Benchmarking and Optimal Scale for Electricity Distribution Boards in New Zealand
A study for the Commerce Commission

Michael G. Pollitt
Energy Policy Research Group
Judge Business School
University of Cambridge

23 July 2025

The Brief from the Commerce Commission:
'The scope of the study will encompass the advice on the following two questions:

1. ***Benchmarking EDB Efficiency:***
   How should we compare/assess the EDBs in terms of efficiency? This is the benchmarking question – and we would like to identify how we can best create incentives to reduce costs over time (irrespective of the number of EDBs).
   - *Our aim is to support the development of effective regulatory incentives for cost reduction over time.*
   - *We are interested in methods that are more likely to be robust and applicable in our regulatory context.*
2. ***Assessing Optimal Scale:***
   How should we assess/determine the likely most appropriate size for EDBs? (or equivalently how many should NZ have). We are concerned that the current structure (29 EDBs) may duplicate costs and reduce technical efficiency, particularly for smaller firms.'

In what follows I answer both questions in turn.

Q1: How should we compare/assess the EDBs in terms of efficiency?

Benchmarking and Incentive Regulation

Most incentive based regulatory systems for distribution networks rely on CPI-X regulation as popularised by Stephen Littlechild (1983) in the UK and theoretically developed by Shleifer (1985). It is important to point out that the key aspect of this is the fixing of the network revenue formula ex ante. This creates an incentive to cut costs and improve service quality by allowing the regulated firm to keep profits for a period of 3-5 year (the time frame of a price control review). This has since been developed further by Ofgem, the energy regulator for Great Britain, into RIIO (Revenue = Incentives+Innovation+Outputs).[1] This emphasises the need for positive revenue incentives to support desired outputs (e.g. speed of connections) and innovation (e.g. via R+D projects).

It is important to acknowledge that a price control review can be viewed as a negotiation between the regulator and the regulated firm, where the firm is challenged to justify its stated revenue requirements following an initial suggestion from the regulator of what the allowed revenue should be (Beesley and Littlechild, 1989). Benchmarking can inform the initial negotiating position of the regulator. Regulatory benchmarking can be subsequently challenged by the regulated companies during the price control review process or subsequently by higher order regulators or the courts as part of the appeals against the price control review process.[2]

In this context efficiency benchmarking usually plays two key roles. First, benchmarking is about establishing starting points for target revenue at the beginning of a regulatory control period. Individual X factors can adjust for the different levels of potential for cost reduction based on how a regulated entity compares to an efficient version of itself (see Figure A.1). Second, dynamic benchmarking over several years can establish the trend level of productivity growth for a sector and set a general target level of efficiency improvement that can be pursued by all regulated entities, including the most relatively efficient firms.

Another potential role of comparative benchmarking is to create league tables of relative performance, which are then not used directly in setting regulated revenues[3]. League tables are potentially very important in competitive customer oriented businesses and have been strongly advocated for in electricity retailing (Littlechild, 2021). However the way league tables work is clear: league tables directly influence profitability via influencing switching rates.

[1] Ofgem (2010).

[2] In Great Britain, the Competition and Markets Authority (CMA) acts as the appeal body against the decisions of sector regulators on price controls (previously the Competition Commission, CC). The process for appeal is outlined in CMA (2022). For instance, Mid Kent Water appealed against price control on basis of non-use of panel data. CC (2000) ruled is was difficult to decide between efficiency models.

[3] The Commerce Commission is currently prohibited from using comparative benchmarking within its regulatory price setting process for EDBs (see New Zealand Commerce Act 1986, 53P).

Using comparative benchmarking to construct a league table which is not backed up with regulatory financial incentives is not normal practice for regulators.[4] This is because it would be costly to undertake and it would be unclear what *additional* incentive properties this would create – since incentive regulation emphasises the importance of actual financial incentives (as exist with electricity retailing league tables). Regulated companies can and do undertake their own joint benchmarking exercises, where they can share information confidentially, without having to worry about regulatory oversight or the need to engage in regulatory gaming (making them more accurate and tailored to their commercial needs). Smaller companies might find a regulatory benchmarking attractive if it socialised the cost of undertaking an otherwise expensive benchmarking exercise. However, conversely, larger companies can undertake such exercises with a view to a potential takeover of smaller companies, so the overall likely benefit of such a centralised exercise would seem to be questionable. However full data transparency might allow more and better private benchmarking exercises to be undertaken and this would likely achieve the same overall effects at lower costs. Similarly, a one-off benchmarking exercise for the regulator might be useful in establishing whether there are currently wide divergences in performance and hence whether future use of comparative benchmarking might be useful.

There are clearly interesting issues worth exploring with league tables and how exactly these can effect behavioural change in the absence of regulatory financial incentives. These have been explored in the context of public services, like health and education. (see Foley and Goldstein, 2012, for a report on lessons from the UK public sector). In energy, promoting behavioural change with league tables has often focussed on issues like energy efficiency, where the desire is to promote simple changes to routine which can effect energy use rather than deep reductions to costs which require multi-year programmes of strategic re-orientation[5].

Efficiency analysis on a comparator group of firms using multi-year data can give both an estimate of the trend productivity improvement and the relative inefficiency of a given firm. While there are a number of techniques that can provide estimates for both of these, it is not actually essential to do efficiency benchmarking to provide both individualised and general incentives to improve performance.

Performance-based rate making, as practiced in the US, might simply set a real revenue cap per customer with a revenue sharing mechanism (see Jamasb and Pollitt, 2001). The revenue sharing mechanism could adjust revenues each subsequent year in line a

[4] A good example of where league tables are backed by financial incentives is Ofgem's Output Delivery Incentives where relative performance of electricity distribution network companies – on, say, customer satisfaction, is assessed by an independent panel. This is then rewarded with financial incentives (see Ofgem, 2024a). The extensive use of performance league tables on train punctuality by the Office of Rail and Road in Great Britain, is accompanied by direct passenger payments for under-performance. See for example: https://dataportal.orr.gov.uk/statistics/passenger-experience/delay-compensation-claims/

[5] Allcott (2011) provides analysis on the effect of providing household consumers with information on their energy use relative to their neighbours. The effect is small but significant. Appliance energy ratings and household energy performance metrics also exploit a desire to conform to 'good' behavioural norms, with good, but modest, effects.

target rate of return, whereby if the rate of return increased the company would have to return some portion to the customers in lower prices. This would act as a rolling incentive to reduce costs and would actually encourage a more inefficient firm (which would find it relatively easy to increase profits) to reduce costs by more. This mechanism requires no external benchmarking and was, for example, implemented for Southern California Edison by the California Public Utilities Commission for 5 years from 1996, with annual X factors of 0-1.6% and four bands of revenue sharing around a target rate of return. For +/-50 basis points of rate of return, shareholder receives 100% of reward/risk; +/-50 bps to +/- 300 bps sharing between customer and shareholder of 25-100%; +/- 300 bps to +/-600 bps, 100% to shareholder; and above +/-600 bps the 'rates' are reopened by the regulator (see Jamasb and Pollitt, 2001). Joskow and Schmalensee (2025) suggest that Hawaii provides a good recent example of such a Multi-Year Rate Plan (MYRP), which has a similar structure of a 5 year X factor with revenue sharing.[6]

Another simple mechanism that could be implemented as part of CPI-X regulation is an average productivity X factor. This was suggested by Uri (2002) for US telecoms. If the X factor is specific to the industry being regulated and the industry starts off at a normal level of profitability, then in theory setting the X factor equal to the average productivity level will maintain reduce prices in line with costs and maintain profitability. One problem with this is that initially low profitability firms might never be able to hit their target level of profitability, as it does guarantee any individual firm will hit a target level of profitability. However it has got strong incentive properties and can be set to run continuously. In New Zealand, setting a common X factor and adjusting on the basis of profitability (rather than comparative cost benchmarking) at the start of the regulatory period means relative inefficiency can persist (and grow) for long periods and is dynamically underwritten by upward revisions to prices to maintain profitability. Firms which are pursuing non-profit objectives (including a managerial quiet life) are underwritten by this system.

In practice many regulators have preferred to calculate individual firm efficiencies as part of their periodic reviews of revenue allowances for regulated distribution firms (see Jamasb and Pollitt, 2001; CEER, 2025). A good reason for doing this is that this is what all firms do in setting internal performance targets. It has also been recognised that for smaller regulated firms, the regulator is providing information on potential comparators and performance that they might not otherwise have access to. It is also true that to some extent regulatory benchmarking can endogenously improve performance, by setting a target X to aim for, particularly in negotiations with unions or suppliers. The higher the X, ceteris paribus, the likely more cost reduction that can be induced from third parties working with the regulated firm.

Comparative Benchmarking Options

A number of comparative benchmarking options can be pursued.

[6] The details of the plan are here: https://puc.hawaii.gov/wp-content/uploads/2020/12/PBR-Phase-2-DO-5-Page-Summary.Final_.12-22-2020.pdf

A standard management consultancy style of benchmarking is **activity benchmarking.**

This can be functional or task based. Functional is where IT departments or Head Quarters staff are compared directly across firms (who do not need to be in the same industry). Task based might look at the cost/time/labour to fix a fault in the network or install a new transformer. This sort of analysis can be undertaken by Engineering or Economic consultants on behalf of the regulator.

Matching analysis might compare key performance indicators (KPIs) across matched firms. This might look at the number of staff per customer for firms of the roughly the same size, with similar customer types (e.g. two utilities covering large metro areas). This is less popular among regulators who don't have national matched pairs, than it is among the companies themselves.

Model analysis uses an engineering model to build an efficient network to serve the actual customers of a given network. This 'ideal' model or 'norm' model has been used in certain countries (e.g. Brazil[7] and Sweden[8]).

Frontier efficiency analysis (see Coelli et al., 2005; O'Donnell, 2018) has become popular across the world (see for example Jamasb and Pollitt, 2001, and Haney and Pollitt, 2009, 2013a). Three techniques have been popular: data envelopment analysis (DEA), corrected ordinary least squares (COLS) and stochastic frontier analysis (SFA).[9] Historically we have seen a lot of use of DEA (though less than activity benchmarking); some COLS; and less SFA (see Table 1.1).

Table 1.1: Use of different benchmarking techniques across 43 electricity regulators (Haney and Pollitt, 2009).

D= Distribution T= Transmission	Rank	Frequency
Process/Activity	1	10 (D); 8 (T)
DEA	2	8 (D); 8 (T)
COLS	3	3 (D); 3 (T)
SFA	4	2 (D); 1 (T)

Overall, there is weak evidence of experience and neighbourhood effects in best practice in regulation (c.f. Haney and Pollitt, 2009, 2011). Thus similar groups of countries have a tendency to use similar methods in their benchmarking. We discuss why countries have particular methods over certain periods below.

[7] For instance in the Brazilian electricity regulator, ANEEL undertook a benchmarking exercise in 2004 for 61 electricity distribution utilities using an engineering model to assess what their costs should have been given their configuration (see, Silva, 2007).
[8] See Jamasb and Pollitt (2008) on Sweden.
[9] See Figure A.2 for a comparison of DEA and COLS; see Figure A.3 for a comparison of COLS and SFA.

Multi-factor Total Factor Productivity Analysis (MTFP) can also be used to calculate relative productivity and trend productivity growth. This is involves fixing the input and output weights to calculate weighted productivity score and annual rate of change of this score. Such an approach is similar in spirit to DEA with no allowance for stochastic factors, but with fixed weights on inputs and outputs. This is one of the methods used in Australia to cross check other methods (Frontier Economics, 2023).

For an up-to-date summary of regulatory approaches to benchmarking among EU-27 energy regulators, see CEER (2025). Frontier Economics (2023) discusses the details of recent benchmarking experiences in GB, Finland, Australia and Germany.

In addition to the above methods for looking at the efficiency of individual firms, we see other approaches being used to look at productivity trends in setting industry level X factors. For instance the regulated water industry in the UK has had its costs allocated across other industrial sectors to construct a benchmark real price effect index which is a weighted average of non-water sectors (see CEPA, 2024a). It is also possible to set the target cost level for each firm on the basis of the basis of average industry costs (this currently happens in the Netherlands). This is quite close to the yardstick scheme originally envisaged by Shleifer (1985) and involves no direct comparative efficiency calculations.

When resources are not a problem or comparators are difficult to find domestically non-frontier efficiency methods have a lot of value, because they can deal with small samples of comparable domestic regulated firms.

When comparators are large enough in number DEA, COLS or SFA are implementable. Individual regulators normally favour using domestic samples for these efficiency studies, though they periodically may look to international comparisons to check how far from the frontier their domestic firms are. This is because justifying setting price adjustments or X factors on the basis of international comparisons is difficult. A more reliable route to gaining best practice is allowing foreign takeovers or mergers. For instance, National Grid in the UK did takeover US distribution firms, while ENEL have taken over a number of distribution utilities around the world.[10]

Comparing efficiency benchmarking approaches

DEA has the advantage of allowing the data to directly determine the frontier of efficiency and of allowing comparators who are defining the relevant portion of the frontier to which an individual firm is being compared to be identified. Efficiency can be visualised and it is the method which is closest to what firms do when internally comparing their performance to their peers.[11] It does not require a functional form to be imposed on the production process and can incorporate variable returns to scale (or

[10] https://www.enel.com/company/about-us/where-we-are

[11] I was once involved in an exercise where we used DEA to identify one particular firm which could serve as a benchmark for one of the firms we were analysing (on behalf of the holding company CEO). Having initially claimed that there were no relevant peer firms, the relevant subsidiary manager admitted that they were already using the firm we had identified using DEA as a comparator.

impose constant returns to scale[12]). It is also easy to incorporate multiple outputs and inputs and to look at the evolution of productivity over time. Productivity can be further decomposed into frontier shift caused by the leading firms and catch up of an individual firm moving closer to the frontier (see Figure A.4). The advantage of a DEA productivity decomposition being that no assumptions are required about the pattern of productivity growth over time. Adjustments for environmental conditions can be incorporated directly into the DEA or adjusted for in secondary analysis. Frontier stripping where the entire group of frontier firms is removed and scores recalculated, can be practiced to remove outliers and test the stability of the frontier (see Figure A.5). DEA is essentially an extension of single factor productivity measures and simple comparisons of unit costs that might be done by companies, themselves. DEA can be criticised on the grounds that it might exaggerate efficiency scores or productivity growth by allowing frontiers to be defined by data which might reflect one-off reductions in costs or increases in output with no adjustment for the possibility of error. DEA has been adapted to incorporate

SFA takes explicit account stochastic measurement error in the data by assuming that this exists. A functional form can be defined (usually Cobb-Douglas or Translog) on the production function and productivity growth can be identified, though some assumptions are required about the nature of the trend (e.g. whether it is a single value over the period, or rising or falling). SFA does tend to higher efficiency scores as it is likely to attribute deviations from the frontier to measurement error rather than efficiency. Efficiency scores tend to be higher unadjusted than under un-corrected DEA. The more complex the estimated functional form of the cost function, the more likely high efficiency scores are to result. It is worth pointing out that parametric approaches inherently try to maximise fit and hence 'despise' error. SFA can identify time trends in panel data in order to identify productivity growth rates (e.g. Greene, 2005). Thus we can say that while DEA would likely exaggerate efficiency scores and frontier productivity growth, SFA would likely underestimate them.

COLS is another parametric approach which is practiced by regulators, which fits a functional form to the cost function but shifts it to incorporate all of the firms. It can be more or less generous in terms of inefficiency, depending the extent to which the shifted cost function envelops all or only, say, 75% of the less efficient data, leaving the 25% to be deemed efficient. This approach has been used extensively by the GB regulator, Ofgem. COLS tends to be used to calculate efficiency, rather than productivity.

The Finnish regulator has used the Stochastic Non-parametric Envelopment of Data (StoNED) approach (see Johnson and Kuosmanen, 2015). This is an approach which estimates a non-parametric fronter using minimal axiomatic assumptions about the functional form of the production function. It is an SFA type approach that attempts to combine the advantages of DEA (which does not impose a functional form) with the advantages of SFA (in allowing for measurement error).

[12] Constant returns to scale assumes that optimal subdivision is possible at higher than optimal scale.

All frontier efficiency measures can be adapted to incorporate multiple variables which are deemed to be cost drivers (these can be incorporated as inputs, outputs or environmental variables depending on the nature of how they affect costs). However, in general more variables means higher efficiency scores and also the risk of spurious correlation. As we discuss below which variables should be included is an important area of disagreement between regulators and their regulated companies (and among the companies).

Many regulators have used a variety of approaches in order to try and identify efficiency and productivity. These can include DEA, COLS and SFA or some combination of an efficiency approach and other benchmarking approaches as detailed in the previous section.

Lovell (2006) discusses best practices in efficiency analysis in theory and practice. He points out that other advanced variants of efficiency exist, such as Stochastic Data Envelopment Analysis (SDEA)[13] or versions of both DEA and SFA which incorporate environmental variables. Best practice in implementation involves: a large / high quality dataset; panel data; consistency with engineering / well behaved functional forms; use of bootstrapping / confidence interval analysis; checking for consistency with non-frontier methods; inclusion of quality / environmental / input price variables; and clear value added in doing efficiency analysis. However as Weyman-Jones (2006) observes about the use of efficiency analysis in practice: 'The overwhelming impression of regulatory and governance case studies is that sample size, variable choice, model specification and choice of methodology has been governed by different objectives from those in the theoretical literature.' (p.25). Thus actual benchmarking is a lot messier in practice than would be desirable in theory.

Regulatory choices in implementing a benchmarking method

The peer group

One key choice is which group of firms will be benchmarked against which. For a country with a large size variation of firms a decision might be taken to drop the smallest and largest firms from the sample. This is what would happen in most analyses of large samples of data, where the concern was to accurately estimate the best fitting statistical relationships, rather than to include all of the data points. Indeed it might be possible to partition the data into sub-groups and estimate DEA or SFA on each sub-group. Statistical techniques exist to allocate firms to groups (rather than arbitrarily using one of the parameters, such as number of customers). Thus, Latent Class analysis which sorts the sample into a number of groups should these be significantly different (see Llorca et al., 2014).

The outputs, input and environmental variables used

EDBs distribute electricity to a number of customers over a specific area at a given quality of service. They use physical assets, labour and materials do to this. Costs can be measured in real money or in physical units. Customer mixes, degree of

[13] See Figure A.6 which compares DEA and SDEA, based on Olesen and Petersen (2005).

undergrounding, peak demand, peak capacity, geography and weather may all be important drivers of cost. Efficiency analysis has to decide how to incorporate relevant variables and whether it is necessary to do so, because some variables may be highly correlated and hence do not all need to be included. It is also often preferable to work with a smaller number of variables on the initial analysis (because the larger the number of variables the larger the efficiency scores). More variables also increases the chance that estimated cost functions are badly behaved (costs fall in variables which are clearly cost increasing). It is then possible to test whether other variables are significant and make a second-stage adjustment or to exclude them entirely due to insignificance. Including more variables directly in the efficiency analysis biases measured efficiency up, due to effectively reducing the number of firms against which any given firm is being compared. While preferred measures of output tend to favour measures of outputs directly demanded by consumers (i.e. energy and connection), it is possible to use measures of capacity (e.g. total network length) as outputs, to reflect that the network is built ahead of need. One good regulatory strategy is to ask companies which variables they think are materially significant and then test whether there is evidence that this is the case.

Which costs are comparable

An important up-front choice is to what cost data exactly is benchmarking applied. Is it the entirety of the costs or some fraction of them. Is it physical inputs and outputs rather than financial measures of cost, to maintain comparability and avoid accounting problems? When it comes to benchmarking costs, it is usual to focus the benchmarking on operating costs (opex) rather than capital costs (capex), but it is possible to benchmark total costs (totex = opex + capex), or to even benchmark opex + rental cost of capital (which would be driven by the historic fixed assets).

If benchmarking is only applied to opex, this does raises issues of opex-capex trade-offs and still leaves the question of how to ensure capital investment is efficient open. GB has addressed this with menu regulation of capex expenditure.

Cost adjustments to improve comparability

There is also the issue of the extent to which pass-through or firm specific costs are removed or adjusted for in the data, pre-analysis. Ofgem do multiple adjustments to their data in order to make the data comparable before doing the benchmarking analysis, and then have to apply measured efficiency score to that portion of the data which was actually in the efficiency analysis.

In its review for the 2005-10 electricity distribution control review Ofgem had 14 regulated entities (7 independent), used COLS (1 output/1 input), did not use panel data or confidence interval analysis. It did cross check with other methods (DEA,SFA), adjusted the COLS to the upper quartile of regulated firms and made 13 adjustments to the data they received from the companies before undertaking the efficiency analysis.[14]

[14] See Pollitt (2005).

If costs are benchmarked, a key issue is the extent to which local variations in labour costs are taken into account assuming a common cost of capital and materials costs. It would for instance be usual to make an adjustment for major city costs if wages are significantly higher for utilities based their than the national average. Though there remains a good argument for not doing this as higher wages may reflect higher labour productivity and hence outturn unit labour costs might be expected to be the same across a country.

As an example of the overall revenue impact of benchmarking. Here is an example from the GB electricity distribution price control from 2005-06. Operating costs are benchmarked using COLS (See Pollitt, 2005).

Regulated entity: United Utilities
£67.1m (2002-03): Normalised actual costs in benchmarked year.
£54.8m (82%): Efficient costs
£67.0m: Allowed operating costs in 2005-06 (including non-comparable opex with difference being mainly local property taxes)
£220.9m: Allowed revenue in 2005-06

Thus only 33% of actual revenue was frontier efficiency benchmarked. For comparison here are some different sources of regulated revenue impact:

+/- 5% on efficiency score:	+/-	£3.4m
+/- 1% p.a. on frontier shift:	-/+	£2.7m (by year 5)
+/- 1% on rate of return:	+/-	£9.2m
+/- 10% on capex:	+/-	£11.2m

If multiple efficiency scores for a single firm are calculated the issue is the extent to which they are adjusted by averaging or taking the highest efficiency score (usually the lowest would be considered an outlier). In the German example below, the highest score of the three calculated was used.

There then remains the issue of how to translate a final calculated efficiency score of, say, 80% into the X factor within the revenue cap formula.

One extreme approach would be to adjust the revenue from the first year of new price control period down by an amount consistent with a move to 100% efficiency. This would be a maximal P0 (initial price) adjustment. X could then be set equal to the expected productivity growth of the most efficient firms (which might be 1% p.a. in an advanced economy).

Another approach would be determine a final expected inefficiency gap by the end of the price control. Say this was to close 50% of the gap by year 5 of a 5-year price control period. Assume the gap was closed linearly. This might result in an X factor equivalent to improving efficiency by 2% in each of the 5 years to close the gap plus a general productivity growth factor of 1%.

These two paths for X are very different in terms of end points for revenue and the pathway of revenue during the 5 year price control (see Figure A.1) . The second would involve substantially higher revenue for the regulated firm. Both represent different degrees of confidence in the ability of a regulated firm to cut costs over a five year period on the basis of a regulatory assessment of efficiency.

Other approaches are possible. Following the initial price control of distribution companies in the Netherlands in 2000, the regulator (DTe) directly translated efficiency scores into revenue with a simple cap of 8% on X (see below).

An obvious mistake to avoid is misapplying efficiency scores to investments which should not be part of the benchmarking exercise. There has been an issue for both Netherlands and German utilities that the regulator calculates efficiency scores based on expenses occurred in advance of the outputs materialising. Thus if costs are incurred in advance of need this would could lower an efficiency score and have the effect of appropriating the return to an investment. By focussing efficiency analysis on opex (as in GB) this is much less likely to occur.

Regulators have used benchmarking to establish productivity growth in the sector. For instance in 2018 Ofgem undertook an exercise to look at long run productivity trends in the electricity sector using DEA (from 1990). These exercises resulted in the identification of a long run trend of around 1% p.a. in total factor productivity. This number then appeared as an efficiency target for all firms in the subsequent price control[15]. This work also found that including a wider range of quality and environmental outputs improves measured productivity growth in a sector which has been encouraged to improve quality of service and environmental impacts.

Lessons from leading jurisdictions on their use of benchmarking of electricity distribution

GB has made use of COLS based benchmarking of its 14 distribution utilities. This was very successful initially because the privatised companies were initially very inefficient.

Ofgem (and its predecessor, Offer) has put a lot of effort into data cleaning and adjustment to make comparisons of the operating efficiency between the firms. There was some attempt to ensure that 100% efficiency was not expected to be achieved initially. For the price control from 2000-2005, firms only had to remove 75% of their inefficiency by the second year. As we saw from the United Utilities example above, benchmarking was only one regulatory impact on revenue. Incentives for quality outperformance were also significant. Ofgem's extensive engagement process with individual firms has meant that there has been a high degree of trust in the process especially with respect to data cleaning. However the significance of benchmarking exercises for individual company review targets has declined over successive price control reviews.

[15] Ofgem (2022a), *RIIO-ED2 Final Determinations*, Nov. 2022, p.12.

Ofgem has evolved its use of benchmarking towards a pooled OLS approach making use of past and future projected data. This retains an emphasis on a straightforward regression but now incorporates more years of data with a time-trend (see Ofgem, 2022b, c)[16]. A good summary of how it used these approaches within the overall setting of X factors in the current price control (2023-2028) is provided by Oxera (2022).[17]

In the Netherlands, the energy regulator (Dte) undertook efficiency analysis of electricity distribution companies for the first time in autumn 2000 (see Nillesen and Pollitt, 2007). A July 1999 consultation document lays out ambitious plan to benchmark transmission and distribution utilities by Dte (the Dutch energy regulator).

The regulator analysed 19 electricity distribution companies using DEA, with no panel data or benchmarking. They did not cross-check with other methods and made no general adjustments for environmental factors. They capped the 3 year X factor at 8% p.a., however it was assumed that for companies who were not capped that they had to remove 100% of inefficiency by the end of the three year price control. Legal challenges were launched mid-2001 with appeal to the NMa (Dutch competition authority) against DTe by companies. DTe decisions overturned by the NMa and revised in 2002. Legal challenges led to 300 court cases! This reaction was surprising, especially as the network companies have remained in public ownership. It can be attributed to a poor relationship with companies due to the rapid timetable, an initial lack of education about the process and poor regulatory public relations. The regulatory strategy of DTe was not incentive compatible as it did not focus on revealing the scope for long-term cost reduction and was excessively short-termist. The Dutch regulator has not used frontier efficiency benchmarking for individual electricity distribution company X factors in subsequent price control reviews, but moved to a yardstick approach as mentioned above.

In Germany the Bundesnetzagentur undertook its first efficiency benchmarking in 2007.

They analysed 328 electricity and 488 gas distribution companies using DEA, SFA and COLS. They did do bootstrapping to get more robust estimates and did include environmental factors. In using the results of the estimated scores (they used the highest of the scores from each of the methods), they capped the maximum efficiency improvement over 10 years at 50%. This process was overseen by two leading academics of efficiency analysis: Per Agrell and Peter Bogetoft. This approach has evolved, but is still in use and exploits the large dataset that is available to the German regulator.[18]

General observations about the practice of benchmarking electricity distribution companies

There is a role for comparison of the results of benchmarking methods, especially to justify why a given method is representative or fair to the companies. Bauer et al. (1998) argue for observable consistency between approaches, namely: different

[16] For a discussion of the latest methodology see Ofgem (2022b, p.246ff).
[17] https://www.oxera.com/insights/agenda/articles/riio-ed2-final-determinations/
[18] A short summary of the current approach, from the regulator, is here: https://www.bundesnetzagentur.de/EN/RulingChambers/Chamber8/RC8_05_Revenue%20caps_revenue%20regulation/54_Efficiency%20benchmarking/Efficiency%20benchmarking_.html

approaches should have comparable means, standard deviations and distributional properties and rank orders, with the same firms at the top and bottom; demonstrable stability over time; consistency with what the market thinks (i.e. if the market thinks a firm is well-run we might question consistently poor frontier efficiency scores and vice versa); and that frontier method results should be consistent with results from non-frontier methods.

Some countries are doing a decent job with the data – the GB data has traditionally not been extensive - that they have and some simpler methods may serve a useful purpose within the context of trusted negotiation around the price control.

There is an issue with the failure of more sophisticated methods to deliver useable results on small samples of data. While Lovell (2006) is right about what makes for a more robust efficiency analysis, in practice getting data of this quality is difficult and even then it may not be useable with respect to all individual company efficiency scores.

In practice most efficiency exercises are naturally generous to companies because adjustments are almost always in favour of company revenue and because benchmarking exercises don't include all the comparator companies in the world (e.g. the utility in London does compare itself to those in New York and Paris, but the regulator does not).

There is (and should be) evidence that scores are converging over time and hence firm specific efficiency scores are becoming less material to revenue.

The percentage of allowed revenue subject to benchmarking low as comparable operating costs fall relative to depreciation and returns on regulatory asset bases.

There continues to be a lack of use international data in the regulation of electricity distribution.

Benchmarking processes are subject to ongoing legal obstacles and challenges. In the US it is generally illegal for a given regulated firm to be regulated on the basis of another firm's costs (hence the use of PBR arrangements) and comparative benchmarking is not used (see Joskow and Schmalensee, 2025).In other jurisdictions regulators are rightly subject to appeals against their use of the results of efficiency analyses.

The future of benchmarking

A concern of those who do benchmarking is whether repeated benchmarking is becoming harder to justify over time, especially when it comes to using individual firm efficiency scores.[19] This is in addition to those (e.g. Cronin and Motluk, 2007) who think that comparative efficiency benchmarking does not accurately mimic the pressure of

[19] See for example Agrell et al. (2013).

market to improve performance in the same way that a single yardstick does in Shleifer (1985).

This is because the underlying cost function becomes more difficult to model as both the outputs and inputs of the firms proliferate and differentiate. This is especially true with respect to distributed energy resources and innovation. Distributed energy resources (DERs) can vary significantly from network to network (e.g. amount of batteries, PV, small wind, smart demand side response), with difficult to measure impacts on costs. Some distribution networks (particularly larger ones with favourable geography) may have substantially more opportunity to use market based procurement of flexibility to optimise their networks, this may result in them looking very efficient in terms of the provision of their core service, when actually they have outsourced some service provision – reducing regulated opex and capex – to the 'market'. Innovation projects may be funded by specific innovation grants and have the effect of reducing core network costs, however benchmarking may not be capable of distinguishing between companies that have received innovation funding (e.g. to put in an Automated Network Management (ANM) system to better optimise the use of existing assets at no impact on costs) and those that have not. Similarly, if increased stakeholder engagement suggests some EDBs' customers are willing to pay for more undergrounding of wires this may increase costs and measured inefficiency under conditions where this reflects customer preferences. The above are examples where firms are doing fundamentally different things on the basis of differing opportunity and local preferences and cannot be regarded as having access to the same technological choices to be efficient (which is the fundamental assumption of all comparative benchmarking). This would also be an issue for a common total factor productivity (TFP) based X factor, where an EDB in an area with a relatively high preference for undergrounding would have lower TFP growth than the average.

This sort of increasing diversity – which itself reflects different circumstances between EDB areas - means that whole company comparisons of efficiency are difficult to justify without more serious attempts to focus in on core service costs and make relevant adjustments for favourable innovation funding or DER environments. This argues for more bespoke targets for efficiency based on the specific circumstances of the company rather than those based on a common concept of the efficient firm. While DEA and SFA remain of value in testing for general impacts (e.g. do innovative firms have higher efficiency?) and in estimating productivity trends, individual efficiency scores are likely more and more open to stakeholder and legal challenge.

Ofgem's RIIO process downgraded the role of benchmarking by suggesting that negotiated regulatory settlements, procurement auctions, private wires networks and innovation were alternatives to a regulator led process of cost assessment (see Pollitt 2010). More recently, Ofgem has been considering the future of the price control process once again from 2026, with increased emphasis being put on the need for regulatory learning across the price control period, including with respect to uncertainty mechanisms and adaptive regulation. Indeed, Pollitt et al. (2024) emphasise the need for a 'learning regulator' who has the capacity to use past information (as with conventional benchmarking), learn from stakeholders in real time

(about issues arising within price control periods) and be prepared for future learning points in the mid-term (e.g. when large context dependent investments, such as support for EVs or Heat Pumps, might need to be initiated).

Advanced country electricity distribution networks are struggling with rising costs and falling productivity (see: Ajayi et al., 2022, on Great Britain; CEPA, 2024b, on New Zealand; and AER, 2024, on Australia). Headline charges are further increased by the rising cost of capital which has started rising after many years of falling. A good example being the Netherlands where the real weighted average cost of capital (WACC) has had to be increased from a low of 1% in the most recent price control.[20]

Any regulatory benchmarking method is subject to a Lucas critique[21], as with all policy based on previous econometric (or estimated) relationships. It is good to begin with but then it is likely to become less effective over time. It is therefore a good idea to try it for a period and then drop for a period and then revisit. Prior to 1990 frontier efficiency AJanalysis was mostly used for hypothesis testing, rather than firm specific efficiency scores, increasingly we likely to be going back to the former use.

Recommendations

Electricity distribution firms benchmark themselves (or should do), therefore there would need to be a good reason for a regulator not to periodically check the relative efficiency of its regulated firms, against each other and international benchmarks. The New Zealand Commerce Act 1986, as currently drafted, prevents New Zealand from using comparative efficiency measures to set individual X factors (or initial price adjustments at the start of regulatory periods), however it does not prevent initial price adjustments on the grounds of high or low profitability. This means that EDBs with initially large opportunities to improve performance have had little incentive to do so over a prolonged period and that the whole sector has not been pushed to drive up productivity to the extent that might have been possible under comparative benchmarking.

Undertaking a comparative benchmarking exercise to establish whether individual X factors based on such exercises could yield significant benefits for New Zealand electricity consumers would be a useful exercise. This would establish whether a redrafting of the Commerce Act to remove the prohibition on comparative benchmarking was worth proposing.

A variety of benchmarking techniques exist, from activity benchmarking to frontier efficiency analysis. It is good practice to use several techniques to cross-reference the results. This is because this will enhance confidence in the efficiency targets set, either by suggesting that alternative methods give similar results or that there is a good reason why methods actually relied on are more reliable (e.g. a finding that DEA or SFA gives implausible results and should be dropped). It is also to be recommended that where

[20] https://think.ing.com/articles/dutch-utilities-addressing-the-capex-and-opex-challenges/
[21] Lucas, (1976).

multiple methods are used and considered equally reliable the highest efficiency score is used in X factor calculations.

A regulator can and should do reasonable data cleaning in order to ensure improved comparability between its regulated firms. This has the additional advantage of better understanding regulated firm costs and financial reporting. This is likely to drive firms to report on a similar basis making future comparisons easier.

Whichever frontier efficiency techniques for efficiency and productivity calculation are adopted, some allowance should be made for errors in measurement and representativeness of potential non-representativeness of benchmarked years. This can be done statistically within the chosen technique (e.g. via use of SFA, only setting the COLS frontier to the first quartile of efficiency or frontier stripping within DEA) or via adjustments ex post (e.g. setting a maximum value of X or assuming less than 100% closure of the efficiency gap by the end of the regulatory period) or a combination of two.

Direct translation of frontier efficiency scores into allowed revenue without adjustment or allowance of some time to reach 100% efficiency has resulted in issues and is not to be recommended. It is important to remember that regulators who find they have been too generous in setting X for an individual firm can adjust this (using evidence of revealed outperformance) at the next price control period.

Firm specific X factors might be more important in the early periods of the use of benchmarking, where there is both greater scope for efficiency improvement and greater variation in efficiency between firms. Simpler benchmarks (e.g. a common general productivity determined X factor with revenue sharing) can maintain incentives and may achieve the same results for customers in the long-run.

Benchmarking works best in circumstances where the firms are of similar sizes and have similar opportunities, and/or where firms are free to merge. If it were the case that small firms are at a substantial measured cost disadvantage in New Zealand and they were free to merge then the fact that a small firm was not capable of replicating the lower costs of a larger firm, might be less of an issue. However merger costs are real and must be adequately accounted for.

Q2: How many electricity distribution firms should NZ have – are very small firms inefficient?

New Zealand EDBs in a global context

New Zealand (NZ) is unusual because only New Zealand and the Netherlands have ownership unbundling of distribution and retail. This has been studied by Nillesen and Pollitt (2011 and 2021).

NZ has also been unusual in terms of the relatively lax regulation (arguably, zero regulation until the creation of the Electricity Authority in 2003) of distribution utilities for a large period (as documented by Bertram, 2006). This led to inflation of asset bases and, likely, a lack of pressure to cut costs, relative to the incentive regulation trend which took hold across Europe, US, South American and Australia (see Jamasb and Pollitt, 2001).

NZ also unusual because of consumer trust ownership structure which is a globally novel form of ownership for utilities (Haney and Pollitt, 2013b), this may both reduce cost of capital but also create mixed incentives for revenue control, due to the combination of electricity charges and dividends.

The global picture for size of EDBs

New Zealand has 29 electricity distribution firms for a population of 5.2 million. There has been rationalisation in NZ, as there were 60 in the 1980s. However NZ not unusual in terms of the existence of large number of firms with tail of smaller firms. NZ has 1 DSO (Distribution System Operator) per 169 thousand of population. There are at least 10 countries in the world significantly smaller DSOs on average, see Table 2.1 below.

Table 2.1: 10 countries with smallest population per DSO (Kufeoglu et al., 2018)

Country	No of DSOs	Legal Structure	Ownership	Access to electricity (%)	Population per DSO (thousand)
Germany	875	D	mixed	100	94.48
Finland	80	D	mixed	100	68.69
Austria	138	D	mixed	100	63.39
Sweden	170	D	mixed	100	58.25
Iceland	6	D	mixed	100	55.71
Czech Republic	290	D	mixed	100	36.42
Norway	146	D	mixed	100	35.84
Estonia	37	G, T, D, R	mixed	100	35.58
Maldives	35	G, D, R	public	100	11.93
Switzerland	900	G, D, R	mixed	100	9.3

One study calculated that there are 7600 electricity distribution companies across 175 countries (Kufeoglu et al., 2018). The legal entity covering electricity distribution, can be focused on just distribution (D) but it is often combined with retail (R); transmission (T); and generation (G) (see Table 2.2). Of the 7600, roughly 2900 that are legally separated from G,T and R. Most of the rest combine, legally, with at least R.

Globally, Kufeoglu et al. lost count at 7600 firms but in reality there are significantly more (e.g. there are 900 rural cooperatives in the US). If the global population was 7.7bn in 2018, that is roughly 1 per million population, if the 10 largest are excluded[22] (covering 2.2bn people), can say that the coverage is more might be more like 1 per 0.7m population. For New Zealand with a population of 4.9m in 2018, that would be 7 DSOs.

There is something to be said for allowing two-way movement between municipal and private ownership as we see in the US. 88 firms have moved in each direction in recent decades[23] (Pollitt, 2025).

Table 2.2: Organisational variety in electricity distribution utilities across the world

[22] Two in China and one each in Indonesia, Mexico, Vietnam, South Korea, South Africa, Ethiopia, Algeria, France and Thailand. One could also replace a couple of these with DRC and Tanzania, which have a single distribution utility and a large population, but low coverage rates.

[23] US municipalities buy 88 utilities since 1973 and sell 88 since 1980. See APPA, American Public Power Association (website accessed 30 August 2024).

Legal Structure	Ownership *Public*	*Private*	*Mixed*	*Total*	*Example*
T D	1	1	2	4	Japan
G D R	6	2	1	9	Philippines
T D R	8	0	2	10	Kenya
D	12	18	4	34	UK
D R	23	13	2	38	Nigeria
G T D R	67	5	5	77	Indonesia
Total	117	39	16	172	

T=Transmission, D=Distribution, G=Generation, R=Retail
172 countries: Largest DSO in each country.
Type of ownership is categorised as "mixed" only if the private or public shares represent less than 85% of the total
With inspiration from Trimble et al. (2016) and Kufeoglu et al. (2018).

Several countries have had significant historic rationalisations of the number of electricity firms. The Netherlands went from over 158 in 1989[24] to 8 today. GB went from 14 separate firms in 1990 to 7 groups today (and from 626 in 1935[25] to 14 in 1948).

There is a difference between regulated entity numbers and holding company numbers, e.g. in US[26], but also in UK. So important to remember that regulated legal operating unit – which may be what is analysed – is a relic of a former ownership structure, which may not be how the company is organised now. No studies that I am aware of do size analysis for countries with large numbers of distribution utilities at the level of the corporate entity, which is what matters for national productivity.

More recently Pollitt et al. (2025) look at 175 electricity distribution entities covering capital cities and find that even by focussing on one type of city, there is a huge variation in size, degree of integration and level of sophistication.

Economies of scale and scope

[24] See Hesseling and Sari (2006).

[25] Hannah (1979, p.244)).
[26] Approximately 58 companies hold 168 investor owned operating utilities. See https://en.wikipedia.org/wiki/Investor-owned_utilities_in_the_United_States

The definition of economies of scale and scope is based on the following formulae (see Pollitt and Steer, 2012 for a discussion).

Economies of Scale (*Sca)* can be defined by (Panzar and Willig, 1977):

$$Sca = \frac{C(q)}{\sum_{i=1}^{n} q_i C_i(q)}$$

Where q_i is a division of the total quantity of production q and C(q) is total cost and $C_i(q)$ is the marginal cost of producing q_i. If *Sca* > 1 then economies of scale exist and *Sca* describes the extent to which increasing scale reduces costs.

Economies of Scope (*Sco*) can be defined by (Willig, 1979):

$$Sco = \frac{C(q_1, 0) + C(0, q_2) - C(q_1, q_2)}{C(q_1, q_2)}$$

Where q_1 and q_2 are two types of output. If *Sco* > 0, then positive economies of scope exist and *Sco* describes how much costs would go up if jointly produced outputs were produced separately.

It is important to understand that these are both theoretical concepts in neoclassical economics which may not correspond to the level of the legal firm entity. Teece (1980) makes clear that even if both of these relationships hold in physical output they can be achieved with dispersed ownership in theory (his example is of the apple farmer and the sheep farmer doing a deal to graze sheep in an orchard). Thus economies of scale can be replicated by using joint output (e.g. renting part of a larger building), while economies of scope can be replicated by sharing assets (e.g. sharing databases between a separate distributor and retailer). Thus aggregate production can exploit economies of scale and scope without joint ownership in conditions when transaction costs involved in separate ownership are low (in the sense of Hart and Moore, 1990). This means that a small distributor can in theory replicate the unit costs of a larger one with appropriate contractual relationships. Difficulties in contracting with third parties in remote EDB areas to share capital, labour and materials may be offset by high community involvement and collective action leading to efficient and low cost operation (in the sense of Ostrom, 1990).

The fact that we generally don't observe all firms being of one optimal size is well documented in the empirical literature (see Hay and Liu, 1997). One of the reasons why smaller firms survive in electricity distribution is that while their physical inputs per unit output may be higher, their input prices are lower. Local labour may be relatively cheap and indeed companies may have access to very cheap or free labour from retired employees, who feel goodwill towards a local customer-owned enterprise. This may be especially true at peak times, where a company can cheaply call in the goodwill of former employees or contractors, in a way that a larger firm with less social capital may not be able to. Similarly, smaller firms may have access to cheaper local capital than larger firms. Lucas (1978) discusses the size distribution of firms as arising from a diversity of

scarce managerial talent and that some managers are more effective at managing different scales of firms.

Over time, there are reasons why small regulated firms may disappear. These include rising regulation, the need to raise and manage large amounts of capital (for major upgrades) or the rising cost of a lack of capacity to innovate or keep up with best practice. Often it is a lack of capacity to deal with these issues that gives larger firms advantages, meanwhile shareholders (public, private and consumer) see financial advantages in selling their assets rather than putting in more capital to maintain their value.

Another issue for EDBs is whether service areas need to be contiguous to exploit economies of scale and scope. The answer for many cost drivers of the business is no. However for some it is clearly yes. GB has exhibited both. Scottish Power-Manweb and Scottish Hydro Electric-Southern Electric were non-continguous mergers of distribution company areas. These were beneficial in terms of diversifying the exposure of the business to the regional economic variations. However Eastern Electricity-London Electricity-Seeboard was a great success in terms of joint operation over a large and already highly integrated economic region. It is important to say however that the regulator has maintained the setting of separate charges for each of the corporately merged regions, because of different underlying costs.

Parametric cost function estimation -to establish economies of scale (and scope with multiple outputs)- can be done via a tranlog cost function, following Christiansen and Greene (1976):

$$\ln(C) = \alpha_0 + \alpha_1 \cdot \ln(Q) + \tfrac{1}{2} \cdot \alpha_2 \cdot (\ln(Q))^2 + \Sigma\beta_i \cdot \ln(w_i) + \tfrac{1}{2} \cdot \Sigma\Sigma\gamma_{ij} \cdot \ln(w_i) \cdot \ln(w_j) + \Sigma\delta_i \cdot \ln(Q) \cdot \ln(w_i)$$

Where:

- C = total cost
- Q = output
- w_i = price of input i
- α, β, γ, and δ are parameters to be estimated

A simpler Cobb-Douglas functional form could be estimated:

$$\ln(C) = \alpha_0 + \alpha_1 \cdot \ln(Q) + \Sigma\beta_i \cdot \ln(w_i)$$

Economies of scale exist in both models, if following estimation the cost function is differentiated and the following scale variable is calculated:

$SCE = 1\text{-}dC/dQ$; if this is greater than 0 then scale economies exist, if SCE = 0 then there are constant returns to scale.

The evidence on economies of scale and scope

There is some evidence at the cross-sectional level on optimal scale of firms (see Pollitt, 1995, Kwoka 2005, Growitsch et al. 2009). These studies suggest that costs rise significantly below a minimum size threshold (for Kwoka's US sample this would be

around 20,000 customers in 1989). Growitsch et al. (2009) find that economies of scale continuously increase once an international sample is used, suggesting that national market size limits scale economies. Pollitt (1995) looks at a global sample; Kwoka (2005) at US utilities; and Growitch et al. (2009) at a sample of European utilities. A recent study (Soderberg and Vesterberg, 2025) looking at 171 Swedish electricity distribution utilities over 14 years found that optimal scale was achieved around 160,000 customers, suggesting the sector could benefit from substantial rationalisation. This supplements panel studies of the same firms at the national level which frequently shows the existence of strong economies of scale (and scope) for samples of Japanese electric power companies (Nemoto and Goto, 2004) and Chinese utilities (Liu et al., 2019). A review of 35 scale studies for the water industry found a variety of results for economies of scale within different samples, and a similar variation across 21 economies of scope studies (see Saal et al., 2013).

These studies illustrate several problems with the evidence from real world samples and regulated distribution entities. First, it is not clear that the examined sample just relates to the distribution function, for US utilities this is usually combined with retail. Indeed Kwoka explicitly draws attention to retail drivers for his findings. Second, international samples reveal that economies of scale within one country can look very different in a larger sample. UK firms are all much the same size within a European sample, while the finding that larger firms are inefficient in scale would only likely appear in a large sample. Third, panel data studies on the similarly integrated firms are not capable of analysing the counterfactual of whether disintegrated firms could have replicated the apparently falling unit costs of growing large firms. These sorts of studies have been used to justify the continuation of otherwise exceptional large and poorly performing utilities (e.g. TEPCO in Japan).

Nillesen and Pollitt (2010) found that there is a lot of diversity in performance at the operational level for a sample of US utilities, even though various performance factors (e.g. lower heating degree days) are generally correlated with better performance for electricity distribution. Looking at the top performing 25 electricity distribution utilities in their sample, they found they varied widely with respect to environmental conditions. This suggests that above certain thresholds firms with less favourable starting points can perform equally well. A given operational unit can replicate the performance of larger operational units, particularly where it is part of a larger group of units.

The finding that smaller existing firms can perform well is what we would expect, given the process of competition where better firms survive and less efficient ones tend to be taken over (Hay and Liu, 1997). That said, mergers do occur and there is a drift to fewer larger firms or larger holding companies over time in most major jurisdictions.

Are economies of scale and scope likely to change over time?

This is an important question which is explored in Pollitt (2025) in the context of how a move towards a net zero electricity system is likely to impact the size of utilities in generation and retail. For instance, is net zero likely to lead to more decentralisation or larger companies? More specifically, will the value added within the electricity sector

concentrate in larger firms? Distributed energy resources (DERs), with more customer and community ownership might be expected to lead to pressure to disintegrate the network into smaller entities. On the other hand increasing requirements to innovate and for sophisticated IT and procurement might be expected to favour larger firms. Issues such as cyber security and AI might also play out in unexpected ways for the scale of distribution firms. There might also be pressure to reintegrate electricity businesses which have been separated by the vertical unbundling model pursued in Europe, US, Australian and New Zealand utilities. Current economies of scale are based on large numbers of physical assets with distributed employed labour. Future economies might be based on international data processing companies, with little scale advantage at the level of the physical electricity assets.

Currently, however when it comes to innovation larger firms are favoured. Looking at the list of most innovative firms in the world with respect to smart grids[27], no small firm features in this list. Indeed one NZ firm appears and it is Vector (the largest EDB by number of customers).

The value of comparators

An important issue in many utility industries is the extent to which governments should encourage rationalisation of network industries into smaller numbers of larger firms. As Williamson (1968) noted scale economies trade off with rising anti-trust and regulatory issues. One large firm might in theory be efficient, but it reduces comparative competition to zero. Even in network industries this is a problem, because management teams can fail or become relatively less effective over time. Labour and financial market competition can benefit the whole industry and facilitate effective discipline on companies even for network monopolies which do not compete directly in the product market. A good example of apparently successful comparator based competition occurred when the gas distribution network companies (GDNs) were created in GB from the splitting off of gas distribution from a single gas transmission company. For the regulatory period 2008-13 (the first completed after the split), the number of comparators was 8 (4 were not owned by the former incumbent) and the value of comparators was calculated as an extra cost reduction of 3% p.a. (an NPV of saved costs of £310m).[28]

Maintaining a viable number of comparators has dynamic value beyond the static gains from rationalisation. This lead GB regulators to insist on a minimum threshold value to be realised before they would allow DNOs to merge.[29] This value had to be shared between shareholders and customers. This figure is arrived at from an estimate of the loss to the whole industry of estimated inefficiency as a result of losing one comparator. One obvious way to maintain the number of comparators while still allowing mergers to go ahead, it to maintain requirements to maintain the regulated entities post-merger. This has been the approach of Ofgem, who have maintained the original 14 regulated entities and used them for comparison, even though this is

[27] https://www.spgroup.com.sg/our-services/network/overview/smart-grid-index
[28] See Pollitt (2008, slide 25).
[29] The latest version of the policy is in Ofgem (2024b).

potentially gameable by corporate entities owning multiple regulated entities who may be able to reallocate costs between them to reduce their aggregate measured inefficiency.[30]

Comparators are most valuable when they are roughly the same size. This suggests that a rationalisation of the NZ EDBs should pay attention to both the loss of comparators and to the value of creating more effective comparators. Thus if several EDBs merged to create a firm that could be compared to the bigger EDBs this might actually be a good for the industry as a whole in terms of improving comparison. It also suggests that a useful end point might be to think about whether, say, 10-15 EDBs above a minimum threshold size might be the end point that was being aimed for. New Zealand also exempts certain EDBs entirely from CPI-X regulation and bringing all electricity customers under the protection of CPI-X regulation would seem to be desirable and increase the comparability of EDBs.

Comparison could be between groups of EDBs of roughly comparable in size and structure which could then be analysed separately from each other. This would suggest that some attention might be paid to maintaining a viable minimum number of comparators in a certain size class. This is an issue in the water industry in England and Wales.

It is worth saying that the extent to which the regulator can effectively benchmark NZ EDBs internationally would reduce the need to maintain domestic comparators by substituting them for overseas ones, however this is not to suggest that international benchmarking exercises are easy to do. Indeed one exercise that I was involved with European utilities for the Council for European Regulators (CEER) took a whole year to complete.

Final comments

There is no single right answer to the question of optimal scale of an EDB, because scale is a poorly defined economic concept with respect to the operating unit. Scale and scope trade off, thus a EDB that was also integrated with water, gas, waste, broadband etc. would get a 'scale' economies' that lie beyond the number of electricity customers it has.

However that does not mean that smaller firms should not be encouraged to rationalise into bigger groups or to share services. This is a global trend and is getting easier as transaction costs incurred in having shared services fall. Trends towards exploiting bigger scale are potential sources of productivity growth in the electricity sector.

The empirical evidence on optimal scale is very difficult to interpret because it is both contextually and methodologically dependent (see Saal et al., 2013). There does seem to be a minimum size threshold and often quite strong measured economies of scale,

[30] Jamasb et al. (2004).

but these are often static comparisons, which do not actually allow a direct inference that a given merger would improve efficiency.

It is worth saying that tighter regulation forces smaller companies to seek out rationalisation benefits. The mergers of GB firms were definitely encouraged by the pressure to cut costs and find efficiency savings, under incentive-based price control regulation. This has been a major driver across Northern Europe in reducing the number of firms over time.

References

AER (2024), *Annual Benchmarking Report Electricity distribution network service providers November 2024*, Commonwealth of Australia: Australian Energy Regulator.

Agrell, P.J., Bogetoft, P. and Mikkers, M. (2013). 'Smart-grid investments, regulation and organization', *Energy Policy*, 52: 656-666.

Ajayi, V., Anaya, K.L. and Pollitt, M.G. (2022), 'Incentive regulation, productivity growth and environmental effects: the case of electricity networks in Great Britain', *Energy Economics*, 115 (November): 106354.

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics*, 95 (9–10): 1082-1095.

Bauer, P.W., Berger, A.N., Ferrier, G.D. and Humphrey, D.B. (1998), 'Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods', *Journal of Economics and Business*, 50(2): 85-114.

Beesley, M. E., & Littlechild, S. C. (1989), 'The Regulation of Privatized Monopolies in the United Kingdom', *The RAND Journal of Economics*, *20*(3): 454–472. https://doi.org/10.2307/2555582

Bertram, G. (2006), 'Restructuring the New Zealand Electricity Sector 1984-2005', in Sioshansi, F.P. and Pfaffenberger, W. (eds.) (2006), *Electricity Market Reform: An International Perspective,* Oxford: Elsevier. pp.203-234.

CC (Competition Commission) (2000), *Mid Kent Water plc: a report on the references under sections 12 and 14 of the Water Industry Act 1991*. London: Stationery Office.

CEER (2025), *Report on Regulatory Frameworks for European Energy Networks 2024, Incentive Regulation and Benchmarking Work Stream*, https://www.ceer.eu/wp-content/uploads/2025/02/GA191_05_2_IRB-Regulatory-Frameworks-Report-2024-Main-report-1-1.pdf

CEPA (2024a), *PR24 Draft Determinations: Frontier Shift, Real Price Effects and the energy crisis cost adjustment mechanism, Ofwat, 28 June 2024*. London: CEPA. https://www.ofwat.gov.uk/wp-content/uploads/2024/07/CEPA-frontier-shift-real-price-effects-and-the-energy-crisis-cost-adjustment-mechanism.pdf

CEPA (2024b), *EDB Productivity Study: A report prepared for the Commerce Commission 24 June 2024*, London: CEPA.

Christensen, L. R., & Greene, W. H. (1976), 'Economies of Scale in U.S. Electric Power Generation', *Journal of Political Economy*, *84*(4), 655–676. http://www.jstor.org/stable/1831326

CMA (Competition and Markets Authority), (2022), Energy Licence Modification Appeals: Competition and Markets Authority Guide, CMA71, London: Competition and Markets Authority. https://assets.publishing.service.gov.uk/media/635946898fa8f557cfc22c4f/Energy_Guidance.pdf

Coelli, T.J., Prasada Rao, D.S., O'Donnell, C.J. and Battese, G.E. (2005), *An Introduction to Efficiency and Productivity Analysis*, 2nd Edition, Boston, Kluwer.

Cronin, F.J. and Motluk, S.A. (2007), 'Flawed Competition Policies: Designing 'Markets' with Biased Costs and Efficiency Benchmarks', *Review of Industrial Organization,* 31: 43–67. https://doi.org/10.1007/s11151-007-9142-z

Foley, B. and Goldstein, H. (2012), *Measuring Success: League tables in the public sector March 2012*, London: British Academy.

Frontier Economics (2023), *Benchmarking Techniques and Practices for Electricity and National Gas Network Operators, Report prepared for ACM 12 October 2023*, WWW.FRONTIER-ECONOMICS.COM

Greene, W. (2005), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics 126* (2), 269–303.

Growitsch, C., Jamasb, T., & Pollitt, M. (2009). Quality of service, efficiency and scale in network industries: an analysis of European electricity distribution. *Applied Economics*, *41*(20), 2555–2570.

Haney, A. and Pollitt, M. (2013a), 'International Benchmarking of Electricity Transmission by Regulators: A contrast between theory and practice?', *Energy Policy*, Vol.62, November 2013, pp.267-281.

Haney, A. and Pollitt, M. (2013b), 'New Models of Public Ownership in Energy', *International Journal of Applied Economics*, Vol.27, No.2, pp.174-192.

Haney, A. and Pollitt, M.G. (2009), 'Efficiency analysis of energy networks: An international survey of regulators', *Energy Policy*, 37: 5814-5830.

Hannah, L. (1979), *Electricity Before Nationalisation*, The Macmillan Press.

Hart, O. and Moore, J. (1990), 'Property Rights and the Nature of the Firm',, *Journal of Political Economy*, 98:6, 1119-1158. https://doi.org/10.1086/261729

Hay, D.A. and Liu, G.S. (1997), 'The Efficiency of Firms: What Difference Does Competition Make?', *The Economic Journal,* 107, 597–617.

Hesseling, D. and Sari, H. (2006), *The introduction of quality regulation of electricity distribution in The Netherlands*, European Energy Law Report III (EELR III-paer), 2006,

pp.127-145.
https://www.acm.nl/sites/default/files/old_publication/publicaties/11528_the-introduction-of-quality-regulation-of-electricity-distribution-in-the-netherlands-dennis-hesseling-mahir-sari.pdf

Jamasb, T., Nillesen, P. and Pollitt, M. (2004), 'Strategic behaviour under regulation benchmarking' *Energy Economics*, Vol.26, No.5, pp.825-843.

Jamasb, T. and Pollitt, M. (2008) "Reference models and incentive regulation of electricity distribution networks: an evaluation of Sweden's Network Performance Assessment Model (NPAM)." Energy Policy, 36(5): 1788-1801.

Jamasb, T. and Pollitt, M. (2007) "Incentive regulation of electricity distribution networks: lessons of experience from Britain." *Energy Policy*, 35(12): 6163-6187.

Jamasb, T. and Pollitt, M. (2001), 'Benchmarking and Regulation of Electric Utilities: Lessons from International Experience'*, Utilities Policy,* Vol.9, No.3, pp.107-130.

Johnson, A.L. and Kuosmanen, T. (2015), 'An Introduction to CNLS and StoNED Methods for Efficiency Analysis: Economic Insights and Computational Aspects'. In: Ray, S., Kumbhakar, S., Dua, P. (eds), *Benchmarking for Performance Evaluation*. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2253-8_3

Joskow, P.L. and Schmalensee, R. (2025), 'Cost-of-service regulation of electricity distribution services in the US' in Glachant, J-M., Joskow, P.L. and Pollitt, M.G. (Eds.), *Handbook on Electricity Regulation*, Edward Elgar, pp.44-67.

Kufeoglu, S., Pollitt, M. and Anaya, K. (2018), *Electric power distribution in the world: today and tomorrow*. Energy Policy Research Group Working Papers, No.1826. Cambridge: University of Cambridge.

Kwoka, J. (2005), 'Electric power distribution: economies of scale, mergers, and restructuring' *Applied Economics*, 37(20): 2373-2386.

Littlechild, S.C. (2021), *An Overall Customer Satisfaction score for GB energy suppliers*, 12 March 2021, EPRG Working Paper 2027.
https://www.jbs.cam.ac.uk/wp-content/uploads/2023/12/eprg-wp2027.pdf

Littlechild, S.C. (2007), "Bird in hand: stipulated settlements and electricity regulation in Florida," EPRG Working Paper, No. 0705.

Littlechild, S.C. (1983), *Regulation of British Telecommunications' Profitability*, London: HMSO.

Liu, X-Y., Pollitt, M.G., Xie, B-C.,Liu, L-Q. (2019), 'Does environmental heterogeneity affect the productive efficiency of grid utilities in China?', *Energy Economics*, 83: 333-344.

Llorca, M., Orea, L. and Pollitt, M. (2014), 'Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry', *Operations Research Perspectives*, Vol.1 (1), pp.6-17.

Lovell, K. (2006). *Frontier Analysis: Recent Advances and Future Challenges*. Keynote Address, North American Productivity Workshop, Stern School of Business, New York University, New York.

Lucas, R.E. (1978), 'On the Size Distribution of Business Firms' *The Bell Journal of Economics* 9, 508–523.

Lucas, R.E. (1976), 'Econometric Policy Evaluation: A Critique' In Brunner, K.; Meltzer, A. (eds.). The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy. Vol. 1. New York: American Elsevier. pp. 19–46. https://people.sabanciuniv.edu/atilgan/FE500_Fall2013/2Nov2013_CevdetAkcay/LucasCritique_1976.pdf

Nemoto, J. and Goto, M. (2004), 'Technological externalities and economies of vertical integration in the electric utility industry', *International Journal of Industrial Organization*, 22(1): 67-81.

Nillesen, P. and Pollitt, M. (2021), 'Ownership Unbundling of Electricity Distribution Networks', *Economics of Energy and Environmental Policy*, 10(1): 147-158. 10.5547/2160-5890.10.1.pnil

Nillesen, P.H. and Pollitt, M.G. (2011), Ownership unbundling in electricity distribution: empirical evidence from New Zealand, *Review of Industrial Organization*, Vol.38. No.1, pp.61-93.

Nillesen, P.H.L. and Pollitt, M.G. (2007) "The 2001-2003 electricity distribution price control review in the Netherlands: regulatory process and consumer welfare." *Journal of Regulatory Economics*, 31(3): 261-287.

O'Donnell, C.J. (2018), *Productivity and Efficiency Analysis: An Economic Approach to Measuring and Explaining Managerial Performance*, Springer.

Ofgem (2024a), *Distribution System Operation (DSO) Incentive Report for Regulatory Year 1 April 2023 -31 March 2024*, London: Ofgem.

Ofgem (2024b), *Ofgem's approach to energy network mergers and statement of methods*, London: Ofgem.
Ofgem (2022a), *RIIO-ED2 Final Determinations, Nov. 2022*, London: Ofgem.

Ofgem (2022b), *RIIO ED2 Draft Determinations Core Methodology Document, 19 June 2022,* London: Ofgem.

Ofgem (2022c), *RIIO ED2 Final Determinations Core Methodology Document, 30 November 2022,* London: Ofgem.

Ofgem (2010), *Regulating energy networks for the future: RPI-X@20 decision document*, London: Ofgem.

Olesen, O.B. and N.C.Petersen (1995), 'Chance Constrained Efficiency Evaluation', *Management Science*, Vol.41, No.3, pp.442-457.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Panzer, J.C. and Willig, R.D.,(1977), 'Economies of Scale in Multi-Output Production', *The Quarterly Journal of Economics* 91, 481–493.

Pollitt, M.G., Duma, D., Covatariu, A. and Nillesen, P. (2025), *A Global Map of Electricity and Gas Distribution Network Companies*, EPRG Working Papers, forthcoming.

Pollitt, M.G., Covatariu, A. and Duma, D. (2024), *Towards a more Dynamic Regulation of Energy Networks*, Brussels: Centre on Regulation in Europe. https://cerre.eu/publications/towards-a-more-dynamic-regulation-for-energy-networks/

Pollitt, M.G. And Steer, S.J. (2012), 'Economies of Scale and Scope in Network Industries: Lessons for UK water and sewerage sectors*', Utilities Policy*, 21(C): 17-31.

Pollitt, M.G. (2025), 'Scale, governance and net zero: decentralisation vs centralisation in electricity'. In Sareen, S., Inderberg, T.H.J. and Eikeland, P.O. (eds.), *Are low-carbon futures decentralised? The Governance of Collective Electric Systems*, Edward Elgar, forthcoming.

Pollitt, M. (2010), 'Does Electricity (and Heat) Network Regulation have anything to learn from Fixed Line Telecoms Regulation?', *Energy Policy*, Vol.38 (3), pp.1360-1371.

Pollitt, M.G. (2008) "The future of electricity (and gas) regulation in a low-carbon policy world." The Energy Journal, 29(S2): 63-94.

Pollitt, M. (2005) "The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks." Utilities Policy, 13(4): 279-288.

Pollitt, M.G. (1995), *Ownership and Performance in Electric Utilities*. Oxford: Oxford University Press / Oxford Institute for Energy Studies.

Saal, D.S., Arocena, P. and Maziotis (2013), 'Scale and scope economies and the efficient vertical and horizontal configuration of the water industry: a survey of the literature', *Review of Network Economics,* 12(1):93-129.

Shleifer, A. (1985), 'A Theory of Yardstick Competition', *The RAND Journal of Economics*, *16*(3): 319–327. https://doi.org/10.2307/2555560

Söderberg, M. and Vesterberg, M. (2025), 'Scale properties and efficient network structures in the Swedish electricity distribution market', *Journal of Regulatory Economics*, https://doi.org/10.1007/s11149-025-09488-7

Teece, D.J. (1980), 'Economies of Scope and the Scope of Enterprise', *Journal of Economic Behaviour and Organization* 1, 223–247.

Trimble, C., Kojima, M., Perez Arroyo, I. and Mohammadzadeh, F. (2016), *Financial Viability of Electricity Sectors in Sub-Saharan Africa: Quasi-Fiscal Deficits and Hidden Costs*, World Bank Policy Research Working Paper 7788.

Uri, N.D. (2002), 'A Note on Measuring the Productivity Offset for Regulatory Purposes', *Journal of Productivity Analysis* 17: 201–211. https://doi.org/10.1023/A:1015055904013

Weyman-Jones, T. (2006), *Efficiency and productivity analysis in regulation and governance*. *4th* NAPW Plenary Paper, New York.

Weyman-Jones, T.J. (2001), *Stochastic Non-parametric Efficiency Measurement in Electricity Distribution*, mimeo.

Williamson, O. E. (1968), 'Economies as an Antitrust Defense: The Welfare Tradeoffs', *American Economic Review*, 58: 18-36.

Willig, R. D. (1979), 'Multiproduct Technology and Market Structure' *The American Economic Review*, *69*(2): 346–351. http://www.jstor.org/stable/1801671

Appendix

Figure A.1 Example of regulatory of revenue profile over a five year price control from 2010-15

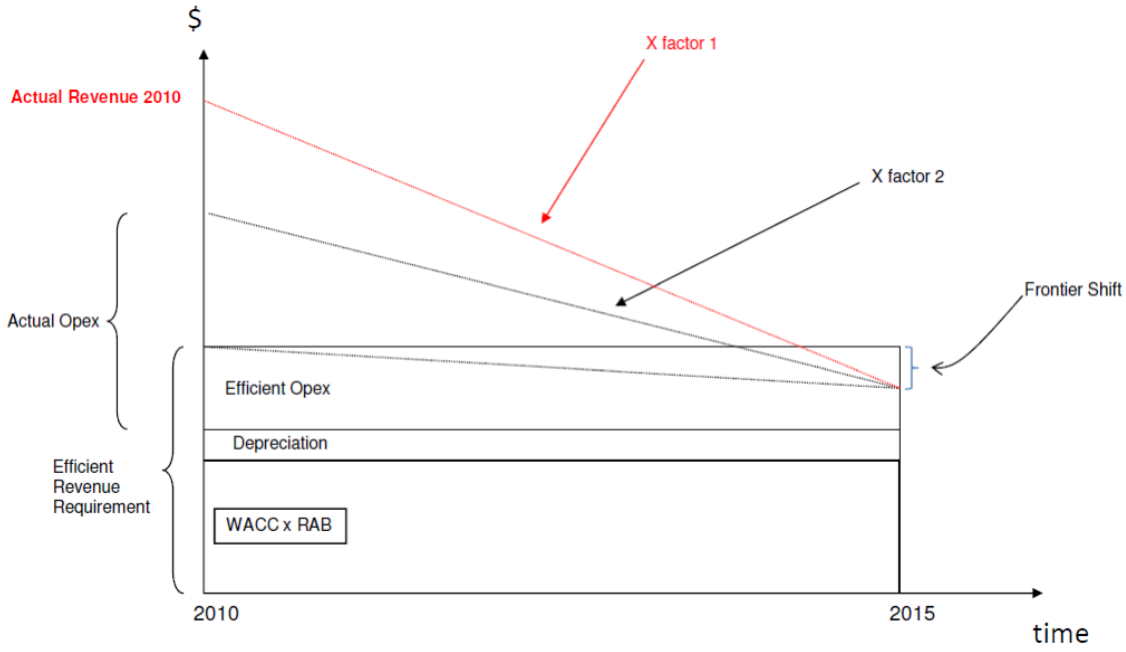


Figure A.2: Comparison of Efficiency Score calculations under DEA and COLS

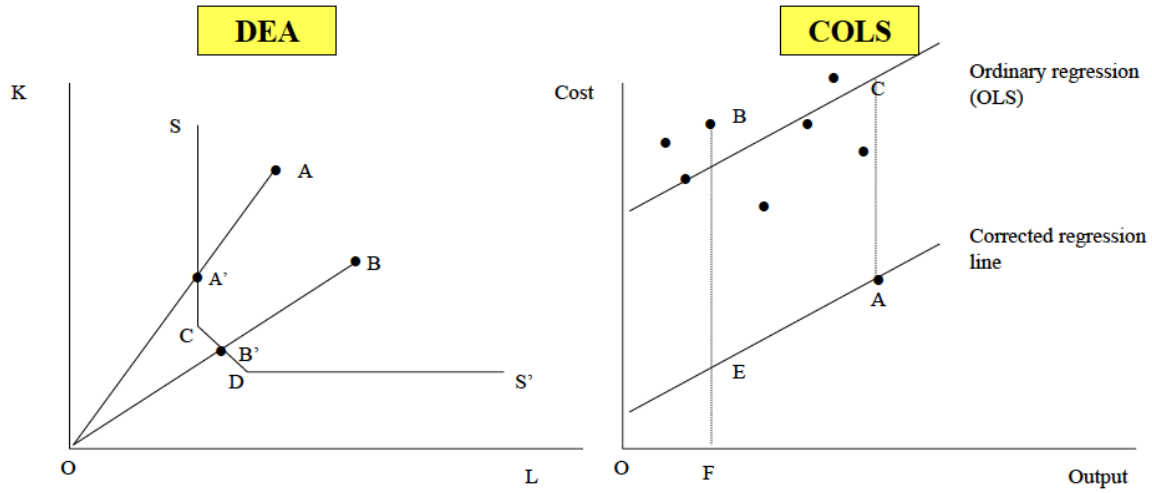


Figure A.3: Comparison of Efficiency Score calculation under SFA and COLS

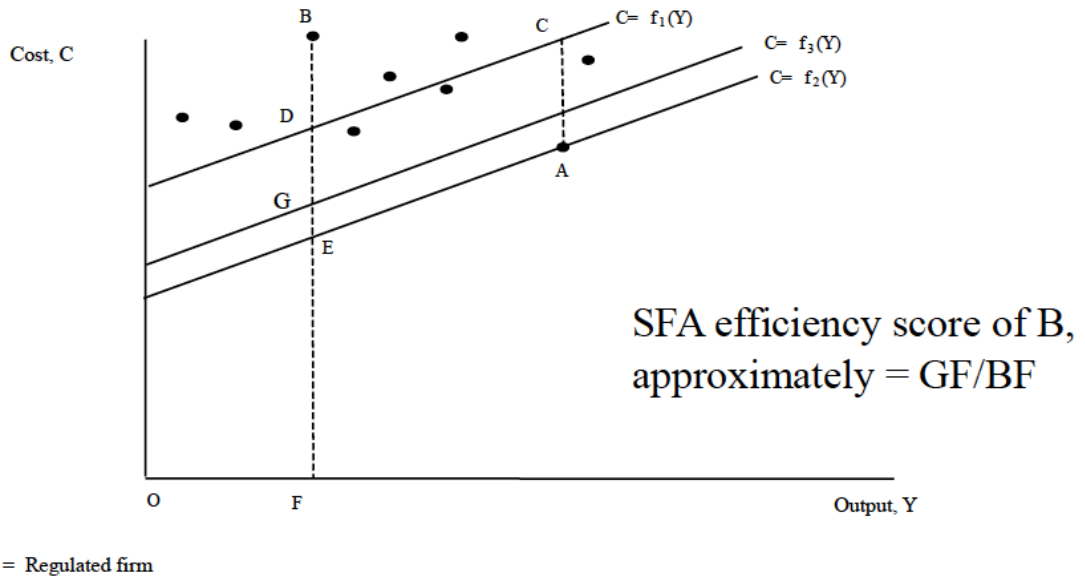
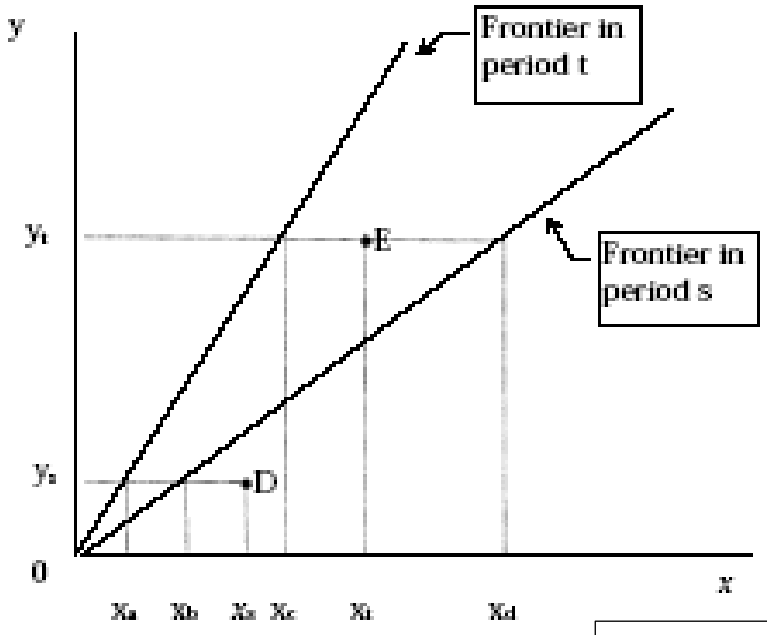


Figure A.4: Calculation of Productivity Growth using Malmquist Index
This shows how a productivity index change can be measured between two periods, s and t.



$$\text{Efficiency change} = \frac{x_c / x_t}{x_b / x_s}$$

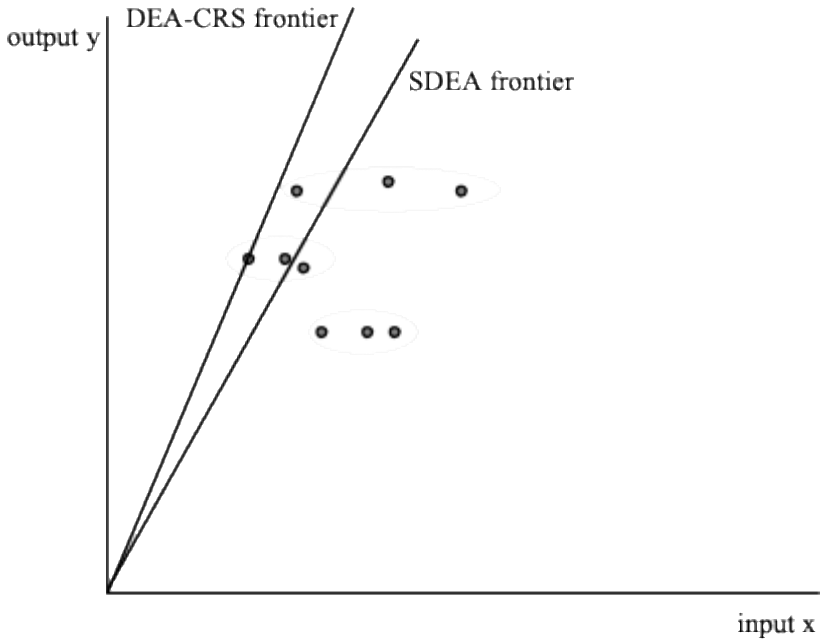
$$\text{Technical change} = \left[\frac{x_d / x_t}{x_c / x_t} \times \frac{x_b / x_s}{x_a / x_s}\right]$$

Efficiency change = EC
Technical change = TC
Malmquist TFP (MTFP) = EC * TC

If this MTFP number was 1.02,
this would indicate productivity had risen by 2% and this would be made up changes in relative efficiency (EC) and a movement in the frontier itself (TC).

Figure A.5: Comparison of DEA Efficiency Scores before and after frontier stripping

Figure A.6: Comparison of Efficiency Score calculation under DEA-CRS nd SDEA. (CRS = constant returns to scale)



DEA and SDEA frontiers; efficiency may exceed unity for SDEA realisation case

Source: Weyman Jones (2001)