf{y}_{r}'\mathbf{l} - y_{ro}\right)\right)^{\frac{1}{2}} \ge Ey_{ro}$$

Turning now to the input constraints, these are initially expressed as:

$$\Pr\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\leq0\right)\geq\mathbf{a}$$

and using the identical algebraic steps as in the output case, together with the normality assumption:

$$\begin{pmatrix} \mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \end{pmatrix} \sim N \left[E \left(\mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \right) \operatorname{var} \left(\mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \right) \right]$$
so that
$$z = \left[\left\{ \left(\mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \right) - E \left(\mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \right) \right\} \right] / \left(\operatorname{var} \left(\mathbf{x}_{i}' \mathbf{l} - x_{io} \mathbf{q} \right) \right)^{\frac{1}{2}} \right] \sim N(0,1)$$

we obtain the transformed probability statement:

$$\Pr\left[z \leq -\frac{E\left(\mathbf{x}_{i}'\mathbf{l} - x_{io}\mathbf{q}\right)}{\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l} - x_{io}\mathbf{q}\right)\right)^{\frac{1}{2}}}\right] \geq \mathbf{a}$$

Proceeding as before we therefore write:

$$\Phi\left(-\frac{E\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)}{\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)\right)^{\frac{1}{2}}}\right) \geq \mathbf{a}$$
so
$$-\frac{E\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)}{\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)\right)^{\frac{1}{2}}} \geq \Phi^{-1}(\mathbf{a})$$
i.e.
$$\frac{E\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)}{\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)\right)^{\frac{1}{2}}} \leq -\Phi^{-1}(\mathbf{a})$$
and
$$E\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right) \leq -\Phi^{-1}(\mathbf{a})\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l}-x_{io}\mathbf{q}\right)\right)^{\frac{1}{2}}$$

This completes the transformation as in the output case, but again we can write the transformed non-linear constraint in a slightly more general form:

$$\mathbf{x}_{i}'\mathbf{l} + (\mathbf{E}\mathbf{x}_{i} - \mathbf{x}_{i})'\mathbf{l} + \Phi^{-1}(\mathbf{a})\left(\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l} - x_{i0}\mathbf{q}\right)\right)^{\frac{1}{2}} - Ex_{i0}\mathbf{q} \leq 0$$

With these results we can write the stochastic DEA model in the LLT formulation as follows:

 $\min q$

s.t.

$$\mathbf{y}_{\mathbf{r}}'\mathbf{l} + (\mathbf{E}\mathbf{y}_{\mathbf{r}} - \mathbf{y}_{\mathbf{r}})'\mathbf{l} - \Phi^{-1}(\mathbf{a})\left[\operatorname{var}\left(\mathbf{y}_{\mathbf{r}}'\mathbf{l} - y_{r0}\right)\right]^{\frac{1}{2}} \ge Ey_{r0} \qquad r = 1 \text{K s}$$

$$\mathbf{x}_{\mathbf{i}}'\mathbf{l} + (\mathbf{E}\mathbf{x}_{\mathbf{i}} - \mathbf{x}_{\mathbf{i}})'\mathbf{l} + \Phi^{-1}(\mathbf{a})\left[\operatorname{var}\left(\mathbf{x}_{\mathbf{i}}'\mathbf{l} - x_{i0}\mathbf{q}\right)\right]^{\frac{1}{2}} - Ex_{i0}\mathbf{q} \le 0 \quad i = 1 \text{K m}$$

Here we are measuring the efficiency of the mean performance of the firm temporarily labelled with the 0 subscript. We may be more interested in measuring the efficiency of an actual realisation, in which case Ey_{r0} and Ex_{io} are replaced by y_{r0} and x_{io} .

The effect of the Charnes and Cooper algebra is to transform a stochastic linear programming problem into a deterministic non-linear programming problem. There are significant additional data requirements as a result and these concern the means and variances of the outputs and inputs and their covariances across different firms. For each output and each input the following data are required.

The expected value of usage of output and input for each firm:

$$Ey_{ij} \quad r = 1K \quad s; j = 1K \quad n$$
$$Ex_{ij} \quad i = 1K \quad m; j = 1K \quad n$$

and the variance-covariance matrix of usage of each output and each input across all firms, i.e. (s + m) variance covariance matrices:

$$\Psi_{r} = \begin{bmatrix} \operatorname{var}(y_{r1}) & \operatorname{cov}(y_{r1}y_{r2}) & \Lambda & \operatorname{cov}(y_{r1}y_{rn}) \\ \operatorname{cov}(y_{r2}y_{r1}) & \operatorname{var}(y_{r2}) & \Lambda & \operatorname{cov}(y_{r2}y_{rn}) \\ M & M & O & M \\ \operatorname{cov}(y_{rn}y_{r1}) & \Lambda & \Lambda & \operatorname{var}(y_{rn}) \end{bmatrix} \qquad r = 1 \mathrm{K} \ s$$

and

$$\Xi_{i} = \begin{bmatrix} \operatorname{var}(x_{i1}) & \operatorname{cov}(x_{i1}x_{i2}) & \Lambda & \operatorname{cov}(x_{i1}x_{in}) \\ \operatorname{cov}(x_{i2}x_{i1}) & \operatorname{var}(x_{i2}) & \Lambda & \operatorname{cov}(x_{i2}x_{in}) \\ M & M & O & M \\ \operatorname{cov}(x_{in}x_{i1}) & \Lambda & \Lambda & \operatorname{var}(x_{in}) \end{bmatrix} \qquad i = 1 \mathrm{K} \ m$$

The deterministic non-linear constraints which have replaced the probabilistic linear constraints of the original chance-constrained problem contain composite variance terms.

These differ according to whether the efficiency of the firm's mean performance or a realisation is being assessed. Each uses the relevant variance-covariance matrix of an output or input compared across all the firms, i.e.: Ψ_r or Ξ_i . When the mean performance is being assessed the scalar composite variance terms are computed as follows:

$$\operatorname{var}\left(\mathbf{y}_{\mathbf{r}}'\mathbf{l} - \mathbf{y}_{ro}\right) = \mathbf{l}'\Psi_{r}\mathbf{l} + \operatorname{var}(\mathbf{y}_{r0}) - 2\mathbf{\hat{l}}'\Psi_{r}\mathbf{i}^{0}$$
$$\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l} - \mathbf{x}_{io}\mathbf{q}\right) = \mathbf{l}'\Xi_{i}\mathbf{l} + \mathbf{q}^{2}\operatorname{var}(\mathbf{x}_{i0}) - 2\mathbf{q}\mathbf{\hat{l}}'\Xi_{i}\mathbf{i}^{0}$$

In these expressions, \hat{I} is the vector which has $I_j - \frac{1}{2}$ in the position corresponding to the firm whose efficiency is being measured, and I_k , k = 1K $n, k \neq j$ elsewhere, and i^0 is the column of the (n by n) identity matrix which has the value 1 in the position corresponding to the firm whose efficiency is being measured. When a realisation is being assessed the composite variance terms are computed as:

$$\operatorname{var}\left(\mathbf{y}_{\mathbf{r}}'\mathbf{l} - \mathbf{y}_{ro}\right) = \mathbf{l}'\Psi_{r}\mathbf{l}$$
$$\operatorname{var}\left(\mathbf{x}_{i}'\mathbf{l} - \mathbf{x}_{io}\right) = \mathbf{l}'\Xi_{i}\mathbf{l}$$

For clarity of setting up the model, LLT (1993) suggest that the problem can be restated in non-matrix terms. Using Z_{1-a} to denote the critical value of *z* from the standard normal tables, we have for the mean performance case:

min**q**

st.

$$\sum_{j=1}^{j=n} y_{rj} \mathbf{l}_{j} + \sum_{j=1}^{j=n} (Ey_{rj} - y_{rj}) \mathbf{l}_{j} - Z_{1-a} \left[\sum_{j=1}^{j=nk=n} \mathbf{m}_{j} \mathbf{m}_{k} (\operatorname{cov}(y_{rk} y_{rj})) \right]^{\frac{1}{2}} \ge Ey_{r0}$$

$$r = 1 \text{K s}$$

$$\sum_{j=1}^{j=n} x_{ij} \mathbf{l}_{j} + \sum_{j=1}^{j=n} (Ex_{ij} - x_{ij}) \mathbf{l}_{j} + Z_{1-a} \left[\sum_{j=1}^{j=nk=n} \sum_{k=1}^{n} p_{k} (\operatorname{cov}(x_{ik} x_{ij})) \right]^{\frac{1}{2}} - Ex_{i0} \mathbf{q} \le 0$$

$$i = 1 \text{K } m$$

In this restatement:

$$\boldsymbol{m}_{j} = \boldsymbol{l}_{j}$$
, for $j = 1$ K $n, j \neq 0$ and $\boldsymbol{m}_{j} = \boldsymbol{l}_{j} - 1$, for $j = 0$

and

$$\boldsymbol{n}_{j} = \boldsymbol{l}_{j}$$
, for $j = 1$ K $n, j \neq 0$ and $\boldsymbol{n}_{j} = \boldsymbol{l}_{j} - \boldsymbol{q}$, for $j = 0$

This is a non-linear programming problem in the variables: q, l_j, m and n_j .

Specifically it has a linear objective and (s + m) quadratic inequality constraints with additional restrictions on the variables to ensure positive variance terms. For the single realisation case the implementation is shown below, and this version is stated in Lovell (1993, pp 34-5):

min**q**

s.t.

$$\sum_{j=1}^{j=n} y_{rj} \mathbf{l}_{j} + \sum_{j=1}^{j=n} (Ey_{rj} - y_{rj}) \mathbf{l}_{j} - Z_{1-a} \left[\sum_{j=1}^{j=nk=n} \mathbf{l}_{j} \mathbf{l}_{k} (\operatorname{cov}(y_{rk}y_{rj})) \right]^{\frac{1}{2}} \ge y_{r0}$$

$$r = 1 \text{K } s$$

$$\sum_{j=1}^{j=n} x_{ij} \mathbf{l}_{j} + \sum_{j=1}^{j=n} (Ex_{ij} - x_{ij}) \mathbf{l}_{j} + Z_{1-a} \left[\sum_{j=1}^{j=nk=n} \sum_{k=1}^{n} \mathbf{l}_{j} \mathbf{l}_{k} (\operatorname{cov}(x_{ik}x_{ij})) \right]^{\frac{1}{2}} - x_{i0} \mathbf{q} \le 0$$

$$i = 1 \text{K } m$$

This version is slightly simpler because the choice variables in the variance terms are independent of the efficiency score q. To implement this programme we have used the algorithm of Lasdon et al (1978) which is widely available in many spreadsheet and symbolic programming applications, see Kendrick (1996).

What is the intuition behind SDEA? LLT provide one form of insight using the density function of the random error, but we can also borrow another diagrammatic intuition from the paper by Olsen and Petersen (1995). This is shown in figure 3 below.

In this diagram we illustrate observations on a panel of producing or decision making units (DMUs 1 -3) for the case of two inputs and one output. The boundary of the input requirements set is defined by the isoquant. In deterministic DEA the individually most efficient realisations define the frontier shown by the solid line. However, in implementing SDEA we are in effect looking for confidence regions around each producing unit's observations within the panel. These are shown as grouped within the ellipses shown around sets of observations. Olesen and Petersen describe the SDEA frontier as being evaluated relative to the centre of these confidence regions. As a consequence, the SDEA frontier associates extreme outliers with the stochastic error term and this has the effect of moving the frontier closer to the bulk of the producing units. Some realisations will then lie above the frontier and in evaluating the realisation model these observations will have a super-efficiency larger than unity.

FIGURE 3 HERE

In the diagram the DEA frontier passes through the most extreme observations of the three DMUs 1, 2, and 3, while the SDEA frontier passes through the centre of the confidence regions around the observations for these DMUs. We can see that particular observations will have a SDEA efficiency larger than unity. The observation at A has two efficiency scores: OB/OA for the DEA frontier and OB*/OA for the SDEA frontier. The SDEA score will usually be greater but never lower than the DEA efficiency score. The distance between the two frontiers represents the role of the stochastic error term in accounting for the variation in production performance. The larger is the variance of the sample, the larger will be the confidence ranges for the data and therefore the greater will be the distance between DEA and SDEA frontiers. In other words a sample with a wide variation in inputs and outputs observed for each unit will ascribe more of the variation in performance to the stochastic error than a sample with a narrow variation over the panel. In some cases we may find that a widely varying panel has two properties:

The *mean performance* of the units clusters around unity (100 percent efficiency) because the SDEA frontier has shifted so far towards the units which lie below the DEA frontier, and the *extreme performance* or individual realisation of some of the most successful observations lies well in excess of 100 percent. Such results would indicate that the sample contained a very large degree of measurement error and other stochastic influences, and consequently only the mean performance frontier is of relevance in using the results for such purposes as yardstick competition.

Data and case study

To demonstrate the effect of different forms of envelopment including SDEA consider the 1999 electricity distribution price control review (DPCR) carried out in the UK. 14 electricity distribution companies, known as the regional electricity companies or

RECs, featured in the review. An efficiency comparison was made covering their individual operating expenditure performance, OPEX, in 1998. The raw OPEX data was adjusted prior to the exercise in various ways to arrive at a variable called base cost, OFGEM (1999). This was compared with three output variables: customer numbers, units of electricity distributed, and length of lines. Prior to the regression the three outputs were combined into a single composite size variable using a first order binomial expansion with predetermined exponents. The plot of base cost against composite size for the 14 RECs was adjusted by altering its slope coefficient so that it passed through the lowest cost observation while maintaining the intercept value of base cost per unit of composite size at a level judged to be correct by a panel of expert advisers.

A number of problems exist with this procedure. There is some doubt about whether a sensible output vector has been chosen because of the use of *lines* (circuit length) as an output. The difficulty arises because *lines* expenditure is recognised as an important component of total company expenditure. *Lines* therefore has the status of a 'cost driver' in the management literature, but it is not strictly an output in the sense of providing a commodity for customers to purchase. Energy and customer numbers are feasible proxy variables for the service and commodity outputs which customers purchase. *Lines* can represent the difficulty of reaching customers in delivering products but is not an output of the production process in the sense normally used by economists.

Other problems with the OFGEM model concern the question of the appropriate adjustment to the raw OPEX data, the procedure for arriving at the composite output variable is arbitrary, and the fact that the frontier regression does not correspond to deterministic corrected ordinary least squares because the slope rather than the intercept

was adjusted to ensure non-negative errors. However since the primary interest in this paper is the effect of model choice rather than specification, it makes sense to adopt this OFGEM model as a benchmark. The objective is to investigate the effect that different envelopment models would have had if applied to this data set and specification for 1998.

The data therefore consist of the variables described in table 2. Since the base operating cost figure is derived by OFGEM only for the year 1997/8 this paper will only evaluate the single realisation corresponding to that year. The data for base cost was subjected by OFGEM to considerable refinement and will be treated as a set of non-stochastic constants rather than random variables. In other words the input constraint will be treated in this exercise as deterministic. The output constraints are treated as stochastic and the output slacks for each company are therefore normally distributed random variables with constant mean and variances estimated from the data for 1991/2 to 1997/8.

Variable	Category	Units	Source
Base Operating Cost	Input	£mn in 1997 prices	OFGEM 1999 (table 2.6,
		for 1997/8 only	p.19)
Customers	Output	Thousands for	Electricity Association,
		1991/92 to 1997/8	OFGEM and Company
			reports
Electricity delivered	Output	GigaWatthours for	Electricity Association,
		1991/92 to 1997/8	OFGEM and Company
			reports
Length of network	Output	Circuit kilometres	Electricity Association,
(overground +		for 1991/92 to	OFGEM and Company
underground)		1997/8	reports

 Table 2
 sample data

Company	Base Cost (£m, 97/8 only)	Customers	Units (GWh)	Lines (km)
Eastern	74	3151370	30477	87410
East Midlands	81	2244429	24472	66958
London	66	1951571	20246	29801
Manweb	58	1346143	18137	44227
Midlands	94	2204571	24271	63932
Northern	67	1432330	15148	40645
NORWEB	93	2158857	22346	58680
SEEBOARD	62	1999966	18065	44323
Southern	63	2573937	27180	70733
SWALEC	48	953731	11243	31944
South Western	64	1288200	13148	47546
Yorkshire	80	2031268	22725	53406
Scottish Power	71	1781540	21019	68150
Hydro-Electric	49	620574	7716	44883
mean of sample	69	1838463	19728	53760
standard deviatio	n of sample 14	663415	6311	16288

The mean values of the data in the period 1991/2 to 1997/8 are shown in table 3.

Table 3 data summary

As a model selection procedure the array of variables described above has several serious gaps. Customers and electricity delivered are frequently chosen as outputs but maximum demand may also be used to measure the power load which impacts on the network. It is difficult to defend the use of lines as an output other than as a proxy measure for the difficulty of providing distribution service to consumers. More conventionally lines and transformer capacity may be used as inputs representing the capacity of the network to deliver energy and customer services. Line losses are an important resource requirement in well dispersed systems but could be captured by the second power of lines. On the input side, lines and transformer capacity should be supplemented by a measure of short run variable inputs. In state controlled networks with no outsourcing numbers of full time equivalent employees may be used but in privatised

companies operating cost is a better measure because it can capture the cost of contracted input services. Finally, some measures of quality of supply may be included as additional output variables. Such measures could include the number of interruptions per customer and minutes of load lost. The Information and Incentives Project initiated in OFGEM (2000) is designed to resolve some of these issues.

Huang and Li (2001) in a review of different types of input-output disturbance formulation note that the SDEA model is very demanding in its parameter requirements. Although the model used here is less general than the core model in the Huang and Li paper it too requires many parameters. The dilemma can be put in context as follows. A deterministic DEA of *n* companies each making *s* outputs from *m* inputs requires knowledge of n(s+m) parameters, i.e. the outputs and inputs of each of the companies needed to make up the envelopment model's constraints. The whole sample of observations is used to generate these parameters so that a deterministic DEA has in practice zero degrees of freedom. The degrees of freedom is distinct from the number of columns (n+1) minus the number of rows (s+m) in the envelopment constraints matrix. The number of rows (s+m) determines the number of basic variables in a DEA solution, i.e. the number of efficient peer companies less one. In the stochastic model used here based on LLT there are $\{n(s+m) + (n^2(s+m)/2)\}$ parameters to be imposed in the constraint set because the model allows for non-zero but symmetric covariances amongst outputs and amongst inputs but not between outputs and inputs. Unless hypothetical or exogenously calibrated covariance terms are used, these have to be estimated from a panel data sample containing T periods. The number of observations is then nT(s+m) so that the degrees of freedom in the LLT model is: $nT(s+m) - \{n(s+m) + (n^2(s+m)/2)\}$.

This is likely to be much larger than the zero degrees of freedom in the deterministic model and may still be relatively large in econometric terms despite the large number of parameters used in the envelopment constraints. In the primary model to be examined in the next section the observation panel where T = 8, n = 14, (s + m) = 4 provides 448 observations from which $\{n(s + m) + (n^2(s)/2)\} = 350$ independent parameters are constructed leaving 98 degrees of freedom in the statistical sense.

In summary, this paper adopts as a framework for comparing different envelopment techinques the single, well documented model of OFGEM despite the apparent gaps in the range of variables used in that model.

Empirical Results and interpretations

The empirical results consist of the efficiency scores for the 14 companies for 1997-8 from five basic models: the price control DPCR, DEA-CRS, FDH, SDEA (a = 0.95) and SDEA (a = 0.8). These are shown in table 4.

company	DPCR 99	DEA-CRS	FDH	SDEA, 0.95	SDEA, 0.80
1	0.700	0.650	0.793	0.727	0.705
2	0.820	0.577	1.000	0.624	0.607
3	1.040	1.000	1.000	1.118	1.085
4	0.730	0.666	0.995	0.683	0.677
5	0.870	0.773	1.000	0.828	0.814
6	0.760	0.750	0.899	0.778	0.763
7	0.630	0.559	0.680	0.620	0.601
8	0.690	0.536	0.926	0.586	0.568
9	0.640	0.602	0.673	0.674	0.653
10	0.800	0.710	1.000	0.794	0.770
11	0.810	0.723	0.958	0.809	0.785
12	0.770	0.751	1.000	0.769	0.763
13	1.000	1.000	1.000	1.079	1.063
14	0.760	0.729	0.788	0.799	0.775

Table 4 Efficiency scores for UK RECs 1997-8

In table 4 and subsequent tables the numerical company ordering has been randomised. These empirical results confirm the role of the different models. DPCR is taken directly from OFGEM 1999 while the others are all estimated for this paper. As expected both FDH and SDEA models provide a closer envelopment of the data than the DEA-CRS model. SDEA at the 0.95 level is a closer envelopment than SDEA at the 0.8 level. Since this is a particular realisation some of the companies in the SDEA have efficiency scores in excess of unity reflecting the super-efficiency of those companies. The DPCR scores also had this property. At a basic empirical level the closer envelopment models do perform as predicted. However it is interesting to ask whether the differences compared with the DEA-CRS model and the OFGEM DPCR are statistically significant.

Statistical testing of empirical efficiency rankings is a rapidly developing area. For large sample studies Henderson and Russell (2001) suggested non-parametric tests using the integrated mean square error based on kernel densities to test the null hypothesis that the probability density functions underlying two samples of efficiency results are the same. The theory supporting these tests is described in Pagan and Ullah (1999, pp.60-69). For small samples Banker (1996) suggests a variety of tests for the same null hypothesis, including the non-parametric Kolmogorov-Smirnov two sample test which requires no strong assumptions about the probability distributions. This test is based on the maximum distance between the cumulative empirical distributions of the sample results. The procedure and critical values for the test are described in Siegel and Castellan (1988), and this is the test applied here.

The empirical results can be separated into several pairwise comparisons using the nesting categories described earlier. An elementary but little used pairwise test asks whether a set of empirical efficiency rankings could differ from a uniform distribution for which every efficiency score is unity. This one tailed test is applied to each of the empirical envelopments: DEA-CRS, FDH and SDEA. Another set of tests is applied to the null hypothesis that the empirical efficiency scores come from the same probability distribution. The pairs are: [DEA-CRS, FDH], [DEA-CRS, SDEA], and [FDH, SDEA]. The first two comparisons are one tailed again because of the nesting arguments developed earlier, while the last pairing is a two tailed test since there is no theoretical prediction about the nesting of SDEA and FDH. Finally another set of two tailed test can be applied to the comparison of the empirical models with the regulatory scores in DPCR. Siegel and Castellan (1988, p.144)) describe the hypotheses as follows. The null hypothesis, H_0 , is that the samples come from populations with the same probability distribution. In the one tailed test the alternative hypothesis H_I is that values in the population from which one of the samples is drawn are stochastically larger than the values of the population from which the other sample was drawn. In the two tailed test. 'larger' is replaced with 'different'. Formally the test statistic is written (Banker 1996, p.142):

$$D = \max\left\{\hat{F}_{n1}\left(\hat{\boldsymbol{q}}\right) - \hat{F}_{n2}\left(\hat{\boldsymbol{q}}\right)\right\}$$

where $\hat{F}_n(\hat{q})$ is the empirical cumulative relative frequency distribution function of the estimated efficiency scores for sample group *n*. Siegel and Castellan (1988) have tabulated the 1, 5, and 10 percent critical values of n_1n_2D and the relevant figures are shown in table 5.

Critical values for rejection of H_0 in the	One tailed test:	Two tailed test:
Kolmogorov-Smirnov test for two samples of 14		
10 per cent level	84	98
5 percent level	98	112
1 percent level	112	126

Table 5 Kolmogorov-Smirnov test critical values

The comparative pairwise test statistics are presented in table 6 and the results are

illuminating.

DEA-CRS	FDH	SDEA 0.95	SDEA 0.8	uniform
56	112*	28	42	168*
	112*	42	28	168*
		112*	112*	98*
			28	168*
				168*
	DEA-CRS 56	DEA-CRS FDH 56 112* 112* 112*	DEA-CRS FDH SDEA 0.95 56 112* 28 112* 42 112*	DEA-CRS FDH SDEA 0.95 SDEA 0.8 56 112* 28 42 112* 42 28 112* 112* 112* 28 28 112* 28 112* 28

Table 6 K-S test results: * reject H_0 at the 5 percent level

At the most basic level this size of sample does allow us to reject the null hypothesis that all of the companies are equally efficient and on the frontier. OFGEM's measured efficiencies – despite the criticism of the methodology imposed – are not significantly different from any of the other empirical models except FDH. The FDH model on this sample size is the only method which produces an efficiency distribution different from that of the other models including uniform total efficiency. However it only just rejects H_0 compared with uniform total efficiency at the 5 percent level and fails to reject it at the 1 percent level.. All of the non-FDH models including DPCR implicitly assume convexity – i.e. that companies are compared with interpolated efficient points which may not be observed. This results in the FDH model presenting the closest envelopment of the data and consequently the least discrimination amongst the companies. Both of the SDEA models produce an intermediate stage of envelopment – more discriminating than FDH but a closer envelopment than DEA-CRS. It appears that allowing for stochastic error is feasible and produces the expected results that each of the firms improves its position relative to the situation in which stochastic error is ignored. However the improvement relative to DEA-CRS is not so strong that the results fail to find efficiency differences amongst the companies. This suggests that SDEA is a useful and practicable means of overcoming the outliers problem in comparative efficiency studies.

Regulatory Implications

Are there significant regulatory conclusions to be drawn from this exercise? Three may be offered. First it is important to understand what comparative efficiency studies can achieve and why they are used. They do have a link with the theory of principal agent regulation and the critical idea is the nesting of different minimal cost extrapolation models. Second, it is clear that the stochastic DEA model can be implemented and produces results directly in accord with the model's predictions in the form of a closer probabilistic envelopment. Third, in regulatory studies small samples are not particularly helpful. This is a problem for both regulators and companies – but it is more serious for regulators. Companies obviously need to be assured that the comparative efficiency assessment is rigorous and plausible. Increased sample size can give better discrimination to the empirical results but carries two penalties. For the companies already in the comparison any increase in the sample size can never improve the companies' relative efficiency. Increased sample size can only produce more companies which are better or the same. Worse companies are irrelevant. To increase the sample size the regulator may

need to make international comparisons but the issue of whether data is comparable then makes the agents less likely to wish to participate in the game. It remains true that comparative efficiency regulation is not the straightforward task that it might appear at first glance. On the other hand, stochastic non-parametric efficiency measurement can provide feasible and incentive compatible improvements on deterministic models.

This paper has shown that different methods of comparative efficiency measurement can produce different results for the scores of the firms even within the same data set and input-output structure. In particular the case for closer envelopment of the data leads to the use of stochastic and non-convex methods of nonparametric efficiency measurment. It is possible to check very approximately what the effect of these changes in efficiency scores will be on the regulated prices that are applied to the firms. For example OFGEM (1999, annex 2 pp. 72-85) presented a stylised framework for the calculation of the P₀ initial prices for the next review period. Consider just two of the companies, labelled firms 7 and 9 in table 4. They each received low efficiency scores from OFGEM and even lower scores on DEA-CRS. However the FDH model increased their efficiency scores relative to both OFGEM and DEA-CRS and the SDEA model increased the efficiency score of one firm relative to DEA-CRS. Using the stylised calculations in OFGEM's P₀ reports it is possible to conjecture adjusted OPEX figures for these firms and therefore to make a very rough estimate of the effect on the regulated price outcome. The procedure we follow is this. The regulation requires that the present value of benchmarked costs (OPEX + depreciation + return) equals the present value of benchmarked revenues.

PV(OPEX + D + (wacc * RAB)) = PV(REVENUES)

Here *D* is the depreciation, *wacc* is the weighted average cost of capital and *RAB* is the regulatory asset base. The capped average revenue in current prices is given by the formula:

$$P_0(1+RPI-X)^t$$

Consequently:

[(+)]

it is possible to calculate the effect on P_0 of different benchmarks for OPEX. For the two examples in question the results are shown in table 7.

OPEX	Undiscounted OPEX: DPCR	Undiscounted OPEX: DEA	Undiscounted OPEX: FDH	Undiscounted OPEX SDEA
calculations £m	012012100	01 200 2 201	012111211	(0.95)
Firm 7	515	507	520	515
Firm 9	560	555	563	563

Table 7 OPEX recalculated with different models

In the case of firm 7 OPEX is 43 percent of the undiscounted costs and in the case of firm 9 OPEX is 52 percent of the undiscounted costs. The different models make a difference of 1-2 percent in the undiscounted OPEX figures so that the likely impact on the P_0 figure after discounting is probably a variation of around 1 per cent .

In conclusion the paper has investigated the arguments for different types of nonparametric efficiency measurement offering closer envelopment of the data than conventional DEA. In particular the effects of FDH compared with chance constrained DEA have been described. There are powerful arguments for the different models described here. Curiously in the small sample case study to which they have been applied the actual effects of choosing different methods of efficiency measurement have been small, but it would be wrong to expect this effect in larger samples.



timeline for Bogetoft yardstick competition

Figure 1 Time line for the Bogetoft yardstick competition game



Figure 2 boundaries of the FDH and DEA input requirement sets



Figure 3 Olesen and Petersen's model of chance constrained DEA

 \mathbf{X}_1