Ausgrid

Attachment 1.08

Pacific Economics Group (PEG) - Statistical Benchmarking for NSW Distributors, Jan 2015

January 2015

Ausgrid revised regulatory proposal attachment

STATISTICAL BENCHMARKING FOR NSW DISTRIBUTORS

Pacific Economics Group Research, LLC

STATISTICAL BENCHMARKING FOR NSW POWER DISTRIBUTORS

Mark Newton Lowry, PhD
President

Matthew Makos
Consultant

Kaja Rebane
Economist II

19 January 2015

PACIFIC ECONOMICS GROUP RESEARCH LLC

22 East Mifflin, Suite 302

Madison, Wisconsin USA 53703

608.257.1522   608.257.1540 Fax

Table of Contents

Expert Witness Acknowledgment

My name is Mark Newton Lowry. I have a Ph.D. in Applied Economics from the University of Wisconsin-Madison and 30 years of experience as a professional economist.

I have been engaged by Networks NSW to prepare an expert report, in the context of the National Electricity Objective, which outlines the challenges of benchmarking the performance of electricity distributors, reviews the AER's 2014 *Annual Benchmarking Report* and opex benchmarking report, its supporting materials, and the AER's opex proposals for Networks NSW in their price determination, and comments on the use of benchmarking in utility regulation. The detailed terms of reference is attached to this report.

I acknowledge that I have read, understood and complied with the Federal Court of Australia's Practice Note CM 7, "Expert Witnesses in Proceedings in the Federal Court of Australia". I have made all inquiries that I believe are desirable and appropriate to answer the questions put to me. No matters of significance that I regard as relevant have to my knowledge been withheld. I have been provided with a copy of the Federal Court of Australia's Guidelines for Expert Witnesses in Proceeding in the Federal Court of Australia, and confirm that this report has been prepared in accordance with those Guidelines.

I have been assisted in the preparation of this report by Dave Hovde, Matt Makos, Kaja Rebane, Stelios Fourakis, and Gretchen Waschbusch from my Madison, Wisconsin office. However, the opinions set out in this report are my own and are based wholly on the specialized experience set out in Section 2 of the report.

Mark Newton Lowry
16 January 2015

1. Introduction

Spurred in part by recent amendments to the National Electricity Rules, the Australian Energy Regulator ("AER") is making extensive use of benchmarking in its reviews of distribution network service provider ("DNSP" or distributor) costs. An *Annual Benchmarking Report* was released in November 2014 which featured work by an AER consultant, Economic Insights ("EI"). In the same month, the AER released its preliminary recommendations for multiyear revenue requirements of distributors serving New South Wales ("NSW") and the Australian Capital Territory ("ACT"). Recommendations for network services operating expenditure ("opex") revenue are well below those proposed by the distributors. The AER's opex recommendations rely heavily on EI research detailed in its November 2014 report (the "Opex Bench" report) titled *Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs.*

Networks NSW owns the three New South Wales distributors: Ausgrid, Endeavour Energy, and Essential Energy. The company has retained Pacific Economics Group ("PEG") Research LLC to appraise EI's study of their network services opex. We have also been asked to review the evolving role of statistical benchmarking in utility regulation in overseas jurisdictions.

This document is the report on our research. Following a statement of our credentials to undertake this work in Section 2, Section 3 provides a largely nontechnical discussion of benchmarking methods. Section 4 discusses the general challenge of benchmarking network services opex. Our review of benchmarking precedents is discussed in Section 5. Section 6 presents our critique of EI's benchmarking work. An Appendix provides further details of our benchmarking precedent review.

2. Credentials

PEG Research LLC is a company in the USA-based Pacific Economics Group consortium which is active in utility economics. Our staff includes several well-known PhD economists. Larry Kaufmann and Mark Newton Lowry are experts on statistical performance research and modern regulation. Charles Cicchetti, former chair of Wisconsin's Public Service Commission, recently retired as an economics professor at the University of Southern California. Jeff Dubin, previously professor of economics at the California Institute of Technology, teaches econometrics at the University of California Los Angeles.

The PEG Research team based in Madison, Wisconsin includes leading practitioners of statistical research on utility performance and alternatives to traditional utility regulation such as the price control plans (aka multiyear rate plans) used in Australia. We have over sixty man years of experience in these fields, which share a foundation in economic statistics. The University of Wisconsin trained most of our staff and is known internationally for its economic statistics program. We periodically write articles on our research in refereed journals. Our practice is multinational in character and has to date involved projects in twelve countries, including many in Australia and New Zealand. Work for a mix of utilities and regulators has given us a reputation for objectivity and dedication to regulatory science.

2.1 Benchmarking Experience

Statistical benchmarking is the appraisal of performance using statistical methods to fashion benchmarks and make performance comparisons. Good benchmarking methods are encouraged in North American regulation by the availability of abundant, quality data and the high standards of evidence required in proceedings, which have a quasi-judicial character and often involve extensive data requests, technical conferences, and sworn oral testimony. PEG Research has responded to this opportunity by becoming a pioneer in the use of scientific benchmarking in regulation.

Benchmarking the cost and reliability of power DNSPs is a company specialty. We have also benchmarked costs of power generation and transmission, bundled power service, and gas distribution. In addition to our numerous benchmarking studies of operating expenses ("opex"), we have been doing rigorous research on capital and *total* cost performance for more than two decades.

Clients Our personnel have testified on benchmarking for AmerenUE, Atlanta Gas Light, Boston Gas, Central Vermont Public Service, Enbridge Gas Distribution, Fortis Alberta, Hydro One Networks, Kentucky Utilities, Louisville Gas & Electric, the Michigan Public Service Commission, NMGas, Oklahoma Gas & Electric, the Ontario Energy Board, Pacific Gas & Electric, Portland General Electric, Progress Energy Florida, Public Service of Colorado, San Diego Gas & Electric, Southern California Edison, and Southern California Gas. Other clients of our benchmarking services have included AGL Electricity, the Electric Power Supply Association of Australia, Energex, Envestra, Ergon, ESCOSA, Multinet, the National Electricity Distributors' Forum, Powercor, Powerlink Queensland, Transend, the Queensland Competition Authority, TXU Australia, and United Energy (Australia), the Superintendencia de Electricidad (Bolivia), the Canadian Electricity Association and Hydro-Quebec Trans-Energie (Canada), Aqualectra (Curacao), EDF London, EDF Eastern, EDF Seeboard, Northern Electricity Distribution, Yorkshire Electricity Distribution, and United Utilities (England), Jamaica Public Service (Jamaica), the Central Research Institute for the Electric Power Industry (Japan), NGC, Powercor, United Networks, and Vector (New Zealand), and Central Maine Power, Commonwealth Edison, Delmarva Power and Light, Niagara Mohawk Power, Pennsylvania Power & Light, and Public Service Electric & Gas (United States).

Our benchmarking practice began with research and testimony for Southern California Edison, one of the largest US utilities, in 1994. Here are brief descriptions of benchmarking projects we have subsequently undertaken which are especially relevant to this report.

- In the United States, we have provided research and testimony on power distributor opex and reliability for Portland General Electric (2010) and on power distributor total cost for Central Vermont Public Service (2006), San Diego Gas and Electric (2000, 2002, & 2006), and Oshawa PUC Networks (forthcoming).

- In Canada, we prepared for the Canadian Electricity Association a white paper on the challenge of power distribution benchmarking and its role in regulation (2006) and a review of their distribution benchmarking program (2008). We testified on the role of power distribution benchmarking in regulation for Fortis Alberta (2006).

- For the Ontario Energy Board we have twice prepared cost benchmarking studies for more than 70 provincial power distributors (2008 and 2013). The first of these studies addressed their opex. The second addressed their total cost. These studies have been used to set X factors in price control plans. We recently developed new reliability and total cost benchmarking models for the Board using US data.

- In Britain, we advised the Northern and Yorkshire power distributors on a benchmarking policy proposal for a price control update (2004). We have provided confidential opex benchmarking studies for three British DNSPs.

- In work for the Superintendencia de Electricidad we benchmarked the cost performance of four Bolivian power distributors (2003).

- In Germany, we prepared a review of the use of benchmarking in regulation for the Bundesnetzagentur (2006).

- In Australia, we have benchmarked the cost performance of power distributors in work for Victorian distribution businesses (1998, 1999) and the Queensland Competition Authority (2000).

- We prepared a white paper on benchmarking principles and applications for Victorian electric DNSPs (2000).
- We prepared a white paper on the cost structure of power distribution for the Electricity Supply Association of Australia (2000).
- Using US data, we have benchmarked the power transmission cost of Powerlink Queensland (2000) and Transend (2002).
- Using US and Australian data, we benchmarked the cost of Australian gas DNSPs in work for several distributors (2002).
- We reviewed benchmarking studies filed by Victorian DNSPs in work for the Essential Services Commission (2004).
- We have authored an article on the use of benchmarking in regulation in *Energy Policy* (2009) and an authoritative study on the benchmarking of power distributor total cost and its regulatory application for the *Energy Journal* (2005).
- In July 2014 we completed a project for the AER that involved the gathering of US operating data and an appraisal of its usefulness in benchmarking Australian DNSPs. Our report featured illustrative econometric benchmarking models developed using only Australian data as well as a transnational US/Australian dataset.

2.2 Price Control Experience

We are the leading practitioners of the North American approach to price control (aka multiyear rate plan) design in which rates (or revenues) are escalated by indexes based on cost trend (e.g., input price and productivity) research. We have testified on our input price and productivity research in numerous proceedings. Here are some projects that are especially relevant to this report.

- On behalf of TXU Australia, we argued for the legality of permitting a DNSP to operate under the North American approach to price control design before the Supreme Court of Victoria.
- In work for SP AusNet, we developed a "rate of change" formula for escalating opex that was approved for use by the Essential Services Commission.
- We have twice done the input price and productivity research which the Ontario Energy Board has used to set X factors for provincial power distributors.
- We have performed statistical cost research for the Electricity Networks Association in the three most recent price control updates for New Zealand DNSPs.
- We assisted the Essential Services Commission in Melbourne in several price control updates.

2.3 Key Personnel

Here are brief discussions of PEG Research personnel who participated in this project.

Mark Newton Lowry Dr. Lowry is the President of PEG Research and serves as principal investigator for the project. He has thirty years of experience as a professional economist. The economics of utility regulation and statistical research and testimony on utility cost has been his chief professional focus for over twenty years. He has testified dozens of times on his research.

Before joining PEG, Dr. Lowry was a Vice President at Christensen Associates and directed a Regulatory Strategy group there.[1] He has also served as an Assistant Professor of Mineral Economics at the Pennsylvania State University and as a visiting

[1] All of the key members of this group now work for PEG Research.

professor at l'Ecole des Hautes Etudes Commerciales in Montreal, Canada. His academic research and teaching featured the use of mathematical theory and econometrics in industry analysis. He can assist clients in French and Spanish as well as his native English.

His resume includes an extensive list of publications and public appearances. For example, he has chaired several conferences on benchmarking. He has been involved in several Australian benchmarking projects, including the lead role in the recently completed AER project. Dr. Lowry attended Princeton University and holds a Ph.D. in Applied Economics from the University of Wisconsin.

Dave Hovde is a Vice President of PEG Research. He supervises our database management and is active in our Ontario, New Zealand, and US statistical work. Dave has two decades of utility cost research experience, including all of our Australia and New Zealand ("ANZ") projects. Before joining PEG, Dave was a Senior Economist at Christensen Associates. He holds an MA in Economics and undergraduate degrees in Economics, Political Science, and International Relations from the University of Wisconsin.

John Kalfayan is a Senior Advisor at PEG Research and our senior econometrician. Before joining PEG Research, he worked as a Senior Economist at Christensen Associates. He earned an ABD status in Economics at the University of Wisconsin.

Kaja Rebane is an Economist II at PEG Research. She has played a leading role in our new Australian benchmarking work. A talented econometrician, database manager, and writer, she has an undergraduate degree from Stanford University and Master's degrees in Land Resources and Applied Economics from the University of Wisconsin. Kaja holds an ABD status in the Environment and Resources program at UW and is working on her PhD.

Gretchen Waschbusch is our Office Manager and often helps out in our research projects. For example, she gathered and processed weather data in our recent AER project. She has an undergraduate in Business Administration from the University of Wisconsin and a Masters of Business Administration from Edgewood College.

Matt Makos Matt has been a Consultant for several years at PEG Research. He played a leading role in gathering precedents for the use of benchmarking in regulation. Matt holds an undergraduate Business degree from the University of Wisconsin.

Stelios Fourakis Stelios is an Economist I at PEG Research. He developed our Australian input price indexes in the AER project and assisted with our discussion of methodological issues. Stelios holds an undergraduate degree in Political Economy from Georgetown University.

3. An Introduction to Benchmarking

In this section, we consider benchmarking methods and concepts that are central to our discussion of EI's benchmarking work. The two benchmarking methods featured in EI's Opex Bench report are explained. The discussion is largely non-technical.

3.1 What is Benchmarking?

The word benchmark comes from the field of surveying. The *Oxford English Dictionary* defines a benchmark as

> A surveyors mark, cut in some durable material, as a rock, wall, gate pillar, face of a building, etc. to indicate the starting, closing, ending or any suitable intermediate point in a line of levels for the determination of altitudes over the face of a country.

The term has subsequently been used more generally to indicate something that can be used as a point of comparison in performance appraisals.

Benchmarking focuses on one or more activity measures, known as key performance indicators ("KPIs"). The value of each indicator achieved by a firm under scrutiny is compared to a benchmark value. The principal focus of benchmarking in studies of utility performance is cost.

Benchmarks are often developed using data on operations of firms involved in the same activity. Statistical methods are useful both for identifying the benchmarks themselves, and for evaluating a utility's performance relative to those benchmarks. An approach to benchmarking that prominently features statistics is called statistical benchmarking.

Various performance standards can be used in fashioning benchmarks. One sensible option is the average performance of utilities in the sample. Alternatives include the apparent best (or "frontier") performance, and the performance typical of top quartile performers.

3.2 External Business Conditions

3.2.1 Cost Drivers

An external business condition is a characteristic of the operating environment that a firm cannot control. Conditions affecting cost are sometimes called cost *drivers*. Differences in the costs of utilities depend on differences in these drivers as well as differences in their operating efficiency.

A utility's cost performance depends on the cost it achieves *given the cost drivers it faces*. Benchmarks must thus reflect cost drivers if they are to be used to evaluate a firm's performance fairly. The identification of relevant cost drivers and the assessment of their impacts are important tasks in a responsible benchmarking study.

3.2.2 Cost Functions

Economic theory is useful for identifying cost drivers. We begin by positing that the cost incurred by a company is the product of the *minimum achievable* cost and an efficiency factor. Under certain fairly reasonable assumptions, mathematical cost functions exist that relate the minimum cost of an enterprise to cost drivers in its service territory.

Two kinds of functions derived from cost theory are useful in benchmarking. One is the *total* cost function, in which the minimum total cost of an enterprise is determined by the prices of all production inputs, variables reflecting operating scale, and additional variables capturing miscellaneous other business conditions. Variables in the latter category are sometimes conveniently called "Z" variables. When the focus of benchmarking is a subset of total cost, *restricted* (aka "short-run") cost functions are useful. For example, the minimum achievable cost of opex depends on opex-specific input prices, output quantities, the amounts of *capital* inputs that a company owns, and other business conditions.

3.2.3 Capital Quantity and Scale Variables

The inclusion of capital quantity variables in an opex function is theoretically justified for several reasons.

- Different kinds of capital equipment have different operation and maintenance ("O&M") requirements. For example, network services opex is typically lower for lines that are underground than for lines overhead.
- It is generally more costly to operate and maintain capital facilities the more extensive they are.
- There are inconsistencies in the way utilities capitalize opex in their financial reports.
- Opportunities exist to substitute capital for O&M inputs. A firm may, for example, have unusually high opex because its capital is in an advanced stage of depreciation, so that it is using comparatively little capital.

It is difficult to measure capital quantities accurately in benchmarking studies. Multiple variables may be required. In addition to a variable indicating the length of lines one might, for example, need an indicator of the *age* of lines.

In a study of opex, variables reflecting the scale of capital inputs are highly pertinent. In the discussion that follows, we use the term "scale variables" to encompass all variables that indicate operating scale, including scale-related capital quantity variables.

3.2.4 Structure of Cost

The relationship of cost to scale and other business condition variables is sometimes called the "structure" of cost. The "elasticity" of cost with respect to a business condition variable is the percentage change in cost that results from a 1% change in the value of the variable. The relationship of cost to operating scale is particularly important and sometimes complex. Economic theory predicts that cost should be nondecreasing with respect to operating scale. This "monotonicity" condition implies that the elasticity of cost with respect to any scale variable should not be negative.

Economies of scale occur when cost has a tendency to grow more slowly than operating scale in the long run. The opportunity to realize incremental scale economies can vary with operating scale. A classic pattern is for the opportunities to be greatest for small companies, and to diminish as companies approach average size. At some point, companies reach a point of "minimum efficient scale" at which incremental scale economies are exhausted. Beyond this point, firms may experience incremental *dis*economies of scale, in which cost tends to grow more rapidly than scale.

3.3 Benchmarking Methods

In this section we discuss the two approaches to benchmarking used in the EI study: econometric modeling and indexing. The econometric approach is discussed first to establish a context for discussing indexing.

3.3.1 Econometric Benchmarking

Econometric Cost Research

The relationship between the cost of utilities and the business conditions they face can be estimated statistically. For example, suppose that for each utility h in year t, network services opex ($C_{h,t}$) has the following functional relationship to the number of customers served ($N_{h,t}$) and the length of lines ($L_{h,t}$).

$$\ln C_{h,t} = a_0 + a_1 \cdot \ln N_{h,t} + a_2 \cdot \ln L_{h,t} + inefficiency_{h,t} + error_{h,t}. \quad [1]$$

In this model, "parameters" like a_1, and a_2 (aka "coefficients") determine the impact of cost drivers on cost. A branch of statistics called "econometrics" has developed procedures for estimating the values of such parameters using historical data on the costs incurred by utilities and the business conditions they faced. Estimating model parameters in this way is called "regression".

The last two items in [1] are error terms. These reflect the fact that the variables included in the model are unlikely to fully explain the variation in cost of sampled utilities. Reasons for error might include mismeasurement of cost and/or the external business conditions, the exclusion of relevant business conditions from the model, the failure of the model to capture the true form of the relationship between cost and the included variables, and operating inefficiency. It is typically assumed that error terms are random variables with probability distributions determined by additional coefficients, such as mean and variance, which can be estimated.

Statistical theory is useful for evaluating the importance of business condition variables in the cost model. For example, a test can be constructed for the hypothesis that the parameter for each business condition variable equals zero. A variable can be deemed a statistically significant cost driver if this hypothesis can be rejected at a high (e.g., 90%) level of confidence. It is also possible to test the statistical significance of a *group* of variables. Parameter estimates are more likely to be significant when the sample size is large relative to the number of business condition variables included in the model.

Multicollinearity exists in a sample of data if some of the included cost drivers are highly correlated with each other. When present, the precision of the parameter estimates for the variables is reduced. This problem can make the affected parameters appear statistically insignificant even if these variables influence cost. In cost research, multicollinearity tends to be a particular problem between scale variables. A utility with a large number of customers, for instance, is also likely to have high peak demand, due in part to the tendency of residential customers to have peaked loads. The problem of multicollinearity can be reduced by increasing the size of the sample, though in severe cases one or more of the correlated variables may need to be dropped from the model.

Parameter estimates will be biased if relevant cost drivers that are correlated with the included drivers are left out of the model. It is therefore important to include as many potentially important business condition variables as possible in a cost benchmarking model.

Form of the Cost Model

Specific forms must be chosen for functions used in econometric research. Forms employed by scholars in cost research include the Cobb-Douglas (aka "double log") and translogarithmic (aka "translog") forms. The cost model in [1] has a Cobb-Douglas form. In this model, the dependent variable (cost) and the cost drivers (customers and line length) have all been logged. The log of cost rises linearly with the log of each cost driver. This makes the parameter corresponding to each cost driver equal to the elasticity of cost with respect to that variable. Elasticities are *constant* in the sense that they are the same for every value that the cost and cost driver variables might assume. This is restrictive, and may be inconsistent with the true form of the cost relationship we are trying to model. For example, it would not capture a tendency for the elasticity of cost with respect to the number of customers served to increase with the number of customers.

Here is an analogous model of translog form.

$$\begin{aligned}\ln C_{h,t} &= a_0 + a_1 \cdot \ln N_{h,t} + a_2 \cdot \ln L_{h,t} + a_3 \cdot \ln N_{h,t} \cdot \ln N_{h,t} \\ &+ a_4 \cdot \ln L_{h,t} \cdot \ln L_{h,t} + a_5 \cdot \ln L_{h,t} \cdot \ln N_{h,t} + \text{inefficiency}_{\text{h,t}} + \text{error}_{\text{h,t}}.\end{aligned} \qquad [2]$$

This differs from the double log form in the addition of quadratic terms (e.g., $\ln N_{h,t} \cdot \ln N_{h,t}$) and interaction terms (e.g., $\ln L_{h,t} \cdot \ln N_{h,t}$). These "second-order" terms permit cost elasticities to vary with the values of cost drivers. The elasticity of cost with respect to a scale variable may, for example, be higher for a large utility than for a small utility. Interaction terms like $\ln N_{h,t} \cdot \ln L_{h,t}$ permit the elasticity of cost with respect to one variable to depend on the value of another variable. In this case, the elasticity of cost with respect to growth in customers may depend on the miles of distribution line.

The translog form can accommodate a greater variety of the possible relationships between cost and the business condition variables. A disadvantage of the translog form, however, is that it can require the estimation of many more parameters compared to the Cobb-Douglas form. The number of second-order terms increases rapidly with the number of translogged variables. For example, while there is only one additional parameter for a single translogged variable, there are three for two, and five for three.

A large increase in the number of parameters to estimate reduces the precision of all parameter estimates. A typical manifestation of this precision is that the elasticities of cost with respect to some scale variables are negative for some companies. It is therefore common to limit the number of translogged variables. In cost research, translog treatment is typically considered only for input price and scale variables, even though it might be appropriate for other variables as well. When sample size is limited, it may be impossible to accurately estimate a cost model of translog form. The need for a translog specification is commonly assessed with a group test of the statistical significance of second-order terms.

Estimation Procedures

Various estimation procedures (aka "estimators") are used in econometric research. The appropriateness of each depends on the assumed distribution of the error term. The most widely known estimator, ordinary least squares, is appropriate if the distribution of the errors is believed to be simple. Alternative estimators are more appropriate under assumptions of more complicated error distributions. In statistical cost research, error term complications include heteroskedasticity and autocorrelation. Autocorrelation occurs when past values of the error term are good predictors of its future values. Heteroskedasticity occurs when error terms for different companies have different variances. For example, large firms are often found to have larger variances than small firms. Least squares estimators have been developed that address these complications.

Another complication in econometric cost model estimation is that the estimated error term includes an efficiency factor. This factor represents the distance between a firm's actual cost and the minimum achievable cost frontier. In the long term, a firm can never maintain cost levels below this frontier (for technological, legal, and other reasons), so there is a ceiling on the company's efficiency. However, there is no theoretical floor for the company's inefficiency. In short, a company cannot be more efficient than the frontier in the long run but can be, in theory, infinitely inefficient. Statistically, this means that the inefficiency term (and therefore the error term) has an asymmetrical distribution.

An approach to estimation called stochastic frontier analysis ("SFA") attempts to estimate this minimum cost frontier and use the results to predict firms' inefficiency. SFA models have stricter statistical assumptions than most least squares models and generally require additional data for accurate estimation. In data sets of limited size, it may be possible to estimate a model using a least squares approach but not using SFA. Another limitation of SFA is that most routines available in standard econometric software packages do not control for autocorrelation or heteroskedasticity in the sample. Should an SFA model fail a diagnostic test for either of these conditions, there is little recourse available.

Cost Predictions

We can use a cost model fitted with econometric parameter estimates to predict a company's cost given local business conditions. These predictions are econometric benchmarks. Suppose, for example, that we wish to benchmark the cost of Endeavour. We can predict Endeavour's cost in period *t* using, for example, the following Cobb-Douglas model fitted with least squares parameter estimates.

$$\hat{C}_{Endeavour,t} = \hat{a}_0 + \hat{a}_1 \cdot \ln N_{Endeavourt} + \hat{a}_2 \cdot \ln L_{Endeavour,t}. \quad [3]$$

Here $\hat{C}_{Endeavour,t}$ denotes the predicted cost of the company, $N_{Endeavour,t}$ is the number of customers it serves, and $L_{Endeavour,t}$ measures its line length. The $\hat{a}_0$, $\hat{a}_1$, and $\hat{a}_2$ terms are parameter estimates. Performance might then be measured using a formula such as

$$Performance = \left(C_{Endeavour,t} \middle/ \hat{C}_{Endeavour,t} \right).$$

Performance Standards

The estimation procedure influences the performance standard embodied in the benchmarking exercise. For example, SFA generates scores indicating the estimated distance from an efficiency frontier. SFA appraisals thus reflect a *frontier* standard of operating efficiency. Alternatively, predictions from an econometric model estimated by a least squares method reflect a *sample average* efficiency standard.

The results obtained from either kind of estimator can also, however, be evaluated according to alternative standards. For example, in SFA the efficiency score of each sampled utility can be compared to the average of the top quartile of efficiency scores, or even to all scores in the sample. Similarly, when a least squares estimator is used, each score can be compared to the most efficient utility's score, or to an average of the top quartile of scores.

An inherent challenge in estimating distance from the efficiency frontier is that firms can, in the short run, incur costs that are considerably below the level that is sustainable in the long run. For example, a distributor may defer periodic expenses such as tree trimming and other maintenance work. In the short run, such a firm may appear to be a top cost performer, but over the long term such behavior would lead to service quality deterioration. Unusually favorable operating conditions may also cause short-run costs to fall below the long-run frontier. Another problem with the use of a frontier performance standard is that it is unusually sensitive to irregularities in the data. More stable results can be achieved by taking an average of performance comparisons over several years.

Accuracy of Benchmarking Results

A cost benchmark is our best *single* guess of a company's cost given the business conditions it faces. In other words, it is a "point" estimate. Such a prediction may differ from the true benchmark for several reasons.

For example, the variables used in the model may be imperfect measures of the company's cost, or of the business conditions faced by the company. Relevant business conditions may also be left out of the model entirely. Another potential source of error is the functional form of the model (e.g., if a model assumes a relationship is linear when it is not). The sample used may also be too small to produce accurate estimates, especially if the estimation of a large number of parameters is attempted. Any of these factors can contribute to inaccuracies in the benchmarks produced by the model.

Some of these sources of error may not be detectable based on the model results alone, and therefore must be guarded against through the careful application of economic theory and sector-specific knowledge. However, econometric methods do provide ways to judge the likely accuracy of model results. A common approach is to construct confidence intervals around point estimates, indicating the ranges in which the true values are expected to fall with a given level of confidence (e.g., 90%), given the data and model assumptions.

A confidence interval is wider the greater is the uncertainty about the true benchmark value. In general, confidence intervals are wider to the extent that:

- The model is not successful in explaining the variation in cost in the historical data used in its development
- The size of the sample is small
- The number of cost drivers considered is large
- The business conditions of the sampled companies vary little
- The business conditions of the subject utility are dissimilar to those of the typical firm in the sample.

These results suggest that econometric benchmarking will in general be more accurate to the extent that it is based on a large, varied sample of good operating data. In cost research most variation in the values of business conditions occurs *between* companies rather than *within* companies over time. It is thus especially desirable for the sample to include data for numerous companies. When the sample is small, it will be difficult to identify all of the relevant cost drivers and accommodate the appropriate functional form. In benchmarking a firm facing unusual operating conditions it is desirable to include in the sample data for several firms facing similar conditions.

3.3.2 Benchmarking Indexes

Another approach to benchmarking involves the construction of indexes. Benchmarking indexes are commonly employed by utilities in internal performance reviews, and are also used sometimes in regulation. We begin our discussion with a review of index basics, and then consider unit cost and productivity indexes in turn.

Index Basics

An index is defined as "a ratio or other number derived from a series of observations and used as an indicator or measure (as of a condition, property, or phenomenon)."[2] In the context of benchmarking, indexing involves calculating the ratio between a KPI value of a focal utility and the corresponding value for a sample of utilities. The companies in such a sample are known as the peer group.

Economic indexes can be designed to summarize multiple comparisons. Such "multidimensional" indexes involve weighted averages of the comparisons.[3] To better appreciate the advantages of multidimensional indexes in cost benchmarking, recall from our discussion in Section 3.3 that multiple variables are often needed to accurately measure utility operating scale. These variables can have different cost impacts even if all are worth considering. Suppose, for example, that we are benchmarking the opex of a DNSP. If we separately calculate the company's cost per line km and per customer we might come up with two very different assessments. Therefore, it may be desirable in this case to consider line length as well the number of customers served. The relative importance of different variables within the index is captured by the weights applied to

[2] Webster's Third New International Dictionary of the English Language Unabridged, Volume 2, p. 1148. (Chicago: G. and C. Merriam and Co. 1966).

[3] Consumer price indexes are familiar examples. These summarize the inflation (year to year comparisons) in the prices of hundreds of goods and services.

them. For example, an index of operating scale could be constructed as a weighted average of the individual line length and customer number comparisons.

Scale index weights should reflect the relative importance of scale variables as cost drivers. It is thus sensible to weight each scale variable according to its share of the sum of the cost elasticities of all such variables. These elasticities can be estimated econometrically.

Unit Cost Indexes

A simple comparison of utilities' costs reveals little about their performance because there may be large differences in the cost drivers they face. In index-based benchmarking, it is therefore common to use the ratio of cost to one or more important cost drivers as the KPI. Variation in the operating scale of utilities is usually the greatest source of difference in their costs, so it makes sense to utilize ratios of cost to operating scale when making comparisons. Such "unit cost" measures provide a control for differences in scale, permitting the inclusion of companies with more varied operating scales in the peer group.

However, unit cost indexes have a number of limitations as benchmarking tools. The control they provide for differences in operating scale is crude, since they don't capture the varying potential for companies of different size to realize scale economies. Neither do they control for differences in any of the numerous other cost drivers that are known to vary between utilities. The accuracy of unit cost benchmarking thus depends on the extent to which the additional cost pressures faced by the peer group are similar to those facing the subject utility. It is desirable to have numerous peers to smooth out eccentricities in the operating performance or data of individual peers.

Productivity Indexes

A productivity index is the ratio of an output quantity index to an input quantity index. It is, essentially, a unit cost index in which the cost measure has been adjusted to reflect differences in the input prices faced by utilities. Thus, using a productivity index to benchmark cost entails comparing the portion of unit cost that is not due to variation in input prices.

Productivity indexes can be designed to measure only the trend in a given utility's productivity, or differences in the productivity levels of multiple utilities at a given point in time. *Multilateral* productivity indexes have also been developed that measure productivity levels *and* trends.[4] Productivity indexes also vary in the scope of inputs considered. A "total factor" productivity ("TFP") index considers all inputs. "Partial factor" productivity ("PFP") indexes consider subsets of inputs, such as operation and maintenance ("O&M") or capital inputs.

The input quantity used to compute a productivity index is often measured as the ratio of cost to an input price index. Since opex includes both labor and material and service ("M&S") expenses, an opex input price index should be a cost-weighted average of labor and M&S price subindexes. For maximum accuracy, the weights on these subindexes should be utility-specific. In considering the input prices faced by a utility in a large urban area, for instance, the weight on the labor price should typically be higher.

In computing PFP indexes, it is not unusual to use the same scale index that is used to calculate TFP. However, scale indexes can also be customized to the input subgroup. For example, a scale index for O&M productivity can reflect cost elasticity estimates obtained from an *opex* (rather than a total) cost function.

Productivity indexes are more accurate than unit cost indexes as benchmarking tools because they control for differences between utilities in input prices as well as operating scale. They nonetheless have major limitations as benchmarking tools. Like unit cost indexes, they do not control for differences in the opportunity of utilities to realize scale economies. Neither do they control for differences between companies in the values of Z variables. It follows that the selection of a similar peer group is of great

[4] Caves, D. W, L.R. Christensen, and W.E. Diewert (1982), "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers," *The Economic Journal* 92, 73-86.

importance to the accuracy of a benchmarking study based on productivity indexes. Once again, it is desirable for there to be numerous similarly situated peers.

The measurement of capital quantities has been an area of controversy in total factor productivity research. One option that has been widely used is a "service price" approach, in which the cost of capital is calculated as the product of capital price and quantity indexes. The capital price index reflects the cost of owning a unit of capital. The capital quantity index represents the total units of capital, and is calculated using data on gross plant additions and standardized depreciation formulas. PEG Research has routinely used the service price approach to capital quantity measurement in its TFP research, including its work for the OEB.

An alternative approach to capital quantity measurement is to use physical asset measures. These are equivalent to the "scale-related" capital quantity variables discussed in Section 3.2 above. In the context of power distribution TFP, examples of such variables are line length and substation and line transformer capacity.

Arguments favoring the physical asset approach to capital quantity include the greater availability of physical asset measures, and the assertion that many types of assets used by utilities yield a fairly constant stream of service until they are retired. On the other hand, many assets require higher opex as they age, which undermines the idea that capital inputs provide a constant service stream. Furthermore, physical asset measures do not reflect efforts by companies to contain capital cost by deferring replacement of aging assets or reducing unit capex (e.g., capex per route mile).

Denny, Fuss, and Waverman (1982) developed an econometric method for projecting productivity growth.[5] This can be used to forecast future growth given forecasts of changing business conditions, or to backcast the past growth that would

---

[5] Denny, Michael, Melvyn A. Fuss and Leonard Waverman (1981), "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," in Thomas Cowing and Rodney Stevenson, eds., Productivity Measurement in Regulated Industries, (Academic Press, New York) pages 172-218.

have been realized by typical sampled utilities under historical business conditions. The chief component of the productivity growth forecast/backcast is the trend variable parameter estimate. This is adjusted for changes in the values of business condition variables, and includes consideration of the potential to realize scale economies.

Performance Standards

A productivity index that makes a comparison to the full sample embodies a sample average standard of performance. Alternative standards can also be implemented. We can, for example, compare the productivity of each utility to the highest productivity achieved by sampled utilities, or to the average of the top quartile of productivity scores.

Frontier performance comparisons using indexes are fraught with many of the same limitations as occur in the context of econometric modeling. The apparent best productivity performance may not reflect a sustainable frontier if it results from deferred maintenance, unusual business conditions, or data irregularities.

3.4 Benchmarking Standards

The choice of a performance standard is a key element of benchmarking strategy. We consider here three alternative standards: the sample average, competitive market, and frontier standard.

3.4.1 Industry Average Standard

Under a sample average performance standard, benchmarking focuses on how a company's performance compares to the average amongst sampled utilities. We have seen that several of the benchmarking methods discussed previously lend themselves to this method. A conventional productivity index, for instance, is designed to compare the productivity of a subject utility to that of the sample mean.

The average performance standard is especially suitable when benchmarking is used to screen companies for more detailed prudence inquiry. Companies can be deemed eligible for review if benchmarking suggests a performance that is considerably below the industry norm.

The average performance standard can also be used to reward good cost performance. In a price control plan, for example, a company may be permitted a superior return if benchmarking suggests a performance markedly superior to the norm. Rewards like this can materially improve performance incentives.

3.4.2 Competitive Market Standard

Under a competitive market standard the focus of benchmarking is the typical performance of firms in a competitive industry. The intuition for this approach has some appeal. After all, the standard argument for utility regulation is that competition is absent and that a forced restructuring of the industry is unworkable. It makes sense, then, for utility regulation to have as its goal the simulation of desirable aspects of competition.[6]

In a competitive market, prices reflect the interaction of supply and demand conditions at the industry level. Since individual firms can't influence the prices at which they sell their products, they have strong incentives to improve their performance. In the long run, firms with especially bad performance leave the industry.

These attributes of competitive industries encourage the view that *all* firms in such industries are efficient. However, the reality is that it is the *industry as a whole* that earns a competitive rate of return, while at any point in time the efficiency of individual firms varies considerably. More efficient firms achieve superior rates of return, and less efficient firms earn inferior returns. The firm of typical efficiency may operate at a considerable distance from the efficiency frontier.

Benchmarking studies of firms in competitive industries are useful for assessing the extent to which typical firms are inefficient. Studies that employ a frontier standard

[6] Competitive markets also have undesirable aspects that we may not wish to replicate. These include, in some cases, a high degree of volatility.

are especially relevant. On behalf of two British power distributors, PEG conducted surveys of frontier benchmarking studies in two competitive sectors: banking and farming. Results are reported in Tables 1 and 2. In some cases more than one benchmarking method was used in the study. We present in these cases the results from each method.

Our survey on banking efficiency using frontier methods covers Greek, Turkish, European and U.S. banks. The studies for banks report average efficiency levels ranging from 30% to 92%. The studies for farms report average efficiency scores ranging from 76% to 95%.[7]

[7] Note that the efficiency studies in the farming sector consider only technical efficiency, not all possible sources of inefficiency.

Table 1. Survey of Efficiency Studies of Banking Firms

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

Table 2. Survey of Efficiency Studies of Farming Firms

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

We conclude from this survey that the measured average efficiency level of firms has not been at, or even close to, the estimated frontier in either of these two competitive industries. If these results are representative of competitive industries as a whole, we may conclude that to simulate competitive markets the relevant cost performance standard is one some distance away from the short-run frontier.

A noteworthy disadvantage of the competitive market efficiency standard is the difficulty of making it operational. Unlike the sample average standard, there is no straightforward way to use data from a non-competitive industry such as power distribution to implement the standard. One possible approach is to assume that the competitive market standard is a certain percent higher than the average industry standard, or a certain percent lower than the standard represented by the short-run frontier. For example, a sensible approximation to the competitive market standard might be a level of performance at the lower edge of the top quartile.

3.4.3 Frontier Standard

Under a frontier standard, companies are judged not by their position relative to the sample norm, but instead by their distance from the ostensible efficiency frontier. This paradigm can be criticized on several grounds.

One is that the frontier that is measurable is the short-run frontier, which may not be sustainable in the long run. If this is the case, it is unreasonable to expect utilities to operate permanently at this level. Indeed, companies that lie on such a frontier are likely to need additional revenue in the future to ensure quality service.

In addition, accurately measuring the distance of firms from the short-run frontier is challenging, since the process is quite sensitive to data irregularities. Potential 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 varies with the benchmarking method chosen. For example, some methods such as productivity indexing are more prone to measurement errors and hence tend to exaggerate the distance of companies from the frontier.

A third concern about the frontier standard is its fairness. As we have seen, superior cost performers in competitive industries are entitled to superior returns. If firms must operate at 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 in advanced industrial countries. For example, the ability of labor unions to offset the potential monopsony power of employers is one of the major arguments ventured for legalizing their activities.

4. Benchmarking Network Services Opex

In this section we discuss salient considerations in the benchmarking of network services opex. We first consider important aspects of the provision of network services, and then discuss data issues and international benchmarking.

4.1 The Provision of Network Services

DNSPs deliver power to consumers. Prior to delivery, most of this power flows through a transmission system, which carries it at high voltages in order to reduce line losses. Receipt of power from the transmission system commonly occurs at substations, where voltage is reduced from transmission to distribution levels. Power is carried away from the substation on lines at primary voltage, and typically reduced to the secondary voltage at which consumption occurs by small transformers located close to customer premises. [8] Distribution lines may be placed overhead on poles or underground in conduits; structures of both kinds are likely to carry more circuits in urban areas than in rural areas.

4.1.1 Local Delivery Cost

The cost of local delivery service comprises costs of plant ownership, labor, materials, and services. Capital inputs typically account for between 45 and 60 percent of the total cost of network service. The cost shares of labor and materials and services vary greatly between utilities due to differences in labor prices and the extent of outsourcing.

Prices paid for labor, capital, and other inputs are important drivers of power distribution cost. Prices tend to rise over time, and their levels differ by geographic region. We developed opex input price indexes in our work for the AER that were levelized so that cost varied by region. We found in our research that, on average, O&M

[8] Some large volume customers perform their own voltage step downs. At the extreme, they may take delivery of power from the transmission grid and bypass the distribution system entirely.

input prices were about 2.6% higher in New South Wales than in other NEM states over the 1998-2013 period. In addition, we found that input prices in the Sydney area were higher than in rural NSW.

Some distribution expenditures are periodic, in the sense that they do not have to be undertaken at the same level each year. Overhead line maintenance activities such as tree trimming are a salient example. Distributors may in a given year spend far less on line maintenance than is needed to ensure reliable service in the long run.

4.1.2 Operating Scale

The operating scale of a distributor is sometimes narrowly defined as the units used to compute bills. The three most common billing units used by distributors are delivery volume, peak load, and the number of customers served. These variables do not address all operating scale dimensions, however.   For example, there is no charge for the distance over which power is carried from points of receipt to the customer.

The scales of capital inputs are also important opex drivers. Distribution line lengths and substation capacity are especially relevant. Distributors must operate and maintain these facilities whether or not their sizing is optimal.

Line lengths may be measured with respect to structures or circuits. Circuit lengths are not necessarily more pertinent than structure lengths (aka route lengths) from an opex impact perspective. For example, O&M tasks such as vegetation management are not three times more expensive on a route with one circuit than they are on a route with three circuits. The relative importance of route length and circuit length is an empirical issue that can be evaluated econometrically.

The choice of a line length variable can affect benchmarking results. In Australia's National Energy Market ("NEM"), for instance, the ratio of route miles to circuit miles exceeds 0.90 for Aurora, Essential Energy, Powercor, and SA Power Networks, but is less than 0.60 for United Energy and Jemena.   A circuit length variable may therefore reduce efficiency scores for distributors with more rural service territories.

Economies of scale are possible in power distribution. Opportunities to realize economies of scale (which are reflected in the cost elasticities with respect to scale variables) may vary with the size of utilities. The character of scale economies is an empirical issue that can be addressed with econometric methods. This is conventionally done by considering the statistical significance of the second order terms of a translog cost function. If these are significant, the relationship of cost to output is nonlinear and cost elasticities vary with firm operating scale.

Distributors in Australia's NEM have varied operating scales. In 2013, for instance, the smallest distributor, ActewAGL, had only 177,255 customers while the largest, Ausgrid, had 1.6 million. Only one other distributor, Energex, had a scale similar to Ausgrid's. In contrast, four companies had low customer numbers, and six had customer numbers in an intermediate range. As for route km, Citipower (3,113) had the fewest while Essential Energy (180,741) had the most. Only one other distributor (Ergon Energy) had route kms similar to those of Essential Energy. In contrast, four companies had low line kms, and five had line kms of intermediate length. As a result of these differences, for two of the three NSW distributors it is difficult to establish productivity peer groups or econometric samples with several peers when only Australian distributors are sampled.

4.1.3 Network Service Packages

Distributors vary in the package of network services they provide. These differences can have a sizable impact on the cost of service. Here are some prominent examples.

- DNSPs vary in their involvement in the transformation of voltage from the transmission to the primary distribution level. Where transmission and distribution ("T&D") services are provided by separate companies, as in Australia, Britain, New Zealand, and Canada's populous Ontario province, policymakers have typically decided which kind of company provides this service. In Australia, DNSPs typically own substations that step down most voltage to the primary level but Aurora does not.

- Where T&D are provided by the same company, as is typical in North America, the issue is how these services are *categorized*. In the United States, utilities provide detailed cost information to the Federal Energy Regulatory Commission and reports must conform to a Uniform System of Accounts ("USA"). The USA directs utilities to classify substations that step down most voltage to the primary level as distribution facilities.
- Some T&D systems have lines with voltages intermediate between the high voltages used for long distance transmission (e.g., 220 kV) and those used for primary distribution (e.g., 22 kV). In regions with separate transmission and distribution utilities, these intermediate voltage lines are sometimes owned by the transmission utility and sometimes by distributors. Many British distributors own and operate systems of 100+ kilovolt ("kV") lines. In Australia, distributors in the ACT, NSW and Queensland own such facilities but those in South Australia, Tasmania, and Victoria do not.
- In regions where utilities are engaged in both T&D, the classification of intermediate voltage lines can vary between utilities. In the United States, power distribution systems are defined as beginning at the "entrance" to a consuming area. By this definition, subtransmission lines and substations reducing voltage to the subtransmission level are sometimes categorized as distribution facilities in large urban areas, because they are past the entrance to the city and make only local power deliveries. In rural areas, these same facilities might be deemed transmission assets since they carry power to the entrance of scattered towns.
- Where distributors step down the voltage received from the transmission system to primary level, there can be differences in the voltage stepdowns that distributors perform at their substations. This depends on the voltage

of transmission lines from which power is received. Some distributors may step down voltage only from intermediate levels, whereas others step down voltages from much higher levels.

4.1.4 Other Network Characteristics

Power distribution networks vary in a number of other ways that can affect cost.

- Systems vary with respect to customer density. Density, which is commonly measured as the ratio of customers to line length, is highest in urban locations and lowest in sparsely populated rural areas. The impact of density on network services opex may be nonlinear. Low-density distribution systems tend to be more costly than systems of intermediate density. However, high-density systems may also pose special operating challenges, such as the complications of performing maintenance work where lines are beneath busy city streets. If the impact of density is U-shaped, studies that fail to address this may favor suburban distributors like United and Jemena.

  The density of distribution systems in Australia's NEM varies widely. In 2013, density was lowest for Ergon and Essential Energy and highest for ActewAGL and Citipower. Ausgrid and Energex each serve a large urban area and a surrounding rural area. They may thus be saddled with two kinds of high cost densities.

- Undergrounding generally raises the *total* cost of network services, but can lower network service *opex* due to less frequent line maintenance. Undergrounding is most common in densely settled urban areas (e.g., central business districts) and least common in rural areas. In suburban areas, undergrounding has been increasingly common in recent years and also depends on public policies. For example, in metropolitan Sydney undergrounding is more common in the comparatively new suburbs served by Endeavour than in the comparatively old suburbs served by Ausgrid.

The extent of system undergrounding can be measured by the share of circuit length that is underground. However, this measure tends to underestimate overheading in service territories that include a mix of urban and rural areas, like those of Ausgrid and Energex, to the extent that there is a greater circuit count per structure in urban areas. An alternative measure of undergrounding is the share of undergrounded lines in the total *value* of lines. The relative importance of these undergrounding metrics as cost drivers can be appraised econometrically. The extent of undergrounding in Australia's NEM varies widely: in 2013 the share of circuit length undergrounded was highest for ActewAGL (53%) and Citipower (48%) and lowest for Essential Energy (3%), Ergon Energy (5%), and Powercor (7%).

- Automated metering infrastructure ("AMI") can lower network services opex, as well as metering expenses, by reducing the cost of addressing faults.
- The shape of distribution systems is influenced by land forms. For example, distribution lines often go around lakes, bays, and other large water bodies.
- Higher opex is generally needed for greater reliability. The cost impact of quality is thus a valid issue in distribution benchmarking. However, estimating this impact is challenging. One reason is that cost tends to be higher in years when reliability dips due to severe weather or other unfavorable events. Another is that reliability influences opex less when levels of undergrounding, grid reinforcement, and AMI are higher. Despite its importance, empirical research on the relationship between cost and reliability is not well advanced.

  There is considerable variation in the reliability of services provided by DNSPs in Australia's NEM. Reliability has generally been greater for urban distributors like Citipower and ActewAGL than for rural distributors like Essential Energy and Ergon.

4.1.5 Other Cost Drivers

Cost research by PEG and others has identified a range of additional business conditions that drive network services opex.

- Opex is generally lower the younger is a distribution system.
- Opex is typically higher the greater is service territory forestation, at least when a sizable share of lines are overhead. An obvious reason is the greater need for tree-trimming.

  The extent of forestation in a service territory can be difficult to measure accurately. Forestation may be more extensive in some urban areas than in surrounding rural territory due to parks and irrigation. We have found in our research that rainfall can serve as a useful proxy, since precipitation tends to encourage vegetation growth in rural and urban areas alike. In our work for the AER, we gathered rainfall data for distributors in Australia's NEM, and found considerable variation. Rainfall was lowest for Powercor, SA Power Networks, and Jemena and highest for Ausgrid and Energex. In econometric models that we estimated using only Australian data, the rainfall variable had a highly significant, positive parameter estimate, suggesting a positive relationship between forestation and opex.
- Opex is generally higher in areas where severe weather is more frequent.
- Opex is higher the more difficult it is to access distribution facilities. ActewAGL, for example, faces special access challenges due to backyard reticulation and the AER has acknowledged that this raises cost.
- Opex may be elevated where distribution service must be provided to hilly areas.
- Another condition affecting opex is the number customers receiving gas distribution service, since providing both electric and gas services presents opportunities for the realization of scope economies. Only one Australian

DNSP (ActewAGL) is also a gas distributor. However, combined gas and electric companies are fairly common in the U.S.

- Opex is affected by policies of state and local governments. For example, distributors in Victoria have since 2011 incurred higher opex to conform to state policies enacted in response to deadly bushfires. However, all jurisdictions in the NEM *except* Victoria have enacted the Work Health and Safety Act and Work Health and Safety Regulations, which could increase costs in these areas. Many municipalities in Victoria undertake vegetation management on public property, which may lessen the burden on DNSPs.

4.2 Data Issues

4.2.1 Reporting Inconsistencies

Research has identified numerous inconsistencies in how DNSPs report operating data. Inconsistencies in reporting cost tend to be especially marked where utilities have some discretion due to lax reporting guidelines, and/or where the itemization of cost is particularly arbitrary. Inconsistencies in the capitalization of O&M expenses can also create problems; the AER has acknowledged that this poses a particular challenge to benchmarking ActewAGL.

Another issue is the breakdown between direct expenses and corporate and business support expenses. The latter category comprises expenses that are difficult to attribute to specific lines of business. Inconsistencies can also arise in the allocation of direct expenses between distributor activities.

Inconsistencies have been encountered in the data gathered for several of the Z variables that the AER requested. EI acknowledges this in its opex report. PEG considered several of these variables in its benchmarking work for the AER, and none had statistically significant and plausible parameter estimates.

4.2.2 Missing Data

Benchmarking can be complicated by the unavailability of important data. A major problem in many countries is the lack of good capital cost data. Good data that can be used to calculate standardized capital costs and quantities are not available for most countries of the world.

4.2.3 Sample Size and Variability

Development of an econometric cost model that properly reflects the impact of external business conditions faced by DNSPs requires a large, varied data set. Such data sets are unavailable in many countries due to a combination of few distributors and few years of high quality, standardized data. In Australia, standardized data are available for only thirteen distributors for the 2006-2013 sample period. A sample with only 104 observations greatly limits the sophistication of econometric cost models that can be developed: it may be impossible to obtain accurate estimates of the cost impact of certain business conditions, or to properly model scale and scope economies.

4.2.4 Econometric Research on Distribution Cost Drivers

A cost model we developed several years ago using U.S. data illustrates the value of a large and varied data set in cost model estimation. The principle source of data was the Federal Energy Regulatory Commission ("FERC") Form 1. The opex considered included that for customer services as well as for local delivery services.

The study was prepared for testimony in a Portland General Electric rate case. Data for 105 companies over the period 1995-2008 were used. The resulting sample size of 1,446 observations is more than 10 times that which is currently available for Australia. The large size of the sample permitted inclusion of numerous business conditions and a flexible functional form.

The econometric results are reported in Table 3. First of all, it is noteworthy that all the parameter estimates for the first-order terms are statistically significant and plausible. The model also has high explanatory power, though this is fairly common in

Table 3

Econometric Model of Distribution, Customer Care, and Administrative O&M Expenses

VARIABLE KEY

WL = Labor Price
N = Number of Customers
VRC = Residential & Commercial Delivery Volume
DSM = Share of CS&I in Distribution and Customer Care O&M
POH = Percent of Distribution Plant Overhead
NG = Number of Gas Customers
G = Net Generation
HDD = Average Heating Degree Days
P = Average Precipitation
Trend = Time Trend

COST DRIVER	PARAMETER ESTIMATE	T-STATISTIC	P-VALUE
WL	0.360	108.99	0.000
WLWL	0.093	2.41	0.016
WLN	-0.009	-0.69	0.489
WLVRC	-0.012	-1.03	0.305
N	0.817	31.06	0.000
NN	0.381	2.88	0.004
NVRC	-0.387	-3.12	0.002
VRC	0.128	4.80	0.000
VRCVRC	0.377	3.17	0.002

COST DRIVER	PARAMETER ESTIMATE	T-STATISTIC	P-VALUE
DSM	0.028	6.742	0.000
POH	0.144	7.732	0.000
NG	-0.003	-2.609	0.009
G	0.059	7.152	0.000
HDD	0.009	10.075	0.000
P	0.019	1.848	0.065
Trend	-0.015	-13.893	0.000
Constant	12.300	918.586	0.000
System Rbar-Squared	0.969		
Sample Period	1995-2008		
Number of Observations	1446		

utility cost research because the dominant source of variation in cost is operating scale, and this can be adequately (if not perfectly) measured.

Input price and scale variables in the model were translogged. Parameter estimates for the second-order terms (e.g., customers x customers) were highly significant as a group, indicating that the relationship between cost and the scale variables was not best approximated by a Cobb-Douglas functional form.

Results for the Z variable parameter estimates are also informative.

- The positive sign for the system overheading parameter suggests that overheading raises opex.
- The positive sign for the precipitation parameter suggests that opex is generally higher when the service territory is more heavily forested.
- The positive sign for the heating degree days parameter suggests that cost tends to be higher in areas with colder winters.
- The negative sign for the gas customers parameter suggests material economies of scope from joint provision of service.
- The negative 1.5% value for the trend variable parameter suggests a fairly brisk pace of O&M productivity growth.

These findings have important implications for the benchmarking of power distribution in Australia. Most notably, benchmarking studies that do not consciously control for differences between utilities in these business conditions may not be very accurate. Also, distribution cost models may need flexible functional forms. Benchmarking studies prepared using only Australian data may not confirm these results, however, since statistical significance can be difficult to achieve when a sample lacks sufficient size, variation, and data quality.

4.3 International Benchmarking Challenges

Certain complications are especially common in international benchmarking. These include, especially

- Difficulties in comparing input prices
- Differences in operating scale
- Differences in the service packages DNSPs provide
- Inconsistencies in cost reporting
- Missing data
- Availability of itemized opex data that can be used to construct a definition of cost that is analogous to network services opex.

5. Benchmarking in Regulation

This section of the report discusses our review of precedents for use of statistical benchmarking to regulate energy utilities. The review covered four countries in the English speaking world: Britain, Canada, New Zealand, and the United States. We begin this section by discussing the basic role of benchmarking in regulation. There follow discussions of the focus of benchmarking, performance standards, the use of benchmarking results in ratemaking, benchmarking methods, and the benchmarking study review process. Lengthier discussions of the precedents may be found in the Appendix.

5.1 Basic Role

The role of statistical benchmarking in regulation varies widely in the English speaking world. Benchmarking is currently used to establish revenue requirements in several jurisdictions overseas that include Britain and the Canadian province of Ontario. In the United States and most of Canada, on the other hand, benchmarking usually plays no role in ratemaking. Benchmarking studies are rarely initiated by North American regulators. Studies filed in rate cases are typically volunteered by utilities or consumer groups eager to make a point about good or bad cost performance, respectively.

There is no discernible trend in the use of benchmarking in ratemaking in the surveyed countries. Several jurisdictions have, like Ontario, been using benchmarking for years. On the other hand, price control regulation has recently been implemented in Alberta without Commission use of benchmarking. Benchmarking's use by New Zealand's Commerce Commission has been barred by law after controversies there.

5.2 Focus of Benchmarking

The focus of regulatory benchmarking depends greatly on the availability of data. In the United States, Canada, and New Zealand, larger data sets have been gathered and capital cost data are sufficiently standardized to do total cost benchmarking. Where data are more limited, the focus is more likely to be on opex and/or capital expenditures ("capex"). In Britain, considerable attention is now paid to the sum of opex and capex (dubbed "totex").

5.3 Performance Standards

Various performance standards have been used in efficiency assessments. They include the sample average, top quartile, and competitive market (edge of the top quartile) standards. In Britain, top quartile performers are eligible for a superior rate of return. Comparisons to the performance frontier in regulation are rare.

5.4 Use of Benchmarking Results in Ratemaking

Regulators vary greatly in their use of benchmarking evidence. At one extreme, evidence is ignored or any adjustments to revenue that are made on the basis of benchmarking are not explicit. This is the most common outcome in US regulation.

In the middle of the spectrum are a range of small adjustments. For example, only a fraction of the difference between a company's cost and the benchmark may be deemed eligible for disallowance. Indicated disallowances may be implemented gradually over five to ten years. Benchmarking may only be used to indicate the need for closer scrutiny, or considered implicitly as one of several inputs to a revenue determination. Disallowances have rarely been equal to the full amount by which cost deviated from the chosen benchmark, and implemented in the first price control year.

Regulators have often been cautious in their use of benchmarking results until they have gained many years of experience with benchmarking. Disallowances by British regulators have varied in size and have trended downward over time. Most recently, disallowances have been between 0 - 11% - the top of this range was applied in the knowledge that the company in question had been subject to comparative benchmarking techniques for some time.

An approach to price control design originated in North America (and is also used in New Zealand) whereby the X factor in the price (or revenue) cap index reflects the trend in the TFP of the industry rather than a company-specific cost forecast. Where this is approach has been used, benchmarking has guided selection of an X factor adjustment called the “stretch factor”.   Stretch factors typically assume a value between 0.2% and 0.5% and rarely exceed 1%.

Ontario is a leading practitioner of this “North American” approach to price control design.   The majority of power distributors there operate under an “incentive regulation mechanism” in which the stretch factor is informed by the results of a cost benchmarking study. Under the latest mechanism, participating distributors are assigned to one of five performance groups. Those in the best performing group have a stretch factor of zero, while those in the worst performing group have a stretch factor of 0.6%. The penalty for being in the worst performance group is therefore to have revenue grow 0.6% more slowly each year. There is no initial disallowance.

5.5 Benchmarking Methods

Various benchmarking methods are used in regulation. Unit cost and productivity indexing and econometric modelling are the most common methods used in the English speaking world.[9] Revenue adjustments are sometimes based on averages of efficiency appraisals using several methods. Most studies are based on national data, but transnational samples have been used in several studies.

Use of econometrics and the sophistication of econometric models has often been limited by the size of national samples and the lack of consistent national data for a long sample period. In Britain, for example, the sample size is small and econometric modelling is crude. However, extensive work is done prior to model estimation to normalize data. More sophisticated cost modelling has been encouraged in the United States by the unusually large and diverse set of standardized data available. These advantages of US data have encouraged their use in transnational benchmarking studies.

[9] An alternative benchmarking method called data envelopment analysis is (“DEA”) popular in Europe. This is similar to productivity indexing in considering quantities of outputs and inputs and often involves a second-stage regression on a range of additional business conditions.

Large data sets also encourage use of custom (rather than full national sample) peer groups in unit cost comparisons. The OEB, for example, used both unit cost indexes and econometric modelling to establish stretch factors in one benchmarking study. Data for more than seventy utilities made it possible to develop numerous peer groups reflecting special business conditions.[10] PEG often presents an econometric model and a custom unit cost peer group in its benchmarking studies for US clients.[11]

5.6 Review of Benchmarking Work

Extensive review of benchmarking work is undertaken in many jurisdictions. In Britain, for instance, at least two preliminary revenue proposals are issued by the regulator before a final decision. A lengthy period is allowed in Ontario and several other jurisdictions for review of commission-initiated benchmarking studies.

Litigated proceedings are the norm in the United States and Canada. If a benchmarking study is presented in direct testimony, it will (in the absence of a rate case settlement) be followed by the submission of detailed working papers, rounds of data requests, oral testimony, and sometimes a technical conference. Other parties may present counterstudies that are subject to similar scrutiny.

---

[10] Lawrence Kaufmann et. al., *Calibrating Rate Indexing Mechanisms for Third Generation Incentive Regulation in Ontario: Report to the Ontario Energy Board*, February 2008.

[11] See, for example, Mark Newton Lowry, et. al. , *Benchmarking PS Colorado's O&M Revenue Requirement*, June 2014.

6. EI's Benchmarking Study

6.1 Summary of EI's Work

6.1.1 Productivity Research

The AER's *Annual Benchmarking Report* for 2014 addressed the "core poles and wires" component of distributor cost. The report features results for multilateral total factor productivity ("MTFP") indexes, and also presents results for multilateral partial factor productivity ("MPFP") indexes for O&M and capital inputs.[12] The indexes were calculated using Australian data for the eight year 2006-2013 period. The opex PFP results are also presented in EI's Opex Bench report.

The output index used to calculate all three multilateral productivity indexes is the same. It includes output variables for delivery volume ("energy delivered"), customer numbers, total circuit line length, a "ratcheted" peak demand variable, and a reliability metric (minutes off supply). Ratcheted peak demand is the highest value of peak demand a distributor faced in the current and previous years of the sample period. It is used because the year-to-year variation in *actual* peak demand makes it less reflective of the *expectations* that drive utility construction decisions related to system capacity.

Development of the output index is discussed at some length in the opex benchmarking report. Evidently, a specification was previously considered that included throughput, customer numbers, a reliability metric, and a system capacity variable consisting of the product of line and cable circuit length and the total installed capacity of distribution-level transformers. The four non-reliability variables ultimately chosen for this study had previously been used in a productivity trend index prepared by PEG Research for the Ontario Energy Board.

[12] There are also results for partial factor productivity indicators that use one-dimensional scale metrics.

EI attempted to develop elasticity-based weights for the output subindexes econometrically using a translog cost function. This failed because the first-order terms for some scale variables had negative parameter estimates. This may reflect the modest sample size and multicollinearity between the scale variables, and likely also reflects the lesser importance of the variables with negative parameters as cost drivers. Instead, EI developed scale-index weights using an alternative method that did not consider the impact of scale variables on opex. The reliability metric was treated as a "negative output" with a scale-index weight based on the value of consumer reliability. A distributor with good reliability was thus credited with more output.

In the multilateral *total* factor productivity index, capital quantities were measured using the physical assets approach. There are capital quantity subindexes for overhead distribution lines, overhead subtransmission lines, underground distribution cables, underground subtransmission cables, and transformers and substation capacity. The opex input quantity is measured as the ratio of opex to an opex input price index.

EI acknowledges that productivity indexes potentially fail to control for differences in numerous cost drivers that utilities face. To address the extent of distortions caused by this limitation, they undertook a second-stage regression of the opex MPFP indexes on the following business condition variables: customer numbers; customer, energy, and demand densities; the share of underground cable length in total circuit kilometers; SAIDI; and a variable indicating the prevalence of single stage distribution substations. None of these variables had a statistically significant individual parameter estimate. The authors conclude on page 24 of their report that "the opex specification used thus appears to adequately allow for these environmental factors."

Salient results of the productivity work are as follows.

- Average productivity levels during the sample period were generally higher for distributors in Victoria and South Australia. These are the privately owned distributors in the sample. Ausgrid and Essential Energy had two of the four lowest productivity levels with respect to opex, capital cost, and total inputs in 2013. The productivity level of Endeavour was in the intermediate range.

- Opex productivity trended downward over the sample period for most distributors. The companies with the largest percentage declines were Citipower, United, and SA Power Networks. The productivity trends of Ergon and Ausgrid were positive. The relative productivity of the latter utilities thus improved during the sample period.

6.1.2 Econometric Cost Modeling

EI's Opex Bench report also presents results from econometric cost modeling. Cobb-Douglas and translog functional forms were considered in the econometric work. Two procedures were used to estimate model parameters: SFA and a least squares method that corrected for autocorrelation and heteroskedasticity. All models were estimated using Stata statistical software.

The econometric research was based on a sample that included data from distributors in New Zealand and Ontario in addition to Australia. The transnational sample used for the featured results contained data from 68 companies. The sample period for the study was 2006-2013 for Australia and New Zealand, and 2005-2012 for Ontario. The total size of the sample was 544 observations.

Apart from functional form, the econometric models had broadly similar specifications. The dependent variable was real opex (the ratio of nominal opex to an opex input price index). This enforces a prediction of cost theory and reduces the number of parameters requiring estimation. Three scale variables were featured: circuit kilometers, ratcheted peak demand, and the number of customers. Delivery volume was also considered but found to have a negative parameter estimate.

Each model also featured a trend variable and one Z variable: the share of undergrounded circuit in total circuit kilometers. The estimates of the trend variable parameter were (positive) 1.8% in the Cobb-Douglas SFA model and 2.0% in the Cobb-

Douglas and translog least squares models. These estimates suggest that O&M productivity was declining at a rapid rate over the sample period.

A group test of the statistical significance of second-order term parameter estimates in the least squares translog model indicated that the relationship of cost to operating scale was not well represented by the simpler Cobb-Douglas functional form. When the translog model was estimated using SFA, however, the scale variable parameter estimates for several Australian distributors suggested negative cost elasticities, in violation of cost theory. This suggests that EI's transnational data set may be inadequate to accurately estimate a translog model with SFA when there are three scale variables.

All the cost models presented by EI included country binary (aka dummy) variables for New Zealand and Ontario. EI explains its inclusion of these variables on page 31 of the Opex Report with the following comments:

> We cannot be certain that we have exactly the same opex coverage across the three countries so we have included country dummy variables for New Zealand and Ontario to pick up differences in opex coverage (as well as systematic differences in operating environment factors such as the impact of harsher winter conditions in Ontario). The country dummies will also pick up differences in conversion factors not adequately captured by our use of [Organization of Economic Cooperation and Development gross domestic product] purchasing power parities to convert financial variables to Australian dollars.
>
> We have... explicitly included country–level dummy variables (for New Zealand and Ontario) in our cost functions to control for possible cross–country differences/inconsistencies in accounting definitions, price measures, regulatory and physical operating environments, etc.

The parameters of both country-level dummy variables were generally significant in the econometric models presented. The models were then used only to consider the relative rankings of the Australian companies. In addition, the least squares models included dummy variables for each Australian distributor. The parameters associated with these variables were interpreted as representing the relative efficiency levels of the Australian firms over the sample period.

The SFA method directly estimates inefficiency, so the SFA Cobb-Douglas model produced average efficiency scores for each utility. Confidence intervals constructed around these estimates were quite narrow. For instance, the point estimate of Ausgrid's efficiency score was 0.447, while the lower and upper bounds of the confidence interval were 0.418 and 0.478.

Comparable efficiency scores were computed for the least squares and opex MPFP efficiency appraisals by measuring the distance from the efficiency of the best performing distributor. Like the efficiency scores obtained via SFA, these scores represented average efficiency during the entire sample period.

The efficiency scores obtained via the various methods were broadly similar. Distributors in Victoria and South Australia generally had the highest scores, while distributors in ACT, NSW, and Queensland had the lowest scores. In several cases, however, there was a considerable difference between the scores obtained from the MPFP indexes and the translog cost function. In particular, distributors with more rural service territories like Essential Energy, Powercor, and Ausnet Distribution all did considerably better with the translog cost function than with the MPFP indexes. Distributors with more urban service territories like Actew and Citipower did considerably worse in the translog appraisal.

EI's research suggests that Australian DNSPs differ widely in their opex efficiency. The worst performing utility had an efficiency score of only 0.399 using the Cobb-Douglas SFA model, 0.357 using the Cobb-Douglas least squares model, and 0.320 using the translog least squares model. The analogous lowest efficiency score using the opex MPFP index was 0.422.

The proposed opex revenue adjustments are based on the efficiency scores from EI's preferred method, the Cobb-Douglas cost model estimated with SFA. The real (constant-dollar) opex budget for each distributor in the 2013/14 base year was obtained through a multistep process. The first step was to adjust the company's average real opex over the full sample period downwards based on the discrepancy between its efficiency score and 0.86% (a weighted average of the efficiency scores of the five DNSPs with scores greater than 0.75%). This is tantamount to using a top quartile performance standard. An adjustment was then made in each companies' favor for various additional considerations not addressed in the statistical benchmarking. These adjustments were 30% for ActewAGL and 10% for the NSW distributors. The final step was to escalate cost to 2013 using company-specific adjustments for growth in productivity and operating scale, and standard adjustments for opex input price inflation. The productivity growth targets and the output quantity index were developed econometrically using the Denny, Fuss, and Waverman method discussed in Section 3.3.2 above.

EI's methodology produced the following proposed adjustments to the network services opex reported by NSW and ACT distributors for 2013:

ActewAGL	45%
Ausgrid	33%
Endeavour	13%
Essential Energy	35%

6.2 Critique of EI's Work

6.2.1 Productivity Indexing

General limitations of productivity indexing as a cost benchmarking tool were discussed in Section 3.3.2. We noted that these indexes make only crude adjustments for differences in operating scale, since they do not handle variations in opportunities to realize scale economies well. EI's evaluation of the significance of the second-order parameters in its least squares translog opex model indicated that variations in these opportunities were significant.

Moreover, productivity indexes do not account for the cost impacts of miscellaneous Z variables directly, making the careful selection of peer groups with numerous peers a matter of great importance. However, the small size of the Australian sample makes the construction of such peer groups impractical, especially for companies like Ausgrid and Essential Energy that face atypical operating conditions. EI appraised the productivity of all Australian distributors using the same peer group.

For these reasons, we do not believe that multilateral productivity indexing should be the featured methodology in the AER's Annual Benchmarking Reports. Index results should, at a minimum, be supplemented by results using econometric methods.

We also have concerns about certain details of EI's productivity work. Chief among these is the manner in which EI developed the scale index.

- We believe that scale indexes used in opex benchmarking should feature the scale variables that affect *opex* rather than *total* cost, and that they should be combined using elasticity weights that reflect their relative opex impacts. A sound specification is best achieved by an econometric opex study that considers numerous candidate scale variables, as well as a range of potentially important Z variables, in order to reduce the likelihood of omitted variable bias. EI did not use this approach, assuming at the outset that the scale index would include a particular set of scale variables.
- EI used the same scale index for opex and capital productivity that it developed to measure *total* factor productivity. This erodes the usefulness of its *partial* factor productivity indexes for benchmarking, however, since the key drivers of total, O&M and capital cost are likely to differ. For example, substation capacity is likely to be a more important *opex* driver than ratcheted peak demand.
- EI's choice of scale variables for the productivity index (and also the econometric model) was rationalized in part by the fact that customers, volumes, line lengths, and ratcheted peak demand were the scale variables in an output index developed by PEG for the Ontario Energy Board. However, this specification was

based on econometric *total* cost research and results were used to develop a *total* factor productivity index. Furthermore, a substation capacity variable was not an option since good substation capacity data were unavailable for Ontario.

- We also question the way in which reliability was included in the scale index. The impact of reliability on opex is a complicated empirical issue. Good reliability may require higher opex, but it also depends on weather, forestation, system undergrounding, AMI, and system reinforcements. EI's approach to reliability unfairly favors urban utilities in Victoria and ACT since these utilities enjoy favorable reliability operating conditions.
- The input price index used to compute the opex input quantity was not levelized to account for variation between states in the levels of labor or M&S prices. Cost share weights were not company specific. For both of these reasons, important information about differences in opex input prices faced by Australian distributors (e.g., the higher input prices faced by Ausgrid) was ignored.
- The second-stage regression on opex MPFP, which EI used to argue that the MPFP specification adequately accounted for scale and several operating environment factors, is unconvincing. First, the analysis fails to address several potentially important factors. For example, no attempt is made to evaluate the impacts of climate, forestation, or the share of lines with voltages greater than 66 kV, despite the availability of Australian data that could be used for this purpose. Second, the presence of multicollinearity between the included variables in the second stage regression could have easily contributed to the statistical insignificance of their individual parameter estimates. We found, for example, that the pairwise correlation coefficients for the three density measures were 0.9095, 0.9261, and 0.9567. For this reason, EI should have performed the analysis using separate regressions, or at least should have tested the joint significance of the parameter estimates. Given that they did neither, their claim that the opex specification "appears to adequately allow" for these variables is unjustified. The small size of the Australian dataset meant that

variables should be chosen carefully for such an exercise. EI's regression included several variables that considered variations in the cost impact of scale even though the Australian data set was not large enough to support development of a cost model with second order terms for scale variable.

- A physical asset treatment of capital quantities was used in the MTFP index. This was noted in Section 3.3.2 to be controversial since it does not consider important dimensions of capital cost management. The fact that "transformer and other" capacity appeared in the denominator of the MTFP index as an *input* quantity index may be one reason why EI was drawn to a peak demand variable in the *scale* index despite its lesser relevance in an opex study.

6.2.2 Econometric Modeling

Econometric cost research is generally a more accurate basis for cost benchmarking than productivity indexing. Unfortunately, it is difficult to develop a fully satisfactory econometric benchmarking model using the data that the AER has thus far gathered. The Australian data are by themselves too limited to develop the kind of detailed opex benchmarking model with numerous Z variables and second-order scale terms that we have shown to be feasible using the large datasets available in the US.

Transnational datasets can be used to increase sample size and variability, but introduce additional complications. Problems encountered with the Ontario and New Zealand data used by EI are illustrative.

- The definition of opex for the overseas companies does not appear to closely match the AER's network services opex. Opex in Ontario, for instance, includes the costs of customer care services such as metering and billing but excludes costs of maintaining substations with incoming voltage exceeding 50 kV.
- As noted above and acknowledged by EI, data on substation capacity comparable to those available in Australia are unavailable in Ontario, and reliability data in Ontario don't exclude major event days. EI notes on p. 32

of its Opex Bench report the general "lack of operating environment data in Ontario." This limits the ability to account for potentially important business condition variables in the benchmarking exercise now and in future years if use of Ontario data continues.

- Most Ontario and New Zealand distributors are much smaller than those in Australia. Many in Ontario, for example, are municipally owned utilities serving small communities. Therefore, adding data from Ontario and New Zealand to the sample does little to remove the outlier status of larger Australian distributors like Ausgrid.
- EI tried to finesse this problem by using a "medium" data set that includes only overseas companies serving at least 20,000 customers. However, only 12 of the 53 companies added serve at least 100,000 customers, which is not even the number served by the smallest Australian DNSP.
- Even with the inclusion of these much smaller overseas firms, the 544 observations in the transnational sample is still far below the size of the US sample used in the Portland study which supported a translog opex function with numerous Z variables.

We also have concerns about details of EI's econometric work.

- A substation capacity variable was not considered, despite its substantial relevance to network services opex. Its omission appears to have been due in part to the desire to include the Ontario data.
- The Z variable specification is inadequate. Variables addressing climate, forestation, system age, AMI, and the share of lines greater than 66 kV should have been considered. The single Z variable that does appear in these models is the share of undergrounded lines in total *circuit* miles. We have seen that this undergrounding variable is biased against utilities like Ausgrid and Energex that serve a mix of urban and rural areas.

- As in the productivity work, important information about how input prices differed between utilities and changed over time was lost, because the input price index was not levelized, and company-specific cost share weights were not computed for any country in the transnational sample.
- We are not confident that EI's use of country-specific dummies is a sufficient remedy for possible differences in opex coverage or operating environment between countries. Dummy variables only affect the intercept term in an econometric model; they do not adjust the slope coefficients. In other words, they can account for a situation in which costs differ by a consistent amount in a given country, but they cannot address differences in the degree to which cost changes in response to a change in another variable.
- A model with a Cobb-Douglas functional form was featured even though the second-order terms were found to be statistically significant as a group in the least squares translog model. A statistical test revealed that these parameter estimates were also significant as a group in the SFA translog model. This implies that the model used in the featured SFA benchmarking exercise was misspecified with respect to functional form. This matters particularly for scale outliers like Ausgrid and Essential Energy.
- The SFA estimation procedure that EI employed does not correct for heteroskedasticity or autocorrelation. These are common features of the data used in statistical cost research, and not accounting for them can bias benchmarking results.
- The estimates of company dummy parameters that EI uses in its least squares models to assess performance are likely to be inaccurate due to

an "incidental parameters" problem.[13] When company dummy variables are included in a cost model alongside other variables, the model is said to have two kinds of parameters. The parameters of the company dummies are called "incidental," and the parameters of the other variables (such as customers or line miles) are called "structural." The structural parameters can be estimated based on information from the entire sample. The incidental parameters, on the other hand, must be estimated using significantly less information since only one company can provide information for each. The effect of this problem on EI's models is that, since the number of time periods per company is not very large, the company dummy parameters do not have good statistical properties and should not be relied upon for statistical inference. In fact, when reporting results from models estimated with subject-specific dummies, practitioners often just omit the list of point estimates for these parameters and briefly note that the model included subject fixed effects, since the estimates of those fixed effects are not meaningful.

- We also question the confidence intervals that EI presents for the SFA Cobb-Douglas cost efficiency measures in Table 5.6 of its opex bench report. According to Horrace and Schmidt's 1996 article that is referenced by the Stata software routine that Economic Insights used to calculate confidence intervals for the inefficiency terms, those confidence intervals are not meant to be used to compare companies with one another. These confidence intervals are called "marginal," meaning that they relate only to the range of possible values for that specific estimate, considered in isolation. In order to rank the companies in the population,

[13] The incidental parameters problem was first described by Neyman and Scott (1948). The first two sections of Lancaster (2000) describe this problem briefly and intuitively. Cameron and Trivedi (2005) pages 750-757 discuss in more detail the effect of the problem on various types of linear panel data models.

joint confidence intervals that consider error in all the estimated firm inefficiencies together should have been used.

PEG's Transnational Benchmarking Study

The benchmarking study PEGR prepared for the AER illustrates how markedly different results can be obtained using alternative and more defensible benchmarking methods. Our research method was similar to EI's in several respects.

- Econometric cost modeling was the featured method.
- The dependent variable was real network services opex.
- The model was estimated using a least squares estimator in the Stata software package that corrects for autocorrelation and heteroskedasticity.
- A transnational sample was employed to estimate model parameters.
- A translog functional form was considered.

However, our study produced quite a different ranking of the Australian DNSPs. This is illustrated in Table 4 and Figure 1 which compares efficiency scores from EI's featured Cobb-Douglas SFA model to analogous scores using our transnational least squares translog model. It can be seen that the NSW distributors performed much better, and Endeavor was one of the best performers. Certain other distributors (e.g., Citipower, SA Power Networks) performed worse. The efficiency scores are more compressed. It is less clear that investor-owned distributors tend to be better opex performers.

Here are some likely reasons for the different results.

- US data were used to enlarge the sample for model estimation, rather than Ontario and New Zealand data.
- Our models featured different scale variables. Good substation capacity data were available for both countries, allowing the inclusion of this variable in the model. Also, we did not assume the output specification ex ante, which

Table 4

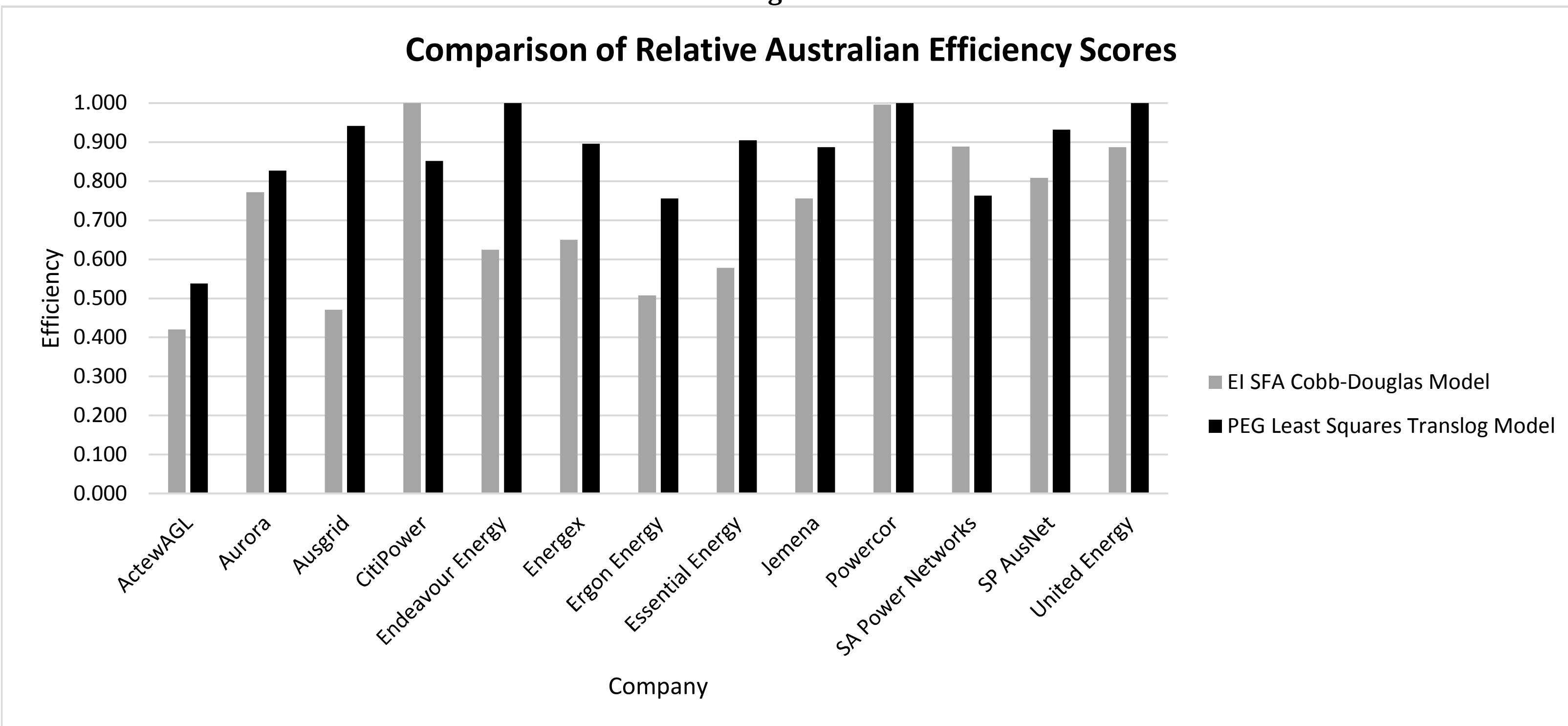
Comparison of Relative Australian Efficiency Scores

Company	Economic Insights SFA Cobb-Douglas Model: Raw Efficiency Score[1]	Economic Insights SFA Cobb-Douglas Model: Standardized Efficiency Score	Pacific Economics Group Least Squares Translog Model: Difference from Benchmark (%)[2]	Pacific Economics Group Least Squares Translog Model: Standardized Efficiency Score
United Energy	0.843	0.887	-3%	1.000
Endeavour Energy	0.593	0.624	-3%	1.000
Powercor	0.946	0.996	-3%	1.000
Ausgrid	0.447	0.471	3%	0.942
SP AusNet	0.768	0.808	4%	0.932
Essential Energy	0.549	0.578	7%	0.905
Energex	0.618	0.651	8%	0.896
Jemena	0.718	0.756	9%	0.887
CitiPower	0.95	1.000	13%	0.852
Aurora	0.733	0.772	16%	0.827
SA Power Networks	0.844	0.888	24%	0.763
Ergon Energy	0.482	0.507	25%	0.756
ActewAGL	0.399	0.420	59%	0.538

[1]Lawrence et al., (2014), *Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs* , Report by Economic Insights for the Australian Energy Regulator, Eden, page 37.

[2]Lowry et al., (2014), *Database for Distribution Network Services in the US and Australia* , Report prepared by PEG Research for the Australian Energy Regulator, Madison, page 27.

Figure 1



Comparison of Relative Australian Efficiency Scores
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
Efficiency
ActewAGL
Aurora
Ausgrid
CitiPower
Endeavour Energy
Energex
Ergon Energy
Essential Energy
Jemena
Powercor
SA Power Networks
SP AusNet
United Energy
Company
EI SFA Cobb-Douglas Model
PEG Least Squares Translog Model

permitted us to select the most relevant variables for analysis. We found substation capacity to be a more important cost driver than ratcheted peak demand, while the volume variable consistently had a negative sign.

- We estimated scale and Z variables simultaneously, rather than attempting to adjust for the influence of the Z variables later.
- Route length was used in our models instead of circuit length.
- Miles of intermediate voltage lines were included as a Z variable. Its parameter estimate was highly significant, indicating that the need to maintain such lines adds substantively to cost.
- Our undergrounding variable was the share of underground *plant* in the value of plant. We found that the parameter of this variable generally had greater statistical significance than that for the analogous variable computed using *circuit lengths.* [14]
- Translog rather than Cobb-Douglas functional forms were employed in the econometric models, permitting the elasticity of cost with respect to scale to vary for firms of different size. This may be particularly important for the rankings of utilities like Ausgrid and Essential Energy that are scale outliers.
- Input price indexes are levelized to reflect differences both between countries and between service territories. The model results therefore reflect the relatively high labor costs in NSW. Additionally, the opex input price index for the United States has company-specific cost share weights.
- Our benchmarking results pertain to the *most recent* years for which data are available, rather than the full sample period. This allows companies to be evaluated on their recent performance, which is arguably much more relevant than their activities many years ago.

[14] This was true for the Australian data. Due to a lack of sufficient circuit lengths data for the U.S. companies, we were unable to make this comparison based on the transnational sample.

We believe that the inclusion of US data in a benchmarking study for Australian DNSPs offers various advantages. However, the AER decided not to use data from US firms in its opex benchmarking report. It provided the following reasons.

- Many US companies do not have consistent data for reliability, line length, system age, peak load, or distribution substation capacity.
- The US companies with more consistent data provide too few additional observations.
- Most investor-owned US power distributors also operate transmission systems, and many also have sizable generation operations.

This dismissal of the merits of including US data in Australian benchmarking is uncharitable in several respects.

- Comparable *transformer* data are unavailable in the United States but comparable data on distribution *substation capacity* are available and we believe that these are more relevant. The substation capacity variable we developed had a highly significant parameter estimate in many models considered, including models estimated using only Australian data. This makes sense, since network services opex is likely to depend on the scale of substation equipment operated and maintained.
- US peak demand data are satisfactory for the numerous US distributors that do not generate power, and peak demand data for distributors that are engaged in power generation can be adjusted downward. However, we found little econometric support for including a peak demand variable in the models using either the Australian or transnational data sets.
- Good quality line length data were available for 15 US distributors, in the form of *total* distribution route miles. However, a much larger group of US utilities have reported overhead route miles and underground circuit miles for many years. A much larger US sample can thus be used in the future were the AER to ask

Australian DNSPs to itemize overhead and underground route miles. This should be a straightforward task. Route length is, in any event, a reasonable alternative to circuit length as a cost driver variable. It may produce better results for distributors with highly rural service territories. The parameter estimate for our route length variable was highly significant.

- Even the current more limited US data set more than doubles the number of sampled companies, and we were able to develop a translog cost model with three scale variables using this dataset.
- The objection about reliability data is surprising since comparable reliability data were unavailable for Ontario and reliability data were not used in EI's transnational research. Similarly, system age data were not used in any of EI's reported research.
- We have in any event identified more than 36 US power distributors with publicly available reliability data that are comparable to those in Australia. Standardized reliability data for a much larger sample are expected to be released by the US Energy Information Administration in February, and new reliability data will be released annually thereafter. Thus, integration of reliability in a benchmarking study using US data for a sizable group of distributors should be possible as early as next year.
- It is straightforward to add variables to a model estimated with US data to address the cost impact of distributor involvement in generation and transmission. We considered such variables but did not obtain encouraging parameter estimates.

As should be clear from the preceding discussion, including US data in an Australian benchmarking analysis is quite feasible. The US data also offer important advantages that are not mentioned in EI's opex report.

- Extensive itemization of opex data permits the construction of a network services opex measure that is similar to that used by the AER.

- Data from the US Bureau of Labor Statistics and the FERC permit calculation of levelized opex input price indexes with company-specific cost share weights. There are, for example, detailed price level data available by metropolitan area.
- The unusually high quality capital cost data in the US make it possible to do reliable transnational capital and total cost benchmarking, if desired. Gross distribution plant addition data are available for more than seventy distributors since the mid-1960s.
- The average operating scale of US distributors is much larger than in the Ontario and NZ samples. In the sample we developed for the AER, for instance, only one company served fewer than 100,000 customers. The average number of customers was around 572,000 --- fairly close to the 700,208 average of the Australian DNSPs. If Australian distributors start itemizing route miles, it will be possible to add numerous utilities with operating scales more like those of Ausgrid and Essential Energy to the sample. Numerous US utilities serve a large metropolitan area and a surrounding rural area, like Ausgrid and Energex do.

In summary, US data in our view have much more potential than data from Ontario (and possibly also New Zealand) to provide the basis for transnational benchmarking of Australian distributors. The value of these data is especially great for benchmarking larger distributors like Essential Energy and Ausgrid. A dataset combining New Zealand and US data merits consideration.

6.2.3 Performance Standard

The use of a top quartile performance standard to benchmark Australian DNSPs is problematic. As discussed above, the short-run frontier may not represent the sustainable long-run opex level. Moreover, even accurately locating the short-run frontier is difficult due to the sensitivity of results to data irregularities. In addition, expecting companies to operate on the frontier is unrealistic, since even in competitive markets, most firms are well away from the frontier at any given point in time.

EI makes several adjustments to lessen the impact of using a frontier performance standard in its benchmarking results, such as averaging the results of companies rated at least 75% efficient, and several post-hoc adjustments to individual utilities' efficiency targets. However, the underlying expectation that companies should be able to attain a performance level close to the frontier remains at the heart of EI's recommendations. At least one of the companies in EI's top efficiency group is requesting sizable opex increases for the next five years.

We recommend that a competitive market standard, like that employed by Ofgem, be used to assess the need for potential opex revenue adjustments. Utilities should be considered for disallowances only to the extent that their measured efficiency fails to reach the *lower bound* of the top quartile of the sample distribution. All performers above this level should be eligible for superior rates of return.

6.2.4 Sample Average Efficiency Focus

The limitations of focusing on average efficiency over the full sample period should also be noted. It is important for the regulatory community to have information about recent performance rankings. This is straightforward to provide with a least squares benchmarking model since it readily yields appraisals for individual years of the sample period.

Among the Australian DNSPs, some state-owned distributors (e.g., Ausgrid and Ergon) improved their relative performance in later years of the sample period while certain privately-held distributors (e.g., SA Power Networks) regressed. Moreover, Endeavour Energy is a top performer in some of our benchmarking exercises. The benefit of private ownership on efficiency is thus unclear.

6.2.5 Size of the Proposed Disallowances

The revenue adjustments proposed by the AER are at the extreme end of the range of adjustments we documented in our survey. Adjustments of this magnitude could compromise the ability of the companies to maintain current reliability standards.

The rates of return of all three companies would fall sharply, and the market values of Ausgrid and Endeavour could be materially reduced at a time when the NSW government plans to sell them. Regulatory risk would increase materially, with possible repercussions on the cost of acquiring funds.

6.2.6 Conclusions

We conclude from this review that EI's benchmarking analysis of network services opex is seriously flawed. Accurate benchmarking will be challenging for years to come, due to the small size of the Australian dataset and the need for further improvements in data reporting. Indeed the number of companies in the Australian sample will always be limited, even as additional years of data accumulate. Ausgrid and Essential Energy will always be outliers in an Australian sample. Transnational data permit development of more complex benchmarking models, but introduce other challenges, such as the comparability of data. Ausgrid and Essential Energy are still outliers in EI's transnational sample. EI has also made some controversial choices regarding benchmarking methods and did not present a thorough sensitivity analysis.

EI has defended its methods on the grounds that there is no apparent bias between the results for rural and urban utilities. The validity of this claim is unclear, since three of the five distributors in the sample serving sizable central business districts perform poorly. Furthermore, there is clear bias against utilities with intermediate voltage systems. There may also be bias against utilities with large systems. The great dispersion of efficiency rankings suggests that important business conditions have been excluded from the analysis.

EI also touts the fact that benchmarking results using alternative methods produce similar results. However, this may be due in part to biases that are shared by all the methods. These include the failure to account for regional differences in input prices, and the omission of several potentially important business conditions from the models, such as forestation and the share of lines with intermediate voltages. In addition, no models are presented that exclude data for Ontario or New Zealand, making it unclear to what degree the data from these countries are driving the results. Given the comparatively small size of most of the overseas firms, this is a real concern.

Research by PEG that used US data and alternative econometric methods produced markedly different benchmarking results. The AER ignored this evidence in developing its proposed opex disallowances. Our model was presented to the AER as an illustration of the potential value of US data but is nonetheless of high quality.

We are also concerned that the AER's process for reviewing its benchmarking work falls short of international best practice. Data requests were permitted and properly answered but there have been no provisions for a technical conference or a second round of consultation should the AER revise its study or proposals. We feel that the point EI has reached in its benchmarking work should have been reached many months before the use of benchmarking in rate setting. Ontario has a better process for reviewing benchmarking research.

For all of these reasons, we believe that EI's current benchmarking results provide an unsatisfactory basis for any reduction to opex revenue other than a standard stretch factor. The large adjustments that are proposed are especially inappropriate given the AER's lack of experience with the new Australian data and the compression of the review window. Even with improved methods, statistical benchmarking will not for the foreseeable future be accurate enough to legitimize the kind of large, immediate disallowances for average and poor cost performers that AER is contemplating. We encourage the AER to upgrade its data and benchmarking methods, accumulate more data, and consider a much more limited role for benchmarking in setting NSW opex revenue.

Appendix: Benchmarking Precedents in Utility Regulation[15]

A.1. Britain

Power distributors in Britain have been regulated since their privatization in 1990. The initial regulatory agency, the Office of Electricity Regulation, was succeeded by the Office of Gas and Electricity Markets ("Ofgem"). During this entire period, distributors have operated under multiyear price controls. Price controls have been periodically updated in distribution price control reviews ("DPCRs"). Controls were updated in 1998/1999 (DPCR3), 2003/2004 (DPCR4), and 2008/2009 (DPCR5). Ofgem's most recent review detailed a new price control approach called Revenue = Incentives + Innovation + Outputs ("RIIO"), and was completed in 2014.

The detailed cost forecasts yielded by the British "building block" approach in these reviews have provided the basis for RPI-X escalation formulas yielding expected revenue streams with equivalent net present value. Statistical benchmarking has played a role in all of the DPCRs. Where cost disallowances have been made due to benchmarking or other appraisal techniques, they have usually been applied to the initial revenue requirement.

A.1.1 Review Process

The earliest price control reviews were somewhat opaque and did not specify the regulators' specific determinations on building block elements. The regulator's determination was instead presented as a "package deal," which companies could either accept or appeal. Over time, British regulatory reviews have become more thorough, based on better information, and followed a more clearly defined and organized process.

---

[15] This appendix draws on material included in a report by PEG to the Bundesnetzagentur, Germany's DNSP regulator.

A British style price review process has now emerged, which commences nearly three years before the date of the next price control period and can last over two years for distributors. The procedure involves the following steps:

- Initial "strategy" consultation summarizing key issues and proposed approaches
- Individual consultations to deal with particular issues
- Initial proposed cost and revenue forecasts by the distributors with accompanying justifications
- Initial revenue proposals by OFGEM to test the reactions of the regulated company and other interested parties
- One or more interim proposals from OFGEM
- Final proposals.

After price controls are set, Ofgem undertakes an assessment of the electricity distribution price control process.

A.1.2 Benchmarking Prior to RIIO

Methods Used

Ofgem has primarily relied on econometric benchmarking in its price reviews. Models with simple functional forms like the Cobb-Douglas have been estimated using ordinary least squares. A scale index called the comprehensive scale variable ("CSV") is typically the only cost driver. Cost data are normalized to ensure these data were defined and collected comparably across all DNOs.

The CSV for DPCR4 was based on each distribution network operator's ("DNO") number of customers served, kWh distributed, and network length. The weights applied to these variables in developing each DNO's CSV were 25%, 25%, and 50%, respectively. These weights were considered roughly proportional to the impact of each scale measure as a "driver" of distribution opex. The final benchmark developed in the DPSC4 review depended on three benchmarking regressions: a base regression and two alternatives.

The DPCR4 review also undertook some data envelopment analysis ("DEA") as a "cross check" on the econometric results. However, Ofgem concluded that the DEA results "are not plausible so [DEA] has not been incorporated directly."[16] Ofgem also undertook DEA in DPCR5 and believed that "the DEA analysis broadly supports the regression analysis we have undertaken. However, because of the . . . limitations of DEA we have not adjusted our view of comparative efficiency scores because of running that analysis.[17] SFA runs were also undertaken in DPCR5 but were not pursued further due to data limitations. Ofgem also undertook some preliminary transnational benchmarking featuring northeastern US utilities. However, this was not used in Ofgem's final proposals due to concerns about data compatibility.

In DPCR5 Ofgem undertook opex benchmarking at several different levels of aggregation. These included a total opex benchmarking study and two kinds of "group" benchmarking studies with regressions addressing disaggregated costs. The regressions featured different cost drivers. The cost drivers identified by Ofgem included metrics like modern equivalent asset value (MEAV), underground faults, and spans cut. To determine opex efficiency scores, Ofgem developed weights for each type of regression to determine overall efficiency shares.

Focus

In DPCR3 Ofgem benchmarked only opex. DPCR4, however, considered total operating and capital expenditures ("totex") as well as opex. DPCR4 also included research on total factor productivity and O&M partial factor productivity. In DPCR5 Ofgem relied on benchmarking of total opex and two different levels of disaggregated opex, as noted above.

---

[16] Office of Gas and Electricity Markets, *Electricity Distribution Price Control Review: Final Proposals*, November 2004, p. 70.

[17] Office of Gas & Electricity Markets, *Electricity Distribution Price Control Review: Final Proposals Allowed Revenue: Cost Assessment appendix,* December 2009, p. 96.

Use of Benchmarking Results in Rate Setting

In DPCR3, Ofgem set opex targets by assuming they were capable of closing 75% of the gap between their efficiency and that of the second most efficient firm by the second year of DPCR3.[18] In DPCR4, each distributor's allowed opex was based on the gap between its efficiency score and that defining the margin between the top and the next best quartile. This tended to reduce the size of disallowances and permitted top-quartile performers to earn a superior rate of return. Ofgem's rationale for this decision was that an "upper quartile benchmark…provides a more robust and sustainable benchmark than a frontier based on one company." The new level of allowed opex was effective immediately, rather than being implemented over a transition period. Ofgem maintained this policy in DPCR5.

Accounting/Data Quality Issues

Standardized accounting requirements were not imposed by OFGEM for many years. However, since coming to appreciate the importance of consistent accounting, Ofgem has put a significant effort into preparing "Regulatory Accounting Guidelines." As a result accounting procedures have become more transparent, and the scrutiny of companies' costs has become much more detailed.[19]

Some major accounting issues encountered in UK regulation have directly impacted the practice of benchmarking. This is particularly true of the normalization of opex costs, which included a great number of specific cost adjustments needed to put all DNO opex costs on a comparable footing. Noteworthy normaliziations in DPCR4 included the following:

- Removal of atypical and one-off costs, e.g., storm costs, storm insurance.

---

[18] Office of Gas and Electricity Markets, *Electricity Distribution Price Control Review: Initial Proposals*, June 2004, p. 66.

[19] The water regulator led the way with its "Book of Numbers," which is a detailed set of accounting and reporting requirements using standardized accounting.

- Adjustments to the capitalization of overheads to offset different overhead allocation methods.
- Adjustments for regional factors to take account of significant geographical, demographic and operational circumstances.

A.1.3 Benchmarking in RIIO

RIIO-ED1 began with a strategy consultation, which focused on what RIIO-ED1 would include and how distributor forecasts would be assessed. Ofgem's initial decision, called the strategy decision, included an outline of the work Ofgem and its consultants would be undertaking on benchmarking issues. Distributors would be given an initial opportunity to present and justify their forecasts to see if they could earn "fast-track" treatment. Fast-track treatment would lead to Ofgem approval of their cost forecasts, allowing them to avoid more rigorous scrutiny of their costs. Distributors that did not earn fast-track treatment were slow tracked and given two additional opportunities to improve their justifications or adjust their cost forecasts before a final decision was issued. The discussion below focuses on the benchmarking that was undertaken for distributors in the slow track.

Ofgem's benchmarking work for RIIO-ED1 featured econometric models, and appears to have used Stata statistical software. Three distinct models were employed, one for "top-down" totex, another for "bottom-up" totex, and a disaggregated model. The top-down totex model featured a Cobb-Douglas form and a least squares estimation procedure.[20] A scale index was employed, with a 12% weight on customer numbers and an 88% weight on the modern equivalent asset value ("MEAV").[21]

The disaggregated research took the form of an activity-level analysis. The analysis undertaken varied by activity, with some activities being assessed with

[20] Ofgem considered but rejected a random effects estimator, because it was overly complex and the results were not changed by its use.

[21] MEAV was defined as the sum of the products of the number of assets owned by a distributor and Ofgem's view of the unit cost of that asset with some exceptions.

regression, and others assessed with age-based modelling, ratio analysis, trend analysis, and technical review by consultants. The bottom-up totex model also utilized a Cobb-Douglas functional form and a least squares estimator. The bottom-up model had different drivers than the top-down model.

In each of the models thirteen years of data were used, consisting of four years of actual data from the current price control, and nine years of forecast data from the final year of the current price control plus the entirety of the RIIO-ED1 period. Data were gathered for 14 British distributors. The data were normalized to account for 4 items: regional labor costs, except for business support costs for three regions; distributor-specific factors; exclusions from totex models; and other adjustments.

Ofgem applied the benchmarking results in the slow-track analysis to totex using a performance standard which was based on the 25th percentile of combined performance in the three models. Ofgem's initial assessment of the models led it to place a heavier weight on the disaggregated model, due to concerns about the limitations of totex modelling. As the review progressed, however, Ofgem's views on the model changed and it proposed using a weighted average of distributors' performance estimates, utilizing a 25% weight for the result of each totex model and a 50% weight for the result of the disaggregated model. Ofgem developed a more favorable view of the totex models partly because, compared to the disaggregated models, they reduced the chance that differences in cost categorization and allocation or tradeoffs between opex and capex would cloud assessments of efficiency. Ofgem's approved totex allowances were based 75% on Ofgem's view and 25% on distributor submitted forecast costs. The largest efficiency disallowance resulting from this approach was around 11%.

A.2 United States

Jurisdiction over electric utility rates in the United States is divided between the FERC and various public service commissions ("PSCs"). Many US electric utilities are still vertically integrated, and generate much of the power they deliver. The FERC regulates the terms of transmission and wholesale generation services, while the PSCs regulate the terms of distribution and any retail power supply services.

Rates for transmission services are often subject to "formula rate" plans that cause revenue to closely track costs. Rates for retail services are typically reset in rate cases that are irregularly timed.[22] A few states (e.g., California and New York) use multiyear price controls. An approach to escalating price controls was developed in the United States that is based on industry cost trends. A formula for an energy distributor's revenue growth might be

$$\textit{growth Revenue} = \textit{Inflation} - X + \textit{growth Customers.}$$

Here X (the "X factor") would reflect the TFP trend of power distributors and a "stretch factor" that may vary with a utility's special opportunity for TFP growth.

Statistical benchmarking studies are quite rare in FERC regulation but are occasionally filed in PSC proceedings. Studies have been fairly common in US proceedings to consider a price control based on indexing research since they may have a bearing on the value of the stretch factor. Studies have typically been filed by the utility, but have sometimes been filed by consumer advocates or commission staff.

US regulation is litigious. Utilities face off against multiple parties, often including a commission- funded residential consumer advocate, a division of the PSC's staff, advocates for commercial and industrial consumers, a low-income residential consumer advocate, and an environmental group. When benchmarking studies are filed, detailed working papers must be submitted. Several rounds of data requests and a technical conference may ensue. If the case is not settled, this is typically followed by rebuttal testimony and additional data requests, oral testimony, briefs, and reply briefs.

[22] Typically, rate cases are initiated by the utilities. However, in some instances concerns about a utility overearning can lead a PSC to order a company to file a rate case. In other instances, parties may agree to a specific date by which a rate case must be filed.

Benchmarking studies, when filed, typically have little impact. This is true for a variety of reasons. First, most commissions regulate only a few electric utilities, which limits the cost savings from benchmarking. Utilities may face special operating conditions, requiring sophisticated and costly benchmarking studies that the commissions and intervenors cannot usually afford to undertake. Second, many commissions assess the performance of their companies through management audits that directly review the processes and systems companies employ. In most cases these audits do not include statistical benchmarking studies. Third, even if a benchmarking study is filed, it is common for a company and other stakeholders to reach a settlement that is approved without relying on the results of the study.[23]

There are several instances in which no settlement between parties was reached, but the benchmarking study that was performed still did not play a direct role in the regulator's decision. Dueling benchmarking studies by Southern California Edison and a consumer advocate had no discernible impact in the California Public Utilities Commission's ("CPUC") decision on a proposed price control plan in the mid-1990s. The CPUC also declined to use the results of an econometric benchmarking study put forward by San Diego Gas & Electric on nuclear generation in 2005 because of concerns it had on the quality of the study.

In the rate case of the Ameren Illinois companies in 2009, the power distributors sponsored O&M unit-cost benchmarking studies that supported the overall reasonableness of the companies' O&M expenses. Two consumer advocates sponsored a study which used econometric benchmarking to produce quite different results. The commission endorsed the companies' benchmarking study over the consumer advocates' study because it believed unit-cost benchmarking was straightforward and

---

[23] Some examples of benchmarking studies that had little direct impact due to a settlement filed between parties include two studies performed by the Maine regulator's staff in response to Central Maine Power's 2000 and 2013 multiyear rate plan proposals, studied sponsored by Oklahoma Gas and Electric in its 2009 and 2011 rate cases, studies sponsored by Florida Power & Light in its 2009 and 2012 rate cases, and a study that Public Service Company of Colorado sponsored as part of its 2009 rate case.

easy to understand. The commission also questioned the consumer advocates' interpretation of their results, and how those results could be applied in the proceeding. The Illinois regulator declined to make any changes to the Ameren Illinois companies' O&M expenses as a result of benchmarking.

There are a few instances in the United States where benchmarking was a substantive issue requiring the PSC's input in a proceeding for an electric utility that wasn't settled. In Vermont, for example, parties agreed to a 2010 price control plan for Green Mountain Power with an inflation-X formula, where the value of the X factor would change over time based on the company's performance in an annual O&M unit-cost benchmarking study.[24] For example, if the company was a bottom-quintile performer in the benchmarking study, it would earn a 1% X factor, while if it was a top-quintile performer, the X factor would be set to zero. This provision was kept in the recent update of Green Mountain Power's PBR plan, and also incorporated into a PBR plan for Central Vermont Public Service.

A.3 Canada

Electric utilities in Canada are regulated at the provincial/ territorial level. Many Canadian utilities are owned by a municipality or province. In most jurisdictions rates for retail services are reset periodically in irregularly-times rate cases, although several provinces (e.g., Ontario, Alberta, and British Columbia) use multiyear price controls. Research on industry productivity trends plays a key role in price control design.

Due to the lack of a national data collection form like that used in the US or Australia, it is difficult to develop large, consistent data sets that can be used for benchmarking. This problem is compounded by the fact that most provinces have very few companies to regulate. Only the province of Ontario has been able to consistently

[24] The company's performance in the benchmarking study could also lead to adjustments in its rate of return on equity.

benchmark its utilities because it has a large number of power distributors and has made strong progress toward a assembling a consistent data set.

Ontario has taken a gradual approach to integrating benchmarking into the regulatory process. The first concerted effort to benchmark the numerous distributors in the province began in 2004, when the OEB was developing a successor to the province's first-generation incentive regulation mechanism (“IRM”) and a government-mandated rate freeze. At the time, the OEB was preparing to review and make decisions on more than eight rate applications in less than two years. A Rates Handbook outlining the methods that were expected to be followed in a rate case proceeding was developed. The Rates Handbook incorporated benchmarking through a section outlining a “Comparators and Cohorts” (“C&C”) study, which a consultant to the OEB’s staff was to undertake. The C&C approach envisioned the use of benchmarking to identify companies that might be less efficient, and which would therefore merit more scrutiny.

The C&C methodology was developed in a working group process that allowed representatives from the distributors and other parties, along with the OEB staff, to meet and verify the data, discuss the benchmarking methods to be undertaken, and determine how the results might be applied to the distributors. The recommendations of the working group were then presented to stakeholder groups, providing an opportunity for more parties to comment. The final product of the working groups was incorporated into the draft Rates Handbook, which led to a more formal process that allowed parties further opportunities to comment on the proposals.

In its approval of the C&C analysis, the OEB explicitly rejected its use as a prudence or efficiency threshold for companies’ costs, partly because the method was untried and partly because the quality of the data was deemed poor. One example of the limitations of the data was that they encompassed only the most recently completed three-year period. Nine distributors did not have comparable data for business conditions and were therefore excluded from the analysis.

The approved C&C methodology featured econometric cost research to identify the drivers of distributors' costs in two discrete areas: wires and interconnections services, and support services. Single-equation Cobb-Douglas models were developed for capital and opex. A translog cost function was also developed that featured numerous cost drivers and second order terms.

The results of the cost modelling were used in a hierarchical cluster analysis to divide distributors into seven cohorts for each of the cost areas. The cohorts varied in size, with some including just a single distributor while others had nearly 40. The cost modelling was also used to identify unit cost and other metrics that would measure the distributors' performance.

To inform its views on the third-generation incentive regulation mechanism ("IRM3") and distributor rate cases, the OEB initiated a proceeding that reconsidered the benchmarking of Ontario power distributors' costs. The proceeding began with an invitation to parties to comment on the benchmarking method, the variables to be included, and the need to collect additional data. OEB Staff's consultant then released a draft econometric and unit cost benchmarking study that was responsive to the initial comments. Only O&M expenses were benchmarked due in part to slow development of the required capital cost data. Stakeholders were provided multiple opportunities to better understand and comment on the study through both written comments and a technical conference. The OEB Staff's consultant then updated the study to incorporate the comments and add an additional year of data. An area of particular concern was the development of appropriate peer groups for the unit cost indexing work.

The X factor for IRM3 consisted of a productivity factor based on the 0.72% annual TFP trend of US power distributors, and a stretch factor which reflected a company's performance in the benchmarking studies. The benchmarking results, after being updated to reflect the additional year of data, were used to determine the stretch factor. Companies deemed significantly superior cost performers in the econometric benchmarking study based on a statistical test, and that were top-quartile unit cost performers received the smallest stretch factor of 0.2%. Companies that were significantly inferior cost performers in the econometric benchmarking study and bottom-quartile unit cost performers received the largest stretch factor of 0.6%. The OEB set stretch factors for the worst performers that it thought would incentivize distributors to improve their productivity without being punitive. Most companies fell between those extremes and earned a 0.4% stretch factor.[25] Each year, the benchmarking data would be updated, the models rerun with the same variables, and new stretch factor assignments made. This plan continued in operation until the end of 2013.

In late 2010, the OEB began a proceeding to develop a new method by which distributor performance could be assessed. A total cost benchmarking methodology was favored. A new working group was formed consisting of representatives from the industry, OEB staff, and other stakeholders. The group met several times to discuss performance measurement, including TFP measurement and benchmarking methodologies. All of the distributors were provided several opportunities to revise or verify the data to be used in the benchmarking and TFP studies, and all the data that the OEB relied upon were posted publicly. The proceeding also allowed for the distributors and other parties to review the report of the OEB Staff's consultant and to offer comments. In an attempt to focus comments on improving methodologies rather than on the results for specific distributors, the OEB Staff's consultant redacted the names of the distributors in the rankings. The OEB Staff had its consultant review and respond to the comments, and experiment with some of the variables that were proposed. The consultant was also tasked with updating the study to include the additional year of data that had recently become available.

The results of the total cost studies were applied to the price control regime outlined in IRM4. However, the OEB now allows companies to opt out of an IRM4 price cap and seek a custom IR plan based on a five-year or longer cost forecast. In either

[25] The OEB rejected viewpoints that a zero or negative stretch factor was appropriate because it believed that all distributors had a chance to experience incremental productivity gains.

case, benchmarking is relied upon. The IRM4 price cap includes an X factor that is the sum of the Ontario power distribution TFP trend of 0%, and a stretch factor tied to a distributor's performance in the total cost benchmarking study. Unlike IRM3, only an econometric benchmarking study is used to determine the stretch factor. Five stretch factors are possible: 0% for companies that have costs 25% or more below the model's prediction, 0.15% for companies with costs 10-25% below the model's prediction, 0.3% for companies within 10% of the model's prediction, 0.45% for companies with costs 10-25% above predicted, and 0.60% for companies with costs 25% or more above the predicted value. The benchmarking study is updated annually to reflect new data, but the model specification is not reconsidered. Companies are able to move between stretch factors during the term of IRM4. If a company opts for a custom IR plan, benchmarking will be used to test the reasonableness of the company's proposal. The benchmarking that will be relied upon includes the annual benchmarking study sponsored by the OEB, and any benchmarking study brought forward in a custom IR proceeding to test the reasonableness of a distributor's proposal.

A.4 New Zealand

The New Zealand Commerce Commission began formal rate and service quality regulation of its power distributors in the early 2000s. Its first effort to regulate the industry, dubbed the "thresholds regime," involved light-handed regulation if the distributor did not breach a price or service quality threshold. The price threshold was adjusted annually by a CPI – X formula. The X factor incorporated the TFP trend of the economy and a two-part stretch factor that reflected a company's relative productivity level and profitability.

A.4.1 Benchmarking Methods

A number of benchmarking analyses were performed in the 2003 thresholds proceeding. The benchmarking method featured by the Commission's consultant was multilateral total factor productivity (MTFP) indexing. New Zealand data were used for the exercise. MTFP indexes were calculated for every distributor in each year from

1999 through 2003. This work was complementary to the TFP trend analysis used to set the X factor, and used the same dataset and the same input and output definitions.

The outputs used in the MTFP indexes were customer numbers, kWh throughput, and a system line capacity measure expressed in megavolt-ampere kilometers ("MVA-km"). Five input quantities were used. These were real (inflation-adjusted) operating expenditures, MVA-km of overhead network assets, MVA-km of underground network assets, kVA of installed transformer capacity, and the value of other distribution assets. This latter asset value was not deflated over time. The initial commission-sponsored report also featured econometric benchmarking analyses. These regressions were estimated using data for New Zealand distributors.

A.4.2 Results

In 2003, the last year of the sample period, MTFP values for the sampled companies ranged from a high of 1.781 (i.e., productivity 78% above the industry average) to a low of 0.674 (i.e., productivity 32.6% below the industry average).

The econometric results were generally unsatisfying. Most of the variables in the cost function regressions were not statistically significant. Because of these problems, the Commission ultimately used the MTFP results as the basis for setting relative efficiency factors.

A.4.3 Translation of Results to X Factors

The MTFP factors were translated into X factor adjustments using a multistage process. The distributors were first ranked from top to bottom with respect to their estimated efficiency. Next, distributors were divided into three groups of similar size based on estimated performance. Companies in the high-efficiency group were given an X factor adjustment of -1%; those in the medium-efficiency group received no X factor adjustment; and those in the low-efficiency group received a 1% X factor adjustment.

A.4.4 2007/08 Thresholds Proceeding

In late 2007 the New Zealand Commerce Commission began a review of its thresholds regime. It released a discussion paper on the form of the next generation of thresholds, and endorsed the continued use of benchmarking to determine the relative efficiency of distributors. Included as an attachment to the discussion paper was an updated benchmarking study by its consultant.

The consultant's study was based on multilateral TFP indexes. These indexes featured three scale variables (volume, system-line MVA capacity, and the number of connections) and five input quantities (operating costs, overhead line capacity, underground line capacity, transformer capacity in kVA, and other capital). The output weights relied on an econometric Leontief cost function presented in the previous review. The passage of time allowed the consultant to acquire the data to form capital-input weights tailored to each distributor. The updated capital input weights were quite different from what had been used before, and this led to large changes in the relative rankings of several companies. Electricity Invercargill, which had previously been found to be the most efficient, was now the least efficient. Nelson Electricity fell from second to 25th, and The Lines Company rose from 21st to 8th.

Before benchmarking could be finalized, the Commerce Amendment Act of 2008 was passed. It states that "The Commission may not, for the purposes of this section, use comparative benchmarking on efficiency in order to set starting prices, rates of change, quality standards, or incentives to improve quality of supply."[26] Benchmarking has not been used in regulation in New Zealand since.

[26] New Zealand Commerce Amendment Act 2008 (Public Act 2008 No. 70), Part 4, Subpart 6, Section 53P, Subsection 10.

References

Caves, D. W., L.R. Christensen, and W.E. Diewert (1982), "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers," *The Economic Journal* 92, 73-86.

Cameron, Colin A. and Pravin K. Trivedi (2005), *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

Camfeld, R., et. al. (2005), *Phase II Report, Part I, Methodology and Study Findings: Comparators and Cohorts Study for 2006 EDR*, Report prepared for the Ontario Energy Board and Staff, October.

Denny, Michael, Melvyn A. Fuss and Leonard Waverman (1981), *The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications*, in Thomas Cowing and Rodney Stevenson, eds., Productivity Measurement in Regulated Industries, (Academic Press, New York) pages 172-218.

Horrace, William C. and Peter Schmidt (1996), "Confidence Statements for Efficiency Estimates from Stochastic Frontier Models," *The Journal of Productivity Analysis*, 7, 257-282.

Illinois Commerce Commission (2010), Dockets 09-0306, 09-0307, 09-0308, 09-0309, 09-0310, 09-0311, Order, April 29.

Kaufmann, L., et. al. (2006), *International Experience with Incentive Regulation and Implications for German Energy Regulation: Report to the Bundesnetzagentur*, February.

Kaufmann, L., et. al. (2008) *Calibrating Rate Indexing Mechanisms for Third Generation Incentive Regulation in Ontario: Report to the Ontario Energy Board*, February.

Kaufmann, L., et. al. (2013) *Empirical Research in Support of Incentive Rate Setting in Ontario*, Report prepared by PEG Research for the Ontario Energy Board, Madison, May.

Lancaster, Tory (2000), “The Incidental Parameter Problem Since 1948,” *Journal of Econometrics*, 95:2, 391-413, doi:10.1016/S0304-4076(99)00044-5.

Lawrence, D., T. Coelli, and John Kain (2014), *Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs*, Report by Economic Insights for the Australian Energy Regulator, Eden, November.

Lowry, M.N. et. al. (2008), *Benchmarking the Costs of Ontario Power Distributors*, March.

Lowry, M.N. et. al. (2014), *Benchmarking PS Colorado’s O&M Revenue Requirement*, June.

Lowry, M.N. et. al. (2014), *Database for Distribution Network Services in the US and Australia,* Report prepared by PEG Research for the Australian Energy Regulator, Madison, August.

Magnus, J.R. (1978), “Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix,” *Journal of Econometrics*, 7, pages 281-312.

Meyrick and Associates (2003), *Regulation of Electricity Lines Businesses, Analysis of Lines Business Performance – 1996-2003,* Report Prepared for Commerce Commission, Wellington, New Zealand, December.

Meyrick and Associates (2007), *Electricity Distribution Business Productivity and Profitability Update,* Report Prepared for Commerce Commission, Wellington, New Zealand, December.

New Zealand Commerce Commission (2007), *Regulation of Electricity Lines Businesses Targeted Control Regime*, Threshold Reset 2009, December.

New Zealand Commerce Amendment Act 2008 (Public Act 2008 No. 70), Part 4, Subpart 6, Section 53P, Subsection 10.

Neyman, J. and Elizabeth L. Scott (1948), “Consistent Estimates Based on Partially Consistent Observations,” *Econometrica*, 16:1, 1-32, 2014, <http://www.jstor.org/stable/1914288>.

Office of Gas and Electricity Markets (2004), *Electricity Distribution Price Control Review: Final Proposals,* November.

Office of Gas and Electricity Markets, *Electricity Distribution Price Control Review: Initial Proposals*, June 2004, p. 66.

Office of Gas & Electricity Markets (2009), *Electricity Distribution Price Control Review: Final Proposals,* December.

Office of Gas and Electricity Markets (2009), *Electricity Distribution Price Control Review: Final Proposals - Allowed Revenue – Cost Assessment*, December.

Office of Gas & Electricity Markets (2014), *RIIO-ED1: Final Determinations for the Slow-Track Electricity Distribution Companies, Business Plan Expenditure Assessment,* November.

Office of Gas & Electricity Markets (2014), *RIIO-ED1: Final Determinations for the Slow-Track Electricity Distribution Companies, Overview,* November.

Office of Gas & Electricity Markets (2009), *Electricity Distribution Price Control Review: Final Proposals - Allowed Revenue: Cost Assessment appendix,* December.

Office of the Regulator General in Victoria (2000), *Electricity Distribution Price Determination 2001-2005: Statement of Purpose and Reasons*, September.

Ontario Energy Board (2005), *2006 Electricity Distribution Rate Handbook*, RP-2004-0188, Report of the Board, May.

Ontario Energy Board (2005), *2006 Electricity Distribution Rate Handbook*, May.

Vermont Public Service Board (2010), Order in Docket 7585, April 16.