CE ELECTRIC FUNDING COMPANY

DISTRIBUTION PRICE CONTROL REVIEW

COMMENTS ON THE FINAL REPORT PREPARED FOR OFGEM BY CAMBRIDGE ECONOMIC POLICY ASSOCIATES: *BACKGROUND TO WORK ON ASSESSING EFFICIENCY FOR THE 2005 DISTRIBUTION PRICE CONTROL REVIEW*

SUMMARY

- Benchmarking can play a part in assessing efficiency for the 2005 distribution price control review.

- To ensure that any benchmarking is carried out in a responsible manner adjustments to raw data must be made for:

    * differences in the reporting of costs;
    * differences in operating conditions faced by each DNO;
    * differences in measurable quality outputs delivered by each DNO; and
    * differences in the risk borne by each DNO as a result of choices made about cost reduction.

- The more robust and complete the analysis, the more reliance may be placed on the results.

- Prices should not be set on the basis of frontier costs which are likely to capture the short-run non-sustainable costs of a particular distribution network operator in the chosen year.

- Using a competitive market standard in benchmarking is informative. This standard indicates that in competitive markets firms operate at some distance (10-20 per cent) away from the frontier.

- Using average costs in setting prices, although not without difficulties, is safer than using frontier costs.

- Using frontier assessments to determine costs is analogous to a regulatory body exercising monopsonist powers. The exercise of monopsony power is generally not regarded as desirable in advanced industrial societies.

- Use of total costs in benchmarking is preferable to using operating costs in isolation because it captures trade offs between operating and capital cost solutions and because it corrects for differences in classification of costs that have not been discerned by the regulator. However the total costs assessment should be used to assist in forming a judgement about efficiency rather than to adjust the regulatory asset value.

- *Pace* Cambridge Economic Policy Associates (CEPA), Stochastic Frontier Analysis (SFA) could play a useful part in the 2005 price control review.

- Data Envelopment Analysis (DEA) has been shown to exaggerate inefficiency scores and, in the Netherlands, it yielded no results on which the regulator could rely.

- Corrected Ordinary Least Squares analysis places undue emphasis on the results of one (or very few) companies. It is unsafe to use this technique to determine allowed income.

- Trends in Total Factor Productivity (TFP) are unlikely to assist in determining efficient costs because the significant rate of TFP gains achieved during the period since privatisation are not an indication of the likely future trend.

- In any case there is no longer (if there ever was) a need to try to *anticipate* future efficiency gains. It should be sufficient to rely on the capital markets to drive out inefficiency. These gains can be passed to customers *ex post*.

- Any benchmarking techniques employed must be capable of hypothesis testing. The relevant hypothesis is that the subject company is on the efficiency frontier. If that hypothesis cannot be demonstrated to be false, with a reasonable level of confidence, it is inappropriate to presume that the company is inefficient.

- If DEA is to be used it should be enhanced to include modules that determine the statistical significance of the efficiency estimates that it produces.

- Panel data could be incorporated to improve the robustness of the benchmarking conclusions.

- Customer numbers cannot be excluded from the scale parameter in the way suggested by CEPA.

- Quality measures should not be excluded from the assessment.

CONTENTS

INTRODUCTION

1 CE Electric UK Funding Company (CE) is the UK parent company of Northern Electric Distribution Ltd (NEDL) and Yorkshire Electricity Distribution plc (YEDL). The views expressed in this submission represent the response of CE, NEDL and YEDL to the final report prepared for Ofgem by Cambridge Economic Policy Associates (CEPA) entitled *Background to Work on Assessing Efficiency for the 2005 Distribution Price Control Review,* September 2003, (the CEPA report).

2 We welcome the initiative taken by Ofgem in commissioning the CEPA report. Benchmarking clearly played an important part in the last distribution price control review (DPCR3) and Ofgem is minded to use benchmarking techniques to inform its conclusions in the current price control review (DPCR4). The CEPA report raises a number of important issues that are central to the responsible use of benchmarking at price control reviews. We are particularly grateful to Ofgem and to CEPA for giving distribution network operators (DNOs) the opportunity to meet with CEPA to gain a deeper understanding of the techniques used by CEPA and some of the important prior assumptions that CEPA made in preparing its report.

3 In our response we have set out our views on each of the items singled out by Ofgem as 'key issues for consideration'. However, we believe that there are some fundamental issues that must be addressed before consideration is given to the details of input data, benchmarking techniques, the use of frontiers or averages and other issues. We have therefore set out our analysis of the fundamental issues that we believe must be addressed if benchmarking is to play a significant part in DPCR4. Much of this analysis draws on the answers given by Ofgem's consultants, Frontier Economics and CEPA, in response to questions put to them at the Ofgem workshops.

PART ONE – GENERAL OBSERVATIONS

PRECONDITIONS FOR THE RESPONSIBLE USE OF BENCHMARKING TECHNIQUES AT A PRICE CONTROL REVIEW

4 The CEPA report follows Ofgem's nomenclature from DPCR3 and refers throughout to the 'efficient' or sometimes to the 'efficiency frontier'.

5 Before a company, or set of companies, may be considered to form an efficient frontier a number of tests must be met. These are considered below. Once those tests have been met there remain issues surrounding the use that can be made of any frontier derived from the data. But it is important that we are clear about our vocabulary.

6 Until the tests have been met we are really talking about the company or companies that have the lowest reported costs (or subset of costs). This is not an 'efficiency frontier' and should not be referred to as such.

The comparability of data

7 The first test that must be satisfied relates to the comparability of the data. The CEPA report proceeded from the assumption that at DPCR3 Ofgem was successful in adjusting the reported costs of the distribution businesses of the public electricity suppliers (PESs) to ensure that all comparisons were carried out on a like-for-like basis. We did not believe at the time that the adjustments made by Ofgem captured all the differences in reporting of the PESs. However, without access to the submissions made by each PES to Ofgem and without access to the studies carried out by Parnell Kerr Forster on each PES (the PKF reports) we were unable to assist Ofgem in its efforts to cleanse the data to ensure that it was truly comparable.

8 Economists anxious to proceed to deploy their models may be prone to assuming that issues of data comparability can be resolved relatively easily and the CEPA report certainly gives little prominence to the problems associated with data cleansing. Yet it is of enormous significance.

9 At DPCR3 adjustments were made by Ofgem to the reported costs of the PESs in order to arrive at the data that formed the basis of the regression exercise. The total value of the adjustments made is set out below under each of the broad categories used by Ofgem:

Category of adjustment	**£m**
Capitalisation policy	80
Allocations, attributions and recharges	264
Recharges	23
Total	**367**

10 These adjustments represented 29 per cent of the total adjusted costs that were used (after some more adjustments totalling a further £276m) to conduct the regression analysis.

11 It is easy to see how any failure to capture all the adjustments might have a significant effect on the outcome. If we take the company that was used to set the frontier (Eastern Electricity), that company's reported operating costs were adjusted as follows:

Category of adjustment	**£m**
Capitalisation	2.2
Allocation, attributions and recharges	30.2
Recharges	0
Total	**32.4**

12 The total adjustments to Eastern's reported costs amounted to 27 per cent of its adjusted costs.

13 The important point here is the scale of the adjustments carried out by Ofgem in order to reach a starting point for regression analysis. The potential for error or incompleteness is large. Indeed, it has recently come to light that several companies (including one that was used to set the efficiency frontier at DPCR3) had been reporting some fault repair expenditure as a capital cost and these differences in accounting treatment went unobserved (and therefore uncorrected for) by Ofgem. As a result the regression

analysis was carried out on data that was manifestly non-comparable. Our purpose in making these observations is not to criticise Ofgem for the way that DPCR3 was conducted, rather to emphasise the importance and the difficulty of the task. In our view the challenge of ensuring data comparability at DPCR4 will not be any easier than it was at DPCR3 and the preparation of regulatory accounting guidelines (RAGs) has done little to ensure that the underlying data has been reported on a comparable basis.

14 The method used by Ofgem at DPCR3 offered some protection against errors arising from non-comparable data. In particular the decision to pivot the regression line and the assumption that companies move only three quarters of the way to the derived frontier mitigated the potential for error arising from non-comparability of data. If a strict COLS assessment had been used the potential for very significant error arising from this source would have been considerable.

Different operating conditions

15 Once Ofgem (and the DNOs) has confidence that the data set is comparable in terms of reported costs consideration has to be given to the different operating conditions faced by each of the DNOs.

16 A company is not 'less efficient' than another company merely because it has higher costs. The operating conditions of the companies need to be taken into account and adjustments have to be made to reflect these differences. To make proper adjustments for these differences is not a simple matter because it requires an understanding of the cost drivers in the distribution business and a knowledge of the operating conditions faced by each DNO. At DPCR3 some effort was made to recognise this with special adjustments for the extra costs of operating in London and in the scale variable which incorporated circuit length. However, these were relatively rough-and-ready adjustments which could not be said to have fully reflected the cost drivers or the diverse operating conditions of each DNO.

Differing quality of supply outputs

17 After adjustments have been made for differences in reporting and for differences in operating conditions account must be taken of the quality of the outputs delivered by each DNO. It is obvious that if there are two companies, facing the same operating conditions, one is not necessarily 'more efficient' than the other, if its lower costs are accompanied by a poorer quality of supply. It *ma*y be more efficient but this cannot be evident until account is taken of the quality of supply being delivered by each company.

18 Although this is obvious it is not easy to factor quality into the analysis. The fact is that the different companies have such different operating conditions and such different system legacies that a *high* cost - *low* quality company may be more efficient than a *low* cost - *high* quality company because the former company has made, on behalf of customers, an efficient trade off between improvements and customers' willingness to pay for those improvements. No account was taken of this dimension in the DPCR3 regressions – although some account was taken of quality of supply in the within range adjustments.

Differences in risk profiles

19 After taking account of differences in reporting of costs, operating conditions, and quality of supply outputs delivered it is necessary to take account of differences in the risk profile of each DNO.

20 Companies may have achieved lower costs by increasing productivity. However, some of the difference in costs may be attributable to the fact that one company may have been more adventurous in shedding costs and have taken on greater risk in so doing. It is important to appreciate that a company that makes this choice is not necessarily doing anything wrong; it is merely making a choice of where it wishes to be on the risk/reward spectrum. But before it is safe to conclude that a company with higher costs is less efficient it is necessary to understand the different degrees of risk being taken by the companies. It is the responsibility of the regulator to ensure that this is fully understood - since no company is well-placed to comment on the risk profile of another company - otherwise the regulator may, inadvertently, encourage companies to seek to emulate a

company that has taken risks that customers would prefer were not being taken on their behalf.

21 At DPCR3 we understand that this question was addressed only with respect to capital expenditure. Ofgem seems to have assumed that operating cost savings could be made by all the non-frontier companies to bring them three quarters of the way to the cost level implied by Eastern Electricity. We have no knowledge of any steps taken by Ofgem to ensure that Eastern Electricity had behaved responsibly with respect to its costs reductions and we note that, since DPCR3, questions have been raised about its avoided costs particularly in relation to vegetation management and its ability to restore supplies following the October 2002 storms.

22 The issue of different risk profiles is not addressed by the fact that all companies face the same disincentive to take undue risk. Companies may respond differently to their (different) perception of this disincentive. A frontier company's perception may discount the risk of any given cost reduction more than it may be discounted by a non-frontier company. Moreover, Ofgem sets a cost of capital appropriate to the sector as a whole. This may not align with the higher risk being taken by the frontier company.

Preconditions for the responsible use of benchmarking – summary

23 The brief analysis set out above demonstrates that only after adjustments are made for:

- differences in the reporting of costs;
- differences in operating conditions faced by each DNO;
- differences in measurable quality outputs delivered by each DNO; and
- differences in risk borne by each DNO as a result of choices made about cost reduction

can a regulator (or a regulatee) be confident that a fair and responsible comparison is being made. Therefore, we invite Ofgem to set out in the second consultation) due in December) how it plans to assess and make appropriate adjustments for each of these items before it proceeds to use benchmarking to inform the review.

24 If Ofgem does not intend, or simply does not think it is feasible, to carry out these assessments then benchmarking may still play *some* part in the DPCR4 process, but it cannot safely be used to set an 'efficiency frontier' which would become the basis of assumed efficiency targets for the DNOs that do not comprise the frontier.

BENCHMARKS - THE COMPETITIVE MARKETS STANDARD AND THE FRONTIER STANDARD

Appropriate benchmarking standards

25 One of the most fundamental issues raised in the CEPA study is the appropriate standard for benchmarking work. We consider here two alternative standards: the competitive market standard and the frontier standard that is recommended by CEPA.

Competitive market standard

26 Under a competitive market standard, the focus of benchmarking is the typical performance of firms in a competitive industry. The intuition for this approach is quite appealing. After all, economic regulation is generally considered to be warranted in cases where competition is absent and monopolistic practices can lead to higher prices (and/or lower production). It makes sense for utility regulation to have as its goal outcomes that are not inconsistent with what we know about outcomes in competitive markets.

27 In a competitive market, the industry as a whole earns a competitive rate of return and prices reflect the interaction of supply and demand conditions at the industry level. Thus, individual firms cannot influence the prices at which they sell the products or services they produce. They have, however, incentives to contain their costs, given prevailing product or service prices, in order to maximize their profits.

28 In the long run, firms that fail to contain their costs will be forced out of the industry. New technologies and management techniques diffuse, permitting all companies to

improve their efficiency. These attributes of competitive markets encourage the view that all firms in competitive markets are efficient.

29 The reality, however, is that at any point in time the extent to which firms succeed in containing their costs can differ greatly in a competitive industry. The more efficient firms earn superior rates of return whereas the less efficient firms earn inferior returns. It is firms of *typical* efficiency under competitive market pressures that earn what is commonly called a competitive rate of return.

30 In considering the functioning of competitive markets it is also worthwhile to draw a distinction between minimum costs in the *short run* and *sustainable* minimum costs. Firms can, in the short run, realise costs that are below what is sustainable in the long run. In the context of electricity distribution, a good example is the deferral of tree trimming and other maintenance expenses. In the long run, distributors that operate at less than sustainable cost will experience service quality deterioration. In the short run this will not be evident from quality of supply data.

31 How can the competitive market standard guide the use of benchmarking in regulation? We start with the assumption that some form of benchmarking method is employed that calculates the difference between a company's operating efficiency and the efficiency frontier. We can then use these frontier benchmarking results to ascertain whether the operating efficiency of a utility bears the same relationship to the cost frontier *as would obtain for a typical utility in a competitive market*.

32 This exercise must be done with some care. One complication that has already been noted is that average performers in competitive industries operate at some distance from the cost frontier. Another is that frontier benchmarking methods are able to identify only the short run non-sustainable (SRNS) cost frontier. A third problem is that even the SRNS frontier is difficult to estimate accurately for the reasons set out above. Problems include the mismeasurement of cost, mismeasurement of output and input quantities, and the exclusion of relevant cost drivers from the benchmarking exercise. The extent of these problems vary with the benchmarking method chosen. For example,

some methods are more prone to measurement errors and hence tend to exaggerate the distance of companies from the SRNS frontier.

33 Our discussion leads to the conclusion that average cost performers in competitive industries operate at some distance from the estimated SRNS frontier identified by benchmarking for that industry. Furthermore, this distance will be specific to the benchmarking method employed. The typical distance will be greater for methods that tend to exaggerate the distance from the frontier.

34 A review of benchmarking studies in competitive industries can shed light on the typical distance from the SRNS frontier. We asked Pacific Economics Group (PEG), a company with considerable US and international benchmarking experience, to conduct surveys of frontier benchmarking studies in two competitive sectors: banking and farming. The surveys reveal average efficiency levels of firms in these two competitive sectors that are only 80-90 per cent of maximum efficiency.

35 PEG's survey on banking efficiency using frontier methods covers Greek, Turkish, European and US banks. The studies for European banks report average efficiency levels from 70-85 per cent using parametric methods. The average efficiency level among US banks ranged from 80-90 per cent using this same approach. In addition, non-parametric approaches show average efficiency levels to be even lower in this competitive industry. The efficiency studies in the farming sector which PEG examined consider only technical efficiency and therefore do not consider all possible sources of efficiency. PEG found average technical efficiency performance in the range of 80-90 per cent in developed countries. The tables below summarise the findings from the two surveys.

Survey of Efficiency Studies of Banking Firms

Study	Data Coverage	Method	Result
Bauer, Berger, Ferrier and Humphrey (1997)	US Banks 1977-1988	Parametric	Average cost efficiency = 83%
		Nonparametric	Average cost efficiency = 30%
Berger and Humphrey (1997)	Survey of 130 efficiency studies of financial institutions	Parametric	Average efficiency = 84%
		Nonparametric	Average efficiency = 72%
Berger and Mester (1997)	US Banks 1990 – 1995	Parametric	Average cost efficiency = 86.8%
		Parametric	Average economic efficiency = 86%
Casu and Girardone (2002)	European Banks 1993-1997	Nonparametric	Average technical efficiency = 65%
		Parametric	Average economic efficiency = 65%
Christopoulous and Tsionas (2001)	Greek Bank 1993-1998	Parametric	Range of economic efficiency = 60%-100%
Christopoulous, Lolos and Tsionas (2002)	Greek Banks 1993-1998	Parametric	Range of economic efficiency = 60%-100%
Clark and Siems (2002)	US Banks 1992-1997	Parametric Method 1	Average cost efficiency = 86%
		Parametric Method 2	Average cost efficiency = 74%
Eisenbeis, Ferrier and Kwan (1999)	US Banks 1986-1991	Parametric	Range of average efficiency level by size = 81%-92%
		Nonparametric	Range of average efficiency level by size = 60%-72%
Fethi, Jackson and Weyman-Jones (2002)	Turkish Banks 1992-1999	Nonparametric (one variant)	Average technical efficiency = 57%
Vennet (2000)	European Banks 1995-1996	Parametric	Average cost efficiency = 80%

Survey of Efficiency Studies of Farming Firms

Study	Data Coverage	Method	Result
Brummer, Glauben and Thijssen (2002)	German, Dutch and Polish Diary Farms 1991-1994	Parametric	Range of average technical efficiency by country = 76%-95%
Hadri, Guermat and Whittaker (2003)	English Cereal Farms 1982-1987	Parametric	Average technical efficiency = 86%
Kumbhakar (2001)	Norwegian Salmon Farms 1988-1992	Parametric	Range of average technical efficiency by specification = 79%-83%
Kumbhakar, Ghosh and McGuckin (1991)	US Diary Farms 1985	Parametric	Range of technical efficiency by size = 66.8%-77.4%
			Range of average allocative efficiency by size = 84.6%-87.6%

36 It is clear from these surveys that the average efficiency level of firms is not at, or even close, to the estimated SRNS frontier in either of these two competitive sectors. Taking these results as representative of competitive industries as a whole, we may conclude that, to be consistent with outcomes in competitive markets, the relevant cost performance standard is one that is at some substantial distance from the production frontier. This distance will, furthermore, be greater for methods that tend to find large efficiency differences.

37 We started this discussion by assuming the use of techniques for comparing the efficiency of companies to a *frontier* standard. It can now be noted that benchmarking methods that can compare the efficiency of utilities to the *norm* for the industry can also be used to implement the competitive market standard. The basic idea here is that utility costs must be found to be at or *superior* to the utility industry norm to be fully acceptable.

38 This approach has a number of practical advantages. Most notably, it side-steps the difficult issue of the appropriate distance from the SRNS efficiency frontier. One disadvantage of the approach is that it is inconsistent with the competitive market paradigm when applied to a group of utilities that operate under the cost pressures typical of competitive markets. Since RPI-X incentive regulation has been preferred because it approximates to the cost pressures typical of competitive markets, clearly some care is needed with the use of yardsticks based on industry norms.

39 Although there has never been a mechanistic application of the average cost standard, there are many examples of its use in regulation in North America. Examples of rate cases where it has been a factor considered in setting prices include AmerenUE, Boston Gas, Enbridge Gas Distribution, Kentucky Utilities, Louisville Gas & Electric, Oklahoma Gas and Electric, San Diego Gas & Electric, Southern California Edison, and Southern California Gas.

Frontier standard

40 The CEPA report appears to advocate the use of a frontier benchmarking standard. Under this standard, companies are judged not for their success at attaining a reasonable

distance from the cost frontier but, rather, for their success at attaining the frontier itself. Companies found to have inferior cost performances may expect to be subject to revenue disallowances.

41 This paradigm can be criticised on a number of grounds. One is that the frontier in question is the SRNS frontier and not the long run sustainable cost frontier. As we have seen, it is unreasonable to expect utilities to operate for a sustained period on this frontier. If anything, companies found to lie on the SRNS frontier may need additional revenues to ensure the sustainability of the service they are required to provide.

42 A second concern about the non-sustainable frontier is its fairness. As we have seen, superior cost performers in competitive industries are entitled to superior returns. If firms must operate on the frontier to earn a competitive return, the regulator is essentially acting as a monopsonist on behalf of customers. Monopsony behaviour is not generally considered to be fair or desirable in advanced industrial societies. For example, the ability of labour unions to offset the potential monopsony power of employers is one of the major arguments for their activities being legitimate.

43 It is inappropriate for regulation to be based on methods that are analogous to the assumption by the regulator of the power of a monopsonist purchaser.

PART TWO - SPECIFIC QUESTIONS RAISED BY OFGEM

44 Having set out our views on the fundamental issues surrounding the responsible use of benchmarking we now set out our views on those items singled out by Ofgem as key issues for consideration.

INPUT DATA

What costs should be benchmarked? e.g. controllable operating costs, total controllable costs, capital expenditure etc

45 The dependent variable is key in cost assessment. OFGEM used standardised controllable operating costs as the benchmark variable in the 1999 price control review.

In a capital-intensive industry, the assessment of *total* cost, which includes both operating costs (opex), current capital expenditures (capex) and sunk capital costs, provides a more accurate picture of efficiency based on long-term optimising strategies employed by the firms in the industry. For instance, firms that are increasing their capital spending to replace their existing capital stock can thereby limit their opex, appearing more efficient when opex alone is the focus of benchmarking. On the other hand, those that undertake more operating and maintenance activities and thereby postpone the replacement of ageing plant can appear to be poor performers. Therefore, it is important to consider opex and capex substitutions when evaluating firms' cost performance.

46 One means of dealing with this problem is to base performance assessment on total costs (totex), which allows the generation of accurate efficiency measures based on long-term considerations. Alternatively, one can examine short-term efficiency by looking at approaches that take into account the impact of capital. One such approach is a short-run cost function, where an appropriate measure of the size of the capital stock can capture the influence of this important cost driver on efficiency.

47 However, it is important that a totex approach is not used in such a way as to undermine the probability of cost recovery for capital invested in the regulatory asset value (RAV). It should be used to inform the debate on operating cost efficiency rather than to determine mechanistically an outcome that might have the effect of disallowing investment in the RAV.

48 Given our reservations about the likelihood that the preconditions for the responsible use of benchmarking will be met at DPCR4 we are reluctant to propose that a further category of costs (i.e. capital investment) should be made subject to any benchmarking exercise. However, as CEPA observes, econometric analysis based on totex assessments reveals companies to be closer to one another in their efficiency scores than when econometric exercises are carried out using opex alone. We believe that this is because the use of totex in the exercise probably corrects for some of the cost categorisation anomalies that Ofgem was unable to adjust for at DPCR3 and that are still

inherent in the data set as we approach DPCR4. We therefore conclude that totex is safer and more likely to yield plausible results that can inform the price control review.

How should controllable costs be defined for the purpose of benchmarking?

49 The approach taken by Ofgem at DPCR3 captures those operating costs that are controllable. Our views on the capital element for any totex assessments are set out below.

What adjustments are required to enable comparisons between the DNOs?

50 In paragraphs 4 to 24 above we set out our observations on the importance of adjusting costs to reflect:

- differences in the way that companies report costs;
- differences in the operating conditions faced by the different DNOs;
- differences in quality of supply outputs delivered by the DNOs; and
- differences in the risk profile of different DNOs.

How should measures of total cost be calculated?

51 Extensive work has been done in the United States on the measurement of total cost for purposes of empirical cost research. Our advisers, PEG, have told us that the basic idea is usually to decompose capital cost in a given period into a capital service price index and a capital quantity index. The same approach could be adopted for use at DPCR4 and we would be happy to provide further analysis from the work carried out by PEG on our behalf.

What adjustments are required for firm specific factors?

52 Special adjustments will need to be made for any cost drivers that affect individual firms that are not properly captured in the scale parameter. The answer to this question therefore depends on the chosen scale variable.

Should international data be used? If so from what sources?

53 In principle we believe that benchmarking could be improved by the use of international data. However, we do not believe that Ofgem has the resources or the time at DPCR4 to cleanse the data that relates to companies operating outside Great Britain to ensure that it is comparable with the data set relating to the DNOs. Indeed, Ofgem's challenge in respect of ensuring comparability between DNOs (over whom Ofgem enjoys a legal power to require information) is a formidable one. We believe that Ofgem's efforts would be better directed at ensuring comparability between DNOs than in striving to achieve a comparable international data set.

Should panel data be used?

54 Given the small size and limited variation in a one year British sample it is difficult to develop benchmarking methods that are at once adequately attentive to the special operating conditions facing the various companies and also permit credible hypothesis testing. Adding additional years of British data to create a panel data set is the best single way to solve these problems. We believe that Stochastic Fontier Analysis (SFA) could yield credible results using even one or two years of British data. Additionally, panel data would provide stronger support for controls for additional business conditions in a Date Envelopment Analysis (DEA) context.

BENCHMARKING TECHNIQUES AND METHODOLOGY

Which techniques should Ofgem use?

55 The CEPA report considers programming techniques (DEA and index approaches including total factor productivity (TFP)), econometric techniques (Ordinary Least Squares (OLS), Corrected Ordinary Least Squares (COLS) and SFA, and process approaches (engineering economic analysis).

56 CEPA concludes (on pages 95 and 96):

- 'The DEA approach is theoretically more appealing than COLS as it determines efficiency using different input and output variables';
- 'one possible approach to establishing an appropriate efficiency frontier would be to use a combination of DEA and COLS'; and
- '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'.

57 We offer our views below on the techniques considered in the CEPA report.

DEA

58 Section 2 of the CEPA report paints a relatively favourable picture of DEA, particularly where a relatively small data set is available. However, CEPA note that in the Netherlands the regulator (DTe) revised its efficiency target for NuonNet from 8 per cent to 2 per cent per annum following an appeal.

59 NERA acted for NuonNet in identifying specific problems with the original DTe DEA analysis. Table 1 shows the scores calculated by the regulator before and after their analysis was challenged.[1]

DEA scores before and after challenge by NuonNet

Company	DEA score before	DEA score after
Defland/Edelnet	0.55	0.85
NUON	0.65	0.95
Midden-Holland	0.50	0.91
Eindhoven/ENET	0.59	1.00

Source: NERA (2002). Figure 3.

60 CEPA concludes that experience in the Netherlands highlights the risk of relying on a single technique for determining the efficiency frontier. A more reasonable conclusion appears to be that the application of DEA yielded little if any meaningful information in relation to setting the X-factor for NuonNet.

61 NERA (2002) set out specific problems with DEA which are reproduced in Box 1. These may usefully be read alongside the problems identified by CEPA on page 20 of its report.

[1] Dr. Enese Lieb-Dóczy and Graham Shuttleworth. May 2002. 'The sense and nonsense of benchmarking'. http://www.nera.com/wwt/publications/5242.pdf

Box 1: NERA (2002) comments on application of DEA to NuonNet

'Advocates of the DEA method point out that (unlike a regression), with a DEA approach no relationship between the input- and output variables needs to be defined and DEA can be used well in the case of statistical problems such as colinearity. Regressions also require a relatively large number of data points. DEA enjoys a certain popularity among regulatory authorities, but brings with it serious methodological problems.

- DEA does not offer a robust approach in determining which variables are relevant and which are not. Deviations from the upper efficiency limit can therefore be the result of errors in the model specification, not inefficiency.
- If only a small number of companies are considered from several aspects (output variables), a company will attain the upper efficiency limit if it scores particularly well in one aspect, irrespective of what values it achieves for the other aspects.
- It is an anomaly that the addition of two particularly efficient companies automatically reduces their DEA score. The result of this is that large companies or merged companies automatically come off worse than small companies or departments within companies.
- DEA results cannot be tested for their robustness. In contrast to a regression, DEA does not offer a statistical approach allowing measuring errors to be taken into account and the significance of differences in the values for two companies to be tested.

These fundamental methodological problems mean that the results of benchmarking on the basis of DEA can only be used with extreme care. The biggest problem is that DEA does not offer a robust methodology for testing for errors in the model specification and for testing the significance of differences between companies.

It can be concluded that simple comparisons can supply valuable information for internal purposes, but in general are only of limited meaningfulness and robustness for regulatory purposes and when it comes to questions of competition policy.'

62 One over-arching problem that emerges is that the small sample attraction of DEA is undermined by the fact that there is no way of knowing with (statistical) confidence what should be included in the DEA analysis unless separate OLS analysis is performed, requiring a larger (and potentially unavailable) sample.

COLS

63 Corrected ordinary least squares (COLS) refers to an approach where the mean regression line is estimated and then shifted vertically to coincide with one or more companies that are judged to form the 'efficiency frontier'. This places considerable weight on a few observations and it therefore no longer makes use of all of the data. Confidence intervals and test statistics calculated at the mean OLS estimate would not apply to the frontier identified by COLS. It is therefore impossible to identify the robustness of the estimated 'frontier'.

64 Commenting on the trade-off between the closeness of the frontier to the edge of the distribution and its robustness Yatchew and Adonis (2001) noted that:

> 'Furthermore, "best practices" are much more difficult to estimate accurately than, say, average or median performance. Indeed, the rationale for incentive-based regulation is grounded in the idea that the regulator *cannot* estimate minimum costs especially accurately.'[2]

TFP

65 CEPA considered both partial and total factor productivity measures and both the possibility of comparing firms at a specific date and also comparing performance over time.

[2] Yatchew and Adonis. 2001. "Incentive Regulation of Distributing Utilities Using Yardstick Competition", *The Electricity Journal*, January- February, p.57.

66 In relation to partial productivity measures CEPA note that:

> 'one cannot sum up the efficiency savings that these measures give for each function and suggest that the total efficiency saving is achievable for the company as a whole. This is to neglect the fact that companies may choose to substitute one type of expenditure for another hence giving them best performance on some measures but not on others leaving best performance on all measures simultaneously unachievable.' (p 25).

67 This point is general and applies to any technique applied to a sub-set of total costs, or potentially to total costs not allowing for quality differences.

68 In relation to TFP CEPA noted that:

- 'As a long term objective tying X to the TFP growth rate in the distribution sector might be a desirable goal.' (p 27); but that

- 'the UK has posted another impressive improvement in efficiency of its electricity distribution sector with high dispersion of productivity growth rates around a high average. This implies that it may be too early to consider TFP in calculating the X factor.' (p 29).

69 We can see the attraction of using TFP rates to assist in price control reviews. However, a robust TFP based assessment requires that the input data is sound and account has to be taken of the fact that the retail price index element of RPI-X regulation already captures economy-wide productivity improvements. The X factor should reflect only those improvements that may be expected above those achieved in the economy as a whole. However, as we approach the fourth price control review since privatisation it is appropriate to ask why it is considered necessary to try to estimate the achievements that may be made *in the future* by DNOs. There is a general consensus that those efficiency gains that were easiest to achieve have now been secured. CEPA's Dr Pollit indicated at the Ofgem workshop that there were two reasons why regulators might try to anticipate efficiency gains at a price control review. The first was to ensure that the future of

incentive regulation was politically sustainable: capturing only the known gains from the prior period left a potential political problem if incentive regulation was perceived by the public to give companies too easy a challenge. The second reason advanced by Dr Pollit was that there was evidence that setting companies harsher Po/X factors correlated with greater subsequent cost reduction.

70 With respect to the first observation, we believe that there is little prospect of significant, let along excessive, out-performance after three price control reviews and after we have already reduced operating costs by 60 per cent since privatisation.

71 With respect to the second observation the only reason why it might be justifiable to set tougher Po/X factors to encourage greater efficiency would be if the capital markets were incapable of driving out efficiencies. The *level* of the Po/X factor does not affect its incentive properties. Of course, setting Po/X factors for all companies by reference to the cost of 'frontier' companies may force companies to adopt a different and higher risk profile. It would, however, be a mistake to direct regulatory policy towards that end without full consideration of whether customers desire such an outcome.

72 We are therefore not convinced of the need to anticipate gains at a price control review.

SFA

73 We note that the CEPA report states that inefficiency scores under SFA are 'biased downwords' (p36). This statement reflects the fact that SFA apportions part of unexplained performance differences to random events. However, it is quite wrong to conclude that SFA understates the inefficiency of the sample Rather the other methods, COLS and DEA, are likely to over-state the inefficiencies by failing to take account of random events in cost appraisal and hence in efficiency measurement.

74 SFA allows for hypotheses tests of efficiency estimates in order to determine if measured inefficiency is statistically significantly different from frontier company efficiency. Since the consequences of relying on a statistically unsound method could have very serious consequences in the electricity distribution sector, a conclusion that a company is inefficient should be reached only if the proposition that the company is at

the efficiency frontier can be falsified. SFA allows such hypotheses testing and therefore can contribute to responsible benchmarking as part of a price control review.

75 In our view CEPA has underestimated the potential to use SFA at DPCR4.

Enhanced DEA

76 If DEA is to become the choice of benchmarking, as recommended by CEPA, it is critically important to form statistical tests to determine if the inefficiency scores generated by the DEA analysis are significantly different from an efficient score. DEA studies commonly yield efficiency score generated that are point estimates with no statistical variance reported. Therefore, the statistical significance of results is unknown. It is, however, possible to determine the statistical significance of scores generated by DEA. Such a determination requires that we first identify the source of variation in this measure. DEA scores are generated from a frontier or envelope calculated using the data. Therefore, the source of variation in DEA scores arises from variation in the calculated frontier, which is a representation of the 'true' frontier. The variability of the frontier can arise due to the randomness of the production technology and/or the randomness of the sample, which is drawn from an underlying population. In the context of the DNOs, if the full sample used in the analysis is the entire population, then variability in the frontier arises from randomness in production technology.

77 With this in mind, we can apply a method that computes the statistical variability of DEA scores called the 'bootstrap'.[3] The bootstrap computes statistical variability of DEA scores based on the assumption that the generated frontier is stochastic. As a result, it simulates the process that generates the observed data repeatedly, applies the estimator to each simulated data and forms confidence intervals for efficiency scores using these estimates. The simulation process is essentially a form of sampling with replacement such that different samples result at each simulation step. The key to the bootstrap procedure is the number of simulations undertaken, as this determines the

[3] See Simar, L and P W Wilson, 1998, 'Sensitivity of Efficiency Scores; How to Bootstrap in Non-parametric Frontier Models,' *Management Science*, 44(1): 49-61; Simar, L and P W Wilson, 2000a 'A General Methodology for Bootstrapping in Non-parametric Frontier Models,' *Journal of Applied Statistics*, 27(6): 779-802; and Simar, L and P W Wilson, 200b, 'Statistical Inference in Non-parametric Frontier Models; The State of the Art,' *Journal of Productivity Analysis*, 13: 49-78

level of statistical significance attached to the efficiency estimates. The greater the number of simulations, the wider the confidence interval and the higher the probability value or confidence level attached to an efficiency estimate. We recommend that such a procedure be used if DEA is going to be used to benchmark efficient costs.

78 The DEA method proposed by CEPA must also be enhanced to incorporate additional business conditions. It is important to try and account for extra business conditions that affect cost and thus efficiency scores. The CEPA report excludes all business conditions based on second stage regressions. However, this conclusion is based on a very small sample. It is important to expand the sample size if the technique is to be robustly applied.

Which costs drivers should be included and how should they be selected?

79 The selection of cost drivers is a critically important dimension of a fair benchmarking study. In this section, we first consider the output quantity specification. We then consider possible additional business conditions.

Output Quantity Variables

80 At DPCR3 OFGEM used a composite scale variable composed of the number of customers (N), volume delivered (V), and length of the distribution system (L). The weights were 50 per cent on customers, and 25 per cent each on volume and line length. The CEPA report questions the inclusion of all three variables in the composite on the grounds that N and V are highly correlated. Instead, it recommends that only V and L be included in a composite scale measure and that they be assigned 50/50 weights.

81 One concern about CEPA's work is the cumbersome and indirect way that CEPA estimates the appropriate output quantity measures and weights using DEA analysis. A more simple and straightforward approach would be to regress opex on alternative output quantity specifications and find out which had the highest explanatory power.

82 At a minimum this approach should be pursued to provide a consistency check with the CEPA methodology.

83 At our request PEG have undertaken preliminary regression work using the 1999 data provided by CEPA. This work does not support CEPA's output quantity recommendation and points to an alternative specification that includes the number of customers. PEG also find, as CEPA does, that results are sensitive to the inclusion or exclusion of data for Eastern. Using either methodology, it is clear that the selection of output quantity drivers would receive greater substantiation if based on multiple years of British data.

84 PEG's work using U.S. data provides additional support for a specification that includes both the number of customers and the delivery volumes of distributors as statistically significant cost drivers. PEG's study shows that the number of customers accounted for nearly 70 per cent of the impact of the scale measures used in the exercise. Preliminary econometric research performed by PEG points to the conclusion that the number of customers is the dominant output-related cost driver *even if customer-related expenses are excluded*. We are therefore sceptical about CEPA's conclusion that 'with the move towards competition in metering, customer numbers are no longer as important a factor in determining the costs of distribution companies.'

85 CEPA can also be criticised for its reasoning that 'the three scale variables in the composite are highly correlated, particularly customer numbers and units distributed. It would therefore appear unnecessary to include all three variables in the composite.' (p74).

86 While correlation is an unfortunate fact of life in this case, it is hardly a reason for the omission of a variable that has been found to be a significant cost driver when doing so can have a material impact on efficiency rankings.

Assessing the composite scale variable and cost drivers

87 It is not possible to separate consideration of the composite variable from consideration of what the actual cost drivers are. The approach CEPA adopts in evaluating alternative cost drivers is to see if they are correlated with the efficiency scores resulting from the 1999 methodology. This method is flawed.

88 Unless the initial regression contains all the significant cost drivers the resulting 'efficiency' scores will be biased if any of the omitted variables are correlated with the included variables. In addition the test statistics for the estimated equation will be invalid.[4] The reason for this is that if the included variables are correlated with omitted variables that are cost drivers the included variables will tend to proxy the omitted variables – giving a false indication of their importance and biasing the resulting efficiency scores.

89 We know from Figure 36 on page 81 of the CEPA report that some of the omitted variables are indeed correlated with the variables included in the composite variable. The results of the base regression may therefore be biased and the regression of the additional variables on the efficiency scores is not informative. It is not therefore valid to conclude 'that utilising a single composite cost driver measuring differences in scale should be sufficient for Ofgem's purposes' (p 83) – assuming the purpose is to find a valid measure of 'efficiency'.

Other Business Conditions

90 The CEPA conclusions would unfairly affect companies whose costs are affected by:

- extensive underground networks; and
- high service quality.

91 The statistical tests performed by CEPA do not provide a satisfactory basis for deciding to ignore special operating conditions. In addition to the awkward multi-step nature of

CEPA's statistical work, the sample has inadequate size and variety to generate adequate results. Note also that CEPA employs a very high (95 per cent) confidence level to render its judgements. A confidence level of 90 per cent is more appropriate given the sample size. At a 90 per cent confidence level an undergrounding variable would be significant using CEPA's methods.

What assumptions should be made if using a regression e.g. functional form, the intercept (fixed costs)

92 We offer no comments on the appropriate functional form for regression techniques. However, we think it is important to understand that the intercept on the Y axis does not correspond to the fixed costs of running a distribution business. It is merely the constant term that results from the slope of a line that, imperfectly, best reflects the costs regressed against the chosen scale variable. This intercept will therefore include the sum of unexplained variables and it should not be confused with the fixed costs.

If using DEA what combination of inputs and outputs should be used? Should the models be input or output orientated or both?

93 Given the reservations we have expressed above on DEA we offer no comments on the preferred combination of inputs and outputs.

What assumptions should be made about returns to scale and economies of scale?

94 The basic choices here are twofold. One option is for cost benchmarks to factor in the scale economies that are typically realised at a particular output level. The other option is not to factor this in so that larger firms get credit for the larger scale economies that they realise and smaller firms are penalised for having realised relatively small economies.

[4] This is a standard result set out in introductory econometrics textbooks. For example, Dougherty, *Introduction to Econometrics* (1992) Section 6.2.

95 Whichever approach is taken, care must be taken to ensure that companies that have merged under the Ofgem merger policy that prevailed at that time will see the benefits of their merger savings for a full five years.

USE OF BENCHMARKING IN THE FINAL COST ASSESSMENT

What is the appropriate benchmark for the DNOs, the frontier firm, the average firm? or something else?

96 The use of average costs in benchmarking is less risky than the use of frontier companies to inform judgements about efficient levels of costs. Given the asymmetry of risks and the value placed by customers on a secure electricity supply it is important to err on the side of caution. Nevertheless, it must be appreciated that even average cost approaches are not without risk and do not equate to the competitive market standard. Care must be taken to ensure that the data is truly comparable and the model specified as accurately and completely as possible. Given the difficulties surrounding this, it will be necessary to apply judgement to the results. In practice this will mean making assumptions, as Ofgem did at DPCR3, about the distance that a company may plausibly move towards any frontier (whether based on average or frontier costs).

How should benchmarking be combined with other analysis particularly the bottom up modelling and TFP analysis?

97 TFP analysis must be used with considerable caution because the significant rate of productivity gains achieved during the period since privatisation will not be an indication of the gains that may be achieved in the forthcoming period.

98 Furthermore, we see no compelling reason why a price control review needs to *anticipate* future efficiencies.

OTHER ISSUES

How should merged firms be treated for the purpose of benchmarking? Should DNO groups be benchmarked as well as the 14 DNOs?

99 It is reasonable that Ofgem's analysis should be informed by both a DNO group and a licensee based approach.

100 However, care will have to be taken to ensure that companies enjoy the full benefit of their merger savings for a period of five years from the date of the merger.

Should measures of quality or other outputs be incorporated into the benchmarking process?

101 CEPA considers benchmarking that includes a combination of costs and quality. CEPA propose minutes lost per customer as the quality measure. Minutes lost varied from 42 for London to 143 for Scottish Hydro-Electric (Annex 8).

102 CEPA found that including quality and opex in the output measure improved the DEA efficiency scores of every firm (p 89). In particular, those firms that looked least efficient in terms of opex alone moved significantly closer to the 'efficiency frontier'.

103 To assess this result CEPA applied the same flawed technique that it applied in Section 7 of its report, namely to regress the efficiency scores on minutes lost per customer. CEPA concludes that 'there does not … appear to be a case for including quality explicitly in the benchmark' (p 90). If minutes lost per customer were an explanatory factor and were correlated with elements of the composite variable (for example, network length) then the analysis and conclusion would be flawed.

104 For quality, however, the problem of inference runs deeper than the above technical point. For an individual firm higher quality clearly implies higher costs. However, across all firms higher cost need not be correlated with higher quality.

105 The reason for this is that at a socially efficient outcome each firm would provide that level of quality that equated the marginal costs of quality with the benefits in terms of customer willingness to pay for quality at the margin. It would therefore be efficient for society if firms with high costs of provision of quality, say those with long lines, had lower quality of supply. Across the sample as a whole this will tend to confound any simple relationship between quality and cost – even though the level of quality for each firm may be an important cost driver. A null result from even a well specified regression including quality would not therefore imply that quality was unimportant as a cost driver.