

**COMMENTARY ON CEPA
BENCHMARKING PAPER**

A Report for EDF Energy

Prepared by NERA

13 November 2003

London

**Project Team:
Graham Shuttleworth
Eнесе Lieb-Doczy**

n/e/r/a

National Economic Research Associates

Economic Consultants

15 Stratford Place

London W1N 9AF

Tel: (+44) 20 7659 8500

Fax: (+44) 20 7659 8501

Web: <http://www.nera.com>

An MMC Company

TABLE OF CONTENTS

1.	INTRODUCTION	1
2.	KEY OBSERVATIONS	2
3.	CEPA COMMENTS ON DPCR3 BENCHMARKING	4
4.	CEPA'S ANSWERS ON OFGEM'S KEY ISSUES	5
4.1.	Use of Benchmarking in Setting Revenues	5
4.2.	Input Data	6
4.3.	Benchmarking Techniques and Methodology	8
4.4.	Choice of Variables	13
4.5.	Benchmarking and Incentives	14
4.6.	The Treatment of Merged Firms and Returns to Scale	15
4.7.	Measures of Quality and Benchmarking	16
4.8.	Use of International Data	17
5.	CONCLUSION	19

1. INTRODUCTION

In this report, we review the report produced by CEPA for Ofgem on benchmarking¹

Ofgem commissioned the CEPA report “*to inform Ofgem in developing its approach on benchmarking and to provide a basis for Ofgem to begin consultation on these matters. Comments in the report reflect CEPA’s views and should not be regarded as Ofgem policy.*”² The report discusses alternative benchmarking techniques, including Ofgem’s approach in DPCR3. CEPA tries to apply the methodology used in DPCR3 to data from 2001/02 but the results suggest that it is not worthy of repeated use. The CEPA report also answers some questions which Ofgem identifies as key issues for consideration.

Ultimately, CEPA comes down in favour of Ofgem using a combination of DEA to benchmark companies and COLS (i.e. regression) to check the validity of the choice of variables. However, CEPA also warns against using any single method to set X-Factors or costs in a mechanistic way, because of the difficulties of applying any method.

This conclusion is unhelpful to Ofgem, which is left having to fund many benchmarking exercises, each of which is potentially of little or no value. It is also unhelpful to the DNOs, which are left not knowing how their performance will be appraised or how revenues will be set. CEPA refers to Dutch experience in this context, but many of the caveats that CEPA expresses about particular techniques of benchmarking undermine CEPA’s advocacy of benchmarking in general.

In the light of CEPA’s report, Ofgem may still feel obliged to carry out some benchmarking. However, Ofgem should not to use any particular results without extensive discussion of their meaning, and should look at other methods of identifying expected cost levels. Our preferred methods would still be to focus on individual and comparable expenditures (e.g. the cost of a standard transformer) as a way of checking capex, and to use TFP to set trend rates of growth in unit costs.

In Section 2 we set out our key observations on the CEPA document. Section 3 discusses general comments that CEPA makes about DPCR3. In Section 4 we discuss the views of CEPA on issues raised by Ofgem. Section 5 contains a short conclusion.

¹ CEPA (2003), *Background to work on assessing efficiency for the 2005 distribution price control review*, Scoping Study, Final report, CEPA, September 2003.

² Ofgem’s introduction to CEPA (2003).

2. KEY OBSERVATIONS

CEPA's discussion of benchmarking often fails to take into account the needs of regulation; in the report, CEPA therefore rejects or recommends particular methods for reasons that should play no part in a regulatory appraisal. We have identified two basic standards for regulatory decisions on price controls:

- The revenue standard: revenues should allow investors to cover their costs and to earn a rate of return comparable with that available from other sectors, in order to attract capital for long-term investment;
- The procedural standard: decision-making procedures should transparent and based on evidence, so that the resulting decisions are stable and hence robust to criticism from different points of view.

Together, these standards mean that regulatory procedures offer a "reasonable prospect of cost recovery", as required to promote efficient investment in and operation of the networks. However, CEPA's analysis violates both these standards:

1. **Failure to abide by the revenue standard:** CEPA explore ways to set X-factors using comparisons with a supposed efficiency frontier; however, Ofgem is likely to offer a rate of return based on normal or average returns in the stock market. This combination is not comparable with the returns on offer in other sectors, where companies which achieve "efficient costs" earn *more than* the normal or average rate of return. If regulated companies cannot offer investors a comparable rate of return, they will be unable to attract capital for the investment needed to maintain an efficient quality and level of service to consumers.
2. **Failures to abide by the procedural standard:** Ultimately, CEPA suggests that benchmarking results cannot be converted directly into X-factors, which is a sound conclusion, but CEPA still makes recommendations in passing which are not supported by experience or logical analysis. CEPA continually refers to benchmarking results as a measure of "efficiency", without acknowledging that the scores could be affected by errors in data measurement and model specification. CEPA also rejects some methods because of "difficulty", but does not consider whether the difficulty is worth overcoming to achieve real regulatory goals. In these cases, CEPA is not arguing logically or transparently and Ofgem should not rely too heavily on CEPA's recommendations.

Because CEPA does not apply proper regulatory standards, the report contains recommendations that have no regulatory or economic logic. Having decided that benchmarking should use a measure of total costs ("totex"), CEPA proposes a definition based on the sum of (1) base opex (as defined by Ofgem), (2) regulatory depreciation of the Regulatory Asset Base (RAB) and (3) the allowed rate of return on the RAB. CEPA rejected other definitions of totex because they were "difficult" to compile. However, the use of RAB

to define capital costs will inevitably produce results that bear no relation to efficiency, since the RAB is determined by numerous past regulatory decisions and historical accidents. Hence, CEPA assigns value to using a particular method when its results will be entirely useless. Ultimately, CEPA is forced to conclude that no single method of benchmarking provides a reliable estimate of efficiency, but instead of recommending that benchmarking be ignored, CEPA concludes that Ofgem should use a variety of methods, as if a number of unreliable results provides useful information in aggregate. In fact, the use of multiple methods just forces Ofgem into an unpredictable exercise of regulatory discretion.

Ofgem may still feel that benchmarking will provide some comfort or will help inform its *discussion* of DNO costs. However, these observations mean that benchmarking is an unsound basis for regulatory *decisions* on revenue allowances. In order to provide clear incentives for efficiency, Ofgem would need to develop a method of setting revenue allowances that only uses other, more objective, sources of information on actual costs.

3. CEPA COMMENTS ON DPCR3 BENCHMARKING

CEPA describes how in 1999 “[Ofgem’s] *final assessment of the efficiency frontier was based on a combination of frontier analysis, expert industry judgement about fixed costs and a decision that the most efficient firm would not be used to determine appropriate levels of efficiency for other firms.*” CEPA concludes, “*Given the uncertainties surrounding benchmarking techniques and the drawbacks of relying solely on a single methodology, such a pragmatic approach appears robust.*”³

However, CEPA then undermines this claim for “robustness” by finding fault with the method that Ofgem used: “*Due to the lack of transparency about the final adjustments made to the methodology and data, we were unable to replicate results (of DPCR 3) precisely.*”⁴ CEPA criticizes in particularly the following aspects of the 1999 approach:⁵

- Unclear adjustments to raw data which introduce a degree of endogeneity into the results.
- Unclear determination of weights
- Unorthodox determination of frontier: “*The regression methodology involved a pivot in the OLS regression line rather than the vertical shift that is standard under COLS methodology. Neither approach can be said to result in a more accurate assessment of the frontier.*”
- Excessive reliance on the position of Eastern
- Use of a purely domestic data set
- Benchmarking opex alone rather than total cost
- Subsequent developments
- Insufficient explanation of outliers

Putting together these arguments, it appears that CEPA only thinks Ofgem’s 1999 method was “robust” because it did not rely entirely on the numerical analysis (or indeed, any particular analysis) and instead allowed for the exercise of regulatory discretion. This is an unorthodox use of the term “robust”, as it clearly has little to do with the transparency, stability or predictability of results.

³ CEPA (2003), page 6.

⁴ CEPA (2003), page 52.

⁵ CEPA (2003), page 52.

4. CEPA'S ANSWERS ON OFGEM'S KEY ISSUES

4.1. Use of Benchmarking in Setting Revenues

CEPA is aware of recent Dutch experience with benchmarking and warns against using benchmarking in a mechanistic way to determine X-factors or future revenues:

“Benchmarking is an important regulatory tool that can be used to assist with judgements about the scope for efficiency improvements across firms. However, it is only a tool and cannot substitute for judgement based on a wider range of evidence. It should therefore only be considered as an input into a pragmatic approach to setting X factors that draws on a range of analysis.”⁶

The trouble with a “pragmatic approach” is that it can become an entirely subjective choice between different data for setting revenues, especially if different methods offer widely differing estimates of key parameters. CEPA recognise this and other criticisms of regulation by benchmarking, but offer no real solution:

“Although widely used, benchmarking is not universally accepted as part of the regulatory process. For example, some commentators have argued that the results of benchmarking are the result of arbitrary choices about details of the techniques, and therefore inappropriate for the determination of price controls. Particular forms of benchmarking approaches can also distort decisions by utilities. If benchmarking is to be used in the regulatory process, the approach must be chosen with care.”⁷

We agree with these observations on the criticisms of benchmarking, but we do not believe that “choosing with care” will overcome the basic flaws in the approach. CEPA suggests the following criteria for the selection of a benchmarking method, but even these criteria do not provide much support for objective decision-making:⁸

- Practical application
- Robustness
- Transparency and verifiability
- Ability to capture business conditions adequately
- Restrictions should be minimized
- Consistency with economic theory

⁶ CEPA (2003), page 99.

⁷ CEPA (2003), page 10.

⁸ CEPA (2003), page 10.

- Consistency with non-frontier approaches
- Non-excessive regulatory burden

Few of these criteria relate clearly to either the revenue standard or the procedural standard. (The last criterion even seems to suggest an acceptance of short-cuts to help the regulator, even if the results are incorrect.) The desire for robustness is laudable, but is not supported by the discussion of benchmarking techniques in CEPA's report. Indeed, although CEPA favours DEA for its perceived practical advantages, CEPA acknowledges that no method is likely to provide results that are sufficiently robust to set X-factors directly:

“DTe’s experience of benchmarking underlines the susceptibility of DEA to the choice of input and output variables and highlights the risks associated with relying on a single technique for determining the efficiency frontier, especially when the results feed directly into the regulatory determination.

In many instances the efficiency gap implied by the regulator is used to set company-specific X factors, e.g. Netherlands electricity distribution. However, in other cases, (e.g. NSW distribution, UK transmission) the results of the DEA analysis are just one of several factors used to determine the X factors or, as in the cases of Finland and Sweden, do not explicitly drive the regulatory process at all.”⁹

Thus, although CEPA favours DEA for benchmarking purposes, the report is quite clear in warning against relying exclusively on one benchmarking methodology only and leaves Ofgem to decide what influence benchmarking results should have on X-factors. This attitude greatly reduces the value of any benchmarking, but CEPA nevertheless analyses the relative merits of different techniques, by reference to its own (limited) set of criteria.

4.2. Input Data

CEPA suggests that Ofgem should benchmark total costs (referred to as “totex”) rather than operating expenditures (“opex”), because CEPA is aware of the distortions caused by looking at only a subset of costs:

“companies that have, in the past, invested in equipment and technology that reduces operating costs will appear to be more efficient than those who have not done so, irrespective of whether the capital /operating expenditure trade-off does actually lower overall costs.”¹⁰

⁹ CEPA (2003), page 23.

¹⁰ CEPA (2003), page 96.

However, CEPA immediately ignores this problem, and allows Ofgem to abandon the use of "totex", if it proves too difficult to put into practice:

*"Ideally, therefore, efficiency should be benchmarked in terms of totex. However, complications involved in measuring the capital expenditure element of totex mean that this is not straightforward. Further analysis is required to assess whether a reasonable totex variable can be constructed that is not itself subject to distortion from gaming. Should benchmarking totex prove inappropriate, an alternative would be to use totex to assess the divergence in opex performance rather than acting as the benchmark for establishing X. factors per se."*¹¹

CEPA does not say what kind of "gaming" is involved (differences in accounting practice, perhaps). However, CEPA argues that even if totex benchmarking is "inappropriate" (an unspecified criterion that can mean anything), Ofgem might still use measures of totex to explain the results of opex benchmarking. CEPA does not say how using totex can be "appropriate" in the second case, when it is "inappropriate" in the first case. Since the only test can relate to data quality, doubts about the validity of totex measures would apply equally to both uses.

CEPA suggests the following definition of totex:¹²

$$\text{Totex}(t) = \text{base opex}(t) + \text{depreciation}(t) + \text{ROC}(t) \times \text{RAV}(t)$$

where ROC is the allowed return on capital (6.5% real) and RAV is the regulatory asset value, and depreciation is regulatory depreciation.

CEPA re-ran the DPCR3 benchmarking on a total cost basis and found that *"the most efficient firm remained the same under the two measures, some of the least efficient firms did improve their scores significantly; indicating that using opex as the measure of efficiency may miss out important factors that are relevant to customers."*¹³ However, CEPA itself points out that Ofgem's adjustments to the raw data on opex lacked transparency and made replication of the DPCR3 benchmarking results impossible.¹⁴

Some of CEPA's criticisms of DCPR3 will apply to methods adopted in DCPR4 and Ofgem should be mindful of them. However, there is no reason for Ofgem to accept all CEPA's recommendations for DCPR4. In particular, the use of the RAB to define capital costs is indefensible. The RAB is distorted by past regulatory decisions and historical accidents. Although the pre-privatisation assets are starting to drop out of companies' RABs, many of the assets are still in use. Moreover, pre-privatisation investment affected subsequent

¹¹ CEPA (2003), page 97.

¹² CEPA (2003), page 85.

¹³ CEPA (2003), page 7.

¹⁴ CEPA (2003), page 52.

investment needs, and hence the current value of the RAB. Thus, the RAB offers little guidance on the efficiency with which each DNO is using its resources, or has invested since privatisation. Although we remain sceptical about the value of any benchmarking technique, we would have expected any attempt to assess efficiency or efficiency growth to use a measure of total assets employed by the company, i.e. either a registry of assets (for DEA or TFP) or an estimate of replacement costs that applies common valuation principles to all assets.

4.3. Benchmarking Techniques and Methodology

CEPA admits that *“the choice of technique can have an impact on the determination of efficiency scores and depends at least partly on the data available and the aims of the benchmarking exercise”*.¹⁵ That admission would seem to undermine the rationale for using benchmarking and any claim to robustness for the results. Moreover, CEPA's appraisal of the different methods contains a lot of judgments that are either not applied universally, or which undermine the rationale for benchmarking.

4.3.1. Stochastic Frontier Analysis (SFA)

CEPA describes SFA as *“statistically the most elegant”* approach to benchmarking, but rejects it in the context of DPCR4. CEPA says that the small number of UK DNOs means that SFA techniques are *“unable to distinguish between the efficiency of firms and noise”*¹⁶

*“Regulators have therefore traditionally been reluctant to use SFA techniques in setting X factors. This is because in small samples the technique is either difficult to implement or gives rise to high efficiency scores.”*¹⁷

Rejecting a statistically sound method because the results are unacceptable would violate the standards of good regulatory procedure. CEPA maintains that SFA encounters problems with small samples, but does not explain why any other method would be better able to overcome the need *“to distinguish between the efficiency of firms and noise”*; in practice, other methods (particularly DEA) just hide the problem. Hence, CEPA's comments on SFA contain a criticism of benchmarking in any form, when the sample size is small.

4.3.2. Parametric Programming Approach (PPA)

CEPA reviews PPA (a method with which we are not familiar), but rejects it on theoretical grounds, arguing that *“PPA suffers from most of the disadvantages of DEA, and in particular that the frontier position is vulnerable to precise variable specification, but it does not have the*

¹⁵ CEPA (2003), page 12.

¹⁶ CEPA (2003), page 94.

¹⁷ CEPA (2003), page 37.

*compensating advantages of econometric approaches like COLS.*¹⁸ Although CEPA is probably justified in rejecting PPA, the report adopts an odd line of reasoning; reversing the words “PPA” and “DEA” in the previous sentence would provide an equally strong reason from rejecting DEA. However, CEPA ultimately chooses DEA as its preferred method. CEPA’s method of appraisal therefore seems to lack an overall consistency.

4.3.3. Constrained Ordinary Least Squares (COLS) Regression

CEPA’s main concern with COLS is that *“the frontier intercept is sensitive to outliers”* – in fact, the level of the COLS frontier depends exclusively on a single company. (The gradient is taken from a regression over all companies.) CEPA suggests that *“the DEA approach is theoretically more appealing than COLS as it determines efficiency using different input and output variables”*.¹⁹ We share CEPA’s concerns about COLS, given the widespread doubts about comparability of data and output measures.

4.3.4. Total Factor Productivity (TFP) indices

CEPA reviews the use of TFP in the United States and elsewhere, but the report exhibits great confusion over the purpose of such analysis. TFP is a tried and tested method for estimating long-run historical trends in productivity growth as the basis for forecasting a reasonable target for productivity growth in the future. U.S. regulators add a “stretch factor” (of only one percent or so) if they can demonstrate that the future is likely to be different from the past in a way that will accelerate productivity growth (e.g. because a price cap is replacing cost pass-through arrangements and is therefore strengthening incentives). However, if companies manage to beat this reasonable target by cutting costs faster, it is understood that the additional profit represents a reward (above the normal or average rate of return) for exceptional performance. Thus, the US method does not violate the revenue standard by trying to impose “efficient” cost targets and a normal or average rate of return.

CEPA’s comments on TFP indicate a large gap between the US view of regulation (and our definition of the revenue standard) and, on the one hand, the current parochial view of regulation within the UK, on the other. For instance, CEPA writes:

*“The index approaches examined do not provide estimates of the efficiency frontier. They do, however, show that companies are improving efficiency at very different rates. In the long-term, X factors should ideally be based on TFP. However, until there is sufficient convergence in firms’ performance, it is not appropriate to use them in this manner. In the meantime, TFP may provide a useful methodology for assessing shifts in the frontier.”*²⁰

¹⁸ CEPA (2003), page 25.

¹⁹ CEPA (2003), page 95.

²⁰ CEPA (2003), page 96.

Here, CERA seems to be implying that TFP methods are only useful if all companies are expected to improve their efficiency at the same, long-run rate. Any such conclusion would be wrong.

As explained above, the revenue standard is inconsistent with focusing on frontier performance and US regulators are required only to set a reasonable target, not to squeeze every last ounce of "potential" performance into the definition of price caps. After all, the incentive for efficiency derives from the fixed nature of price caps, which allows companies to increase profits by cutting costs; setting price caps lower is not a source of additional incentives to reduce costs. CEPA's apparent confusion on this point is visible in other extracts:

"As a longer term objective tying X to the TFP growth rate in the distribution sector might be a desirable goal".²¹

"Analysis of a total factor productivity (TFP) index showed wide disparity in the performance of firms, showing that it is premature to use these directly to set X factors."²²

"TFP would be useful in order to assess the trend in efficiency over time. This is important in assessing whether X-factor targets are being met."²³

Taken together, these comments show that CEPA is trying to set different X-factors by reference to measures of current performance, rather than by trying to estimate a reasonable trend for the future. Implicit in this approach (which is quite common) is the "catch-up hypothesis", which we characterise as consisting of three assumptions about firms identified by benchmarking as high cost:

1. they are without doubt "inefficient";
2. they find it "easy" to reduce costs rapidly; and
3. they do not deserve to benefit from such "easy" cost reductions, or at least not more than other "efficient" firms.

Each of these assumptions is incorrect. First, benchmarking cannot unequivocally identify levels of efficiency; if it did, the regulator should cut P0 now, to disallow the excess costs, instead of waiting for years to eliminate them from prices through an X-factor. Second, high cost firms may not be able to reduce costs more rapidly than others, either because they are not inefficient or because there is no reason to expect them to outperform other companies. Third, in any case, companies need a profit incentive to reduce costs and the extra profits

²¹ CEPA (2003), page 27.

²² CEPA (2003), page 6.

²³ CEPA (2003), page 27.

arising from beating X-factors are never excessive in practice (since such a large proportion of total costs are sunk and cannot be reduced anyway).

CEPA repeats some of the unjustified prejudice that informs regulatory debate within the UK, by arguing that some regulatory methods yield x-factors which are "too low":

"The result of using TFP measures in the US is to impose extremely low efficiency targets, by UK standards. These targets (0-2%) look unchallenging in UK context and coupled with extensive revenue sharing agreements result in a lack of pressure to cut utility costs in the US."²⁴

However, CEPA does not set out by what standard these targets are "too low". Moreover, CEPA fails to explain (1) that the TFP figures quoted are most likely *net* of the TFP growth in the economy as a whole (which appears in the RPI element of RPI-X formulae) so that the overall target is higher; (2) that much of the past reduction in DNO prices has come from the redefinition of costs, and not from efficiency gains, so there is no basis for comparing past X factor and P0 cuts with long-term efficiency gains; and (3) if objective analysis shows that productivity grows slowly, regulators should strive harder to justify high X-factors, rather than rejecting the analysis.

Overall, therefore, we are forced to conclude that CEPA's attitude to TFP is ill-informed and prejudicial and therefore lacks objectivity and credibility.

4.3.5. Data Envelopment Analysis (DEA)

CEPA adopts DEA as its preferred method. Its arguments in favour of DEA are practical rather than theoretical, but unfortunately they are weak.

CEPA suggests that "*DEA can be implemented on a small data set*".²⁵ Of course, regression analysis can also "be implemented on a small data set", but one can test regression for the significance of results derived from small samples. Rather than resolving the problem, DEA hides the problems arising from small samples, by offering no test statistics.

The argument that "*inefficient firms are compared to actual firms rather than some statistical measure*"²⁶ does not hold water either. DEA compares firms with linear combinations of firms on the efficiency frontier, not actual firms. It can be just as difficult for individual firms to identify key differences between themselves and a few others, as between themselves and all others – especially when the range of comparators changes from one model to the next. Hence, DEA does not alleviate the burden of proof borne by benchmarked firms.

²⁴ CEPA (2003), page 28.

²⁵ CEPA (2003), page 19.

²⁶ CEPA (2003), page 19.

CEPA argues that *“DEA can account for factors that are beyond the control of the firms but affect their performance”*²⁷ but this holds equally for regressions any other methods. The suggestion that *“DEA has the advantage that it...can be illustrated easily”*²⁸ is not a serious argument. On the other hand, on pages 20 and 21 CEPA lists eight problems with DEA, which are extensive and well documented.

4.3.6. CEPA's Conclusion: DEA and COLS

By a process of elimination, CEPA concludes that DEA and COLS are the main practical alternatives for determining the efficiency frontier:²⁹

*“[...] one possible approach to establishing an appropriate efficiency frontier would be to use a combination of DEA and COLS. In particular, emphasis could be placed on the DEA scores but with COLS being used to assess the appropriateness of the output variables used, the significance of the DEA efficiency scores obtained and assess whether particular companies were being treated unfairly under DEA. The choice of technique may, however, be less important than the choice of the variables to be included in the benchmarking exercise.”*³⁰

The reference here to COLS is surprising, since the process of testing variables uses ordinary least squares (“OLS”) regression; shifting the frontier as in COLS is entirely superfluous for this purpose, and seems to be intended only to maintain the fiction that regulatory methods should only ever examine “the frontier”.

However, CEPA admits a problem with DEA: *“the calculated efficiency scores are dependent on the variables selected, and the method itself does not provide a test of whether particular variables should be included in the model.”*³¹ Once again, this admission would lead many to outright rejection of benchmarking with DEA, as lacking any practical value in regulation, but CEPA suggests that:

*“[DEA] would seem to be particularly applicable in the early stages of regulation when not much is known about the potential for cost reduction and the underlying efficiency frontier. Electricity Distribution in the UK would seem to be a good candidate for DEA given the existence of similarly organised regional utilities adhering to similar technical standards. If DEA can be used in any regulated industry in the UK electricity distribution would seem to be the one.”*³²

²⁷ CEPA (2003), page 19.

²⁸ CEPA (2003), page 20.

²⁹ CEPA (2003), page 94.

³⁰ CEPA (2003), page 96.

³¹ CEPA (2003), page 95.

³² CEPA (2003), page 21.

Frequent use of “would seem to be” does not inspire confidence in CEPA’s recommendation of the DEA methodology. Moreover, CEPA’s arguments do not bear close scrutiny. The DNOs are not necessarily “similarly organised”. Following “similar technical standards” is not sufficient to allow comparisons, if costs and outputs depend upon a large number of other variables. If these arguments were true, they would contradict CEPA’s suggestion elsewhere in the document that Ofgem should consider using data on networks in other countries. Thus, CEPA’s stated position is compromised by its own uncertainty and internal contradictions.

4.4. Choice of Variables

CEPA provides the following table of the input and output variables used in DEA based regulatory benchmarking exercises.

Figure 5: Examples of input / output variables used for DEA analysis by electricity distribution regulators

Regulator	Inputs used	Outputs used
Norway	Capital (book value and replacement cost), goods/services, losses, labour	Number of customers, energy delivery, length of line and sea cables
Netherlands	Opex	Units, peak demand HV, peak demand LV, network length, customers small, customers large
NSW, Australia	Total operating and maintenance costs, transformer capacity, network size	Electricity sold, customers, peak demand

Source: Jamasb & Pollitt, 2001

The adoption of different variables in different countries suggests that distribution systems differ significantly between countries (which undermines the desire to use data from other countries for benchmarking British DNOs). Alternatively, one might conclude that this table illustrates a degree of randomness in specifying DEA models, due to the lack of any direct method of testing for specification errors. CEPA admits that benchmarking scores depend upon the choice of variables: “**Misspecification of variables can lead to perverse results, potentially with less efficient firms defining the frontier.**”³³ This comment confirms the lack of objectivity and robustness in the results of DEA.

4.4.1. Cost drivers

CEPA acknowledges that numerous factors drive the costs of a distribution business, including customer density, landscape and climate. However, when CEPA conducted a

³³ CEPA (2003), page 18.

second stage analysis of these factors, trying to explain variations in DEA scores, they could not detect any statistically significant effects. CEPA argues that this might be due to the small samples size and suggests that *“this does not means that these factors should be ignored”*.³⁴

Indeed, when re-running DPCR3 to 2001/02 data, CEPA finds that *“the resultant estimated potential for opex improvement varies considerably across companies, far more so than under the 1999 review [...] This raises questions about the appropriateness of the methodology, in particular the selection of cost drivers and the benchmarking of opex”*.³⁵

4.4.2. Panel data

CEPA emphasises that the use of panel data offers numerous advantages, including a larger number of degrees of freedom. Panel data uses observations for (e.g.) the 14 DNOs dating back over (e.g.) the last 10 years. Panel data provides the basis for more reliable estimates of regression and other parameters, although interpretation of the results may require strong assumptions (e.g. over the effect of time trends on individual parameters). CEPA says *“the use of panel data would therefore provide Ofgem with a way to overcome some of the limitations currently encountered due to the small sample size for UK electricity distribution.”*³⁶

Our experience that the use of extensive panel data in TFP methods is the only statistically reliable method for identifying efficiency trends in the distribution sector. However, use of panel data in unreliable methods (that calculate relative efficiency scores rather than trends) will still produce unreliable results. Furthermore, the value of panel data on DNO costs will be reduced, if it does not allow for changes in accounting methods over the relevant period.

4.5. Benchmarking and Incentives

CEPA adopts a highly idealised view of benchmarking:

“[Benchmarking] can be used to strengthen the incentives facing regulated firms by rewarding them financially for closing the gap between their actual and potential efficiency. It may also reduce the cost to regulators of making judgments about efficiency compared to other methods”.³⁷

In practice, benchmarking does nothing to change incentives by itself, since the incentives derive from the price caps and the process of setting them. Since CEPA has warned against using benchmarking directly to set price caps, benchmarking cannot influence incentives (even through price caps) to any great degree. Moreover, a transparent and predictable

³⁴ CEPA (2003), page 98.

³⁵ CEPA (2003), page 57.

³⁶ CEPA (2003), page 99.

³⁷ CEPA (2003), page 9.

regulatory method would not rely on regulators "making judgments about efficiency". Regulation that relies heavily on such judgment will not provide efficient incentives, because regulated companies will be unable to understand what kind of behaviour maximises their returns, and may not decide to adopt efficient methods of operation.

The CEPA report shows some confusion about the source of efficiency incentives, due to its acceptance of the "catch-up hypothesis" (i.e. the notion that companies identified as 'inefficient' show potential for greater efficiency improvements - due to their larger scope to catch up with best practices identified by benchmarking - see above). CEPA's re-run of DPCR3 with 2001/2 data suggests that companies' relative performance did not conform with the "catch-up hypothesis":

*"In retrospect the changes in operating costs across firms have not been as the 1999 analysis would suggest. Firms operating significantly away from the frontier would have been expected to have shown greater improvement in operating efficiency than their counterparts operating close to the frontier in 1999, resulting in convergence in efficiency scores. This does not appear to have been the case from the 2001/02 data and may suggest that the frontier was not correctly defined."*³⁸

CEPA's reference to the incorrect definition of the frontier provides a universal excuse for maintaining the "catch-up" hypothesis. An alternative explanation of the data is simply that no regulator can ever assume that the lowest scoring companies can, will or should be expected to reduce their costs fastest in future, because benchmarking does not measure "potential efficiency gains".

4.6. The Treatment of Merged Firms and Returns to Scale

CEPA poses a question about scale effects as it has arisen before in the Dutch context, but which has up till now played little part in UK regulation:

*"In the context of recent mergers and the increasingly international nature of the industry, is scale considered to be a variable of choice by companies? If so, Ofgem needs to consider whether it should consider restricting the use of benchmarking techniques to those using constant returns to scale to encourage firms to choose an appropriate scale themselves."*³⁹

In a small market like the UK, company size cannot be entirely up to managerial decisions, but is also subject to competition policy concerns. It is unclear how British competition authorities would view further consolidation in the distribution sector. The merger of the DNOs in the East Midlands and Midlands regions removes the last remaining 'sole operator', such that any further mergers would involve major consolidation and not all DNOs can expect to be allowed further mergers. Thus, the answer to CEPA's question

³⁸ CEPA (2003), page 53.

³⁹ CEPA (2003), page 99, see also page 65.

should be plain – that companies now have little scope over their size because scope for further merger is limited. Furthermore, CEPA observes that individual DNOs have little choice over the size of their operations within any region:

“The appropriate approach with respect to benchmarking for electricity distribution is to focus on costs rather than production. This is because, given the universal service obligations and defined territories of the DNOs, the level and mix of output is basically determined exogenously for each company.”⁴⁰

CEPA observes that (merger) ***“savings could clearly impact the efficient level of costs and failure to take the effect into account may result in Ofgem treating firms who have not merged unduly harshly or merged firms unduly leniently.”⁴¹*** Using OLS and COLS approaches based on the DPCR3 methodology CEPA finds, however, that:

“the efficiency scores for the three DNOs that have not merged with another DNO improve dramatically once DNO groupings are assessed rather than individual firms. Indeed, one of the three unmerged DNOs is now the most efficient.”⁴²

This might mean that merged companies are less efficient, or else that the DPCR3 methodology is misspecified. Thus, the analysis adds nothing to overall understanding and its interpretations will depend largely on prejudice.

It is an anomaly of DEA analysis that the combination of two companies automatically reduces their DEA score.⁴³ Thus, if DEA was used to assess the ‘efficiency’ of merged entities, the scores of the merged companies would decline relative to the scores obtained by companies which have not merged.

In practice, therefore, Ofgem has no grounds for making strong assumptions about mergers and the associated potential to reduce costs.

4.7. Measures of Quality and Benchmarking

When CEPA conducted a second stage regression analysis of DEA efficiency scores it found that the impact of quality was insignificant. Hence CEPA suggests that ***“too many other factors are contributing to quality (such as topography), so that quality itself cannot be used as a***

⁴⁰ CEPA (2003), page 11.

⁴¹ CEPA (2003), page 90.

⁴² CEPA (2003), page 92.

⁴³ In the most recent regulatory benchmarking exercise in the Netherlands, Nuon insisted that its constituent companies should be modelled separately, with the result that its benchmarking scores improved considerably.

variable in the analysis. This suggests that maintaining a separate program such as the IIP to incentivise quality improvements is appropriate".⁴⁴

The finding underlines the difficulty of using benchmarking in any practical way, and we agree with CEPA's general conclusion that quality incentives are necessary, whilst noting that it does not contain any specific endorsement of the IIP in particular. Ofgem's attempts to build service quality variables into price caps has highlighted several problem areas:

- Regulatory parameters can capture only some aspects of supply quality and will hence only provide partial incentives.
- The "punishment" implied by revenue reductions often bears no relationship to the costs that companies incur to provide additional supply security or the valuation that customers place on additional security.
- Quality can vary from year to year due to exogenous factors, such as weather conditions. Hence incentives have to be weakened to avoid unjustified revenue shocks to companies.
- Regulatory parameter are part of short- to medium-term price caps, whereas investment incentives are determined by long-term factors, such as the prospects of cost recovery at future reviews.

An attempt to include quality parameters in a benchmarking exercise would face all of these problems and more. As a result, quality differences remain another factor (apart from efficiency) that may explain differences in benchmarking scores.

4.8. Use of International Data

Despite commenting on the desirability of using benchmarking in relation to similar companies, CEPA repeatedly refers to the advantages of using international data.

*"Given the likely uncertainty about the position of the frontier, international data may improve the estimate, provided that the underlying nature of the business is sufficiently similar. Despite the difficulties of international comparison, even the inclusion of a very limited additional sample of companies could have an important impact."*⁴⁵

The use of international data in regulatory benchmarking frequently poses more problems than it solves. The main problem is comparability of data, especially since countries not only differ in the exogenous cost drivers faced by companies but also in the unbundling, cost

⁴⁴ CEPA (2003), page 98.

⁴⁵ CEPA (2003), page 99.

allocation, depreciation and taxation principles used at the distribution level (leading to cost bases which are not comparable).

Attempts to correct for these discrepancies are likely to involve numerous subjective adjustments to the original cost data, leading to non-transparency and lack of objectivity in the benchmarking results. The analysis may also require additional variables to account for differences between working environments, but research has shown that the number of observations required to produce reasonably accurate results under DEA increases dramatically when the number of variables increases.⁴⁶ Adding a new data set may actually provide too few new observations to make up for the necessary additional variables.

Thus, although using international data seems like a useful way to expand the data set, in practice it may create more problems than it alleviates.

⁴⁶ Pedraja-Chaparro F, Salinas-Jimenez, J, and Smith, P (1999) "On the Quality of the Data Envelopment Analysis Model" *Journal of the Operational Research Society*, Vol 50, pp 636-644

5. CONCLUSION

The CEPA report takes an odd view of benchmarking, at some points regarding it as a robust method, producing clear indications of efficiency or potential cost reductions, and at other points undermining such claims by pointing out the deficiencies of the analysis.

CEPA's main conclusion is that Ofgem should carry out benchmarking using a variety of techniques, but should not use any single one to set X-factors. This conclusion is unhelpful to Ofgem, which is left having to fund many benchmarking exercises, each of little or no value. It is also unhelpful to the DNOs, which are left not knowing how their performance will be appraised or how revenues will be set.

However, the implications of CCPA's work stand out clearly. Ofgem should not use estimates of 'efficiency' taken from benchmarking exercises to set cost allowances, revenue allowances or X-factors. Instead, Ofgem should encourage greater efficiency by adopting more transparent and objective methods, based on observable cost data and reasonable target for unit cost reduction. Such methods are consistent with the principles of price cap regulation, which would allow DNOs to earn higher returns by beating the targets.