STATISTICAL BENCHMARKING FOR ALBERTA REGULATION

Pacific Economics Group, LLC
Economic and Litigation Consulting

STATISTICAL BENCHMARKING FOR ALBERTA REGULATION

Mark Newton Lowry, PhD
Partner

13 October 2006

PACIFIC ECONOMICS GROUP

22 East Mifflin, Suite 302
Madison, Wisconsin USA 53705
608.257.1522   608.257.1540 Fax

Table of Contents

Executive Summary

The Alberta Energy and Utilities Board has in recent years evinced growing interest in the statistical benchmarking of jurisdictional energy utilities. Electric utilities in the province are concerned that any use of benchmarking in regulation be just and reasonable. Pacific Economics Group is writing a paper for the Canadian Electricity Association on the responsible use of benchmarking in power distribution regulation. FortisAlberta has retained us to prepare a version of this paper that is relevant to Alberta's situation.

Introduction to Benchmarking

Statistical benchmarking is an approach to performance evaluation in which comparisons are made to benchmarks that represent external performance standards. Statistics on utility operations can be used to calculate benchmarks and draw conclusions from benchmark comparisons. Index and econometric methods are most widely used for this purpose in North America.

A fundamental result of benchmarking science is that differences between utilities in costs and other performance indicators of interest to regulators depend in part on differences in external business conditions. The cost performance of a company is thus a matter of the cost that it achieves given the business conditions that it faces. Benchmarks must reflect external business conditions if they are to reflect a chosen performance standard faithfully. The identification of relevant business conditions and a consideration of their impact on performance indicators are therefore important tasks in a responsible benchmarking study.

Cost theory plays an essential role in rigorous cost benchmarking. It suggests that benchmarking studies must take careful account of differences in the workloads of utilities and in the input prices that they face. Theory also suggests that benchmarking is generally more difficult for comparatively narrow cost categories such as distribution labour.

Statistical theory is useful for assessing the accuracy of benchmarking studies. Statistical tests of efficiency hypotheses can be developed for both econometric and indexing methods. They suggest that the accuracy of benchmarking is generally aided by a large set of standardized, quality data. A data set with these attributes is not currently available in Canada.

Special challenges are encountered in the benchmarking of capital cost. These include the need to control for differences in the depreciation practices and the system age of utilities. Methods designed to address these challenges require many years of capital cost data. Such data are not readily available for Canadian utilities.

Application: Power Distribution

Power distribution services have to date been the most common focus of regulatory benchmarking around the world. Research has identified numerous drivers of power distribution cost. These business conditions can vary between utilities and must be recognized in a responsible benchmarking study. Unfortunately, quality data are not always available for important cost items and business conditions.

The benchmarking of distribution service quality involves many of the same challenges as distribution cost benchmarking. These include inconsistencies in data reporting and the need to control for numerous business conditions that vary between service territories. Service quality has to date received far less attention from benchmarking practitioners as has cost.

International Benchmarking

Special challenges are encountered in benchmarking when international data are used. Utilities in different countries often provide different services and have different cost accounting practices. International comparisons of input prices are difficult to make accurately. Data on the operations of U.S. electric utilities, while abundant and of high general quality, have important limitations in benchmarking applications. Important data on costs and cost drivers are not readily available in the public domain. Guidelines for the categorization of data between transmission and distribution services are vague. Data for many companies have been compromised in recent years by the transfer of assets between transmission and distribution.

Role of Benchmarking in Regulation

Benchmarking has a number of potential uses in utility regulation. It can be used to inform decisions concerning the initial rates and the rate adjustment mechanisms of multiyear regulatory plans. The specific use of benchmarking in rate setting can range from formulaic mechanisms to the use of results only to help identify areas where a more intensive traditional review is needed.

The precedents for the use of statistical benchmarking in regulation around the world are mixed. While used fairly extensively in Western Europe and New Zealand, it has played little role to date in North America. A notable exception is Ontario, where regulators have jurisdiction over numerous power distributors. Benchmarking has fallen out of favour in Australia after controversial experiments. Where benchmarking is pursued around the world, the focus is typically on broad-based performance indicators such as total cost and total operation, maintenance, and administration ("OM&A") expenses.

Advantages of benchmarking in the regulatory arena include its ability to reduce the cost of regulation. Benchmarking can also improve operating performance by strengthening performance incentives and reducing concerns about cross-subsidization that can prompt regulators to discourage efficient diversification. The chief disadvantage of benchmarking is its potential to reduce the quality of regulatory decisions and raise operating risk.

The balance between these effects is a key consideration in deciding whether to use benchmarking. Responsible regulators are more likely to use benchmarking where its advantages clearly outweigh its disadvantages. The potential for benchmarking to improve regulation is greatest where regulators have jurisdiction over numerous utilities; have limited experience with traditional rate regulation; and have access to abundant, standardized, and high quality data.

Regulators have several levers at their disposal to control the balance of risk and return when benchmarking is introduced. Risk can be reduced by making rates less sensitive to benchmarking outcomes. Expected return can be bolstered by giving utilities a chance to benefit from superior returns to counterbalance the risk of penalties for poor performance. Regulators should also consider using awards to the exclusion of penalties.

To the extent that there are concerns about the accuracy of benchmarking, our analysis suggests that its use in ratemaking should be more limited. For example, a first generation benchmarking study could be deemed largely experimental and have a limited role in ratemaking.

In choosing a benchmarking method a reasonable balance must be struck between accuracy and cost. A benchmarking method that generates bad appraisals can violate the just and reasonable standard and, by raising operating risk, can materially offset any benefits from net cost savings in the regulatory process. It therefore makes sense to use relatively sophisticated benchmarking methods in the regulatory arena.

Implications for Alberta Regulation

Our analysis makes it possible to make some constructive comments concerning the use of benchmarking in Alberta regulation. Performance incentives may be strengthened, and unusually bad or good performances may be discovered. However, benchmarking has a limited ability to improve the efficiency of Alberta regulation since the Board does not, like its Ontario counterpart, have jurisdiction over dozens of utilities.

Methods will have to be improved considerably before benchmarking can play an important role in rate-setting. For example, it is difficult to draw statistically significant conclusions concerning operating performance from simple unit cost comparisons that are based on the small amount of data that can be gathered from Alberta utilities. Better benchmarking will require better statistical methods and data drawn from other Canadian provinces or the United States. Good benchmarking studies are costly, as are the revisions in Alberta data reporting requirements that are needed to achieve the requisite standardization for ex-provincial comparisons. It will be difficult to benchmark capital cost accurately without years of historical plant addition data.

The development of a Uniform System of Accounts would be a constructive step towards data standardization. However, it should be recognized that some of the micro level cost data typically gathered in such accounts are difficult to use in regulatory benchmarking. That is because it is especially difficult and costly to accurately benchmark such indicators. Standardized data on the operations of utilities in other Canadian provinces are, unfortunately, not yet available in the public domain. While more abundant data of good quality are available in the United States, these data have noteworthy limitations in international benchmarking work. The limitations include inconsistencies in the classification of transmission and distribution, the difficulty of making accurate international price comparisons, and the non-availability in the public domain of important performance indicators and business condition variables.

If the AEUB intends to use benchmarking in rate-setting it should do so in a way that preserves the balance of operating risk and return. Balance is encouraged by good scientific methods. Benchmarking results should be interpreted cautiously. Rate adjustments may not be desirable in early benchmarking efforts. If and when benchmarking is used in rate setting, rewards should be possible for superior performance.

1. Introduction

Statistical benchmarking has in recent years become an accepted tool in the assessment of utility performance. Managers look to benchmarking studies for indications of how well their companies are doing. Benchmarking is also playing a growing role in regulation. Such studies have, for example, been used to inform decisions concerning the initial rates and the rate adjustment mechanisms of multi-year regulatory plans.

Benchmarking is facilitated by the extensive data that utilities report to regulators and industry associations. Accurate performance appraisals are nonetheless challenging. For example, there are important differences between companies in the services provided, the prices of inputs used in service provision, and in other business conditions that influence their cost. The sample of quality, standardized data available for benchmarking is sometimes small and data on key variables needed for benchmarking are sometimes unavailable.

The Alberta Energy & Utilities Board ("AEUB") has shown a growing interest in the statistical benchmarking of jurisdictional utilities. It has encouraged cost comparisons between utilities in the province and is considering the development of a uniform system of accounts ("USA") and associated minimum filing requirements ("MFRs") that can facilitate benchmarking.

Pacific Economics Group ("PEG") is a leading practitioner of energy utility benchmarking and has advised dozens of clients on benchmarking issues. We have more than forty man-years of experience in the field of performance measurement and pioneered the use of scientific benchmarking in U.S. regulation. We have benchmarked power generation, transmission, distribution, customer, and administrative and general services, bundled power service, and gas distribution.

Most of our benchmarking work has been conducted using FERC Form 1 and other data that conform to the U.S. USA. However, our practice is international in scope and has to date included projects in twelve countries. Our Canadian clients have included the Canadian Electricity Association, the Ontario Energy Board ("OEB"), and well known electric and natural gas utilities. We have on several occasions undertaken studies that compare the cost performance of U.S. utilities to those in Canada and other countries.

Alberta's electric utilities are understandably concerned that any use of benchmarking in regulation be just and reasonable. FortisAlberta has retained PEG to prepare a white paper on benchmarking for electric utility regulation and its potential use in AEUB proceedings. Our paper builds on a report that we are writing for the Canadian Electricity Association.

Here is the plan for the paper. An introduction to statistical benchmarking is provided in Section 2, which includes discussions of salient benchmarking methods. There follows in Section 3 a discussion of the challenges encountered in benchmarking power distribution. The use of benchmarking in regulation is considered in Section 4. The paper concludes with a discussion of the prospects for statistical benchmarking in Alberta regulation.

2. An Introduction to Benchmarking

In this section, we consider some important benchmarking concepts. The benchmarking methods most widely used in North America are explained. The discussion is largely non-technical.

2.1 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 embodies a performance standard and can be used as a point of comparison in performance appraisals.

A quantitative benchmarking exercise commonly involves one or more gauges of activity. These are called, variously, comparators and performance indicators. The values of the performance indicators achieved by an entity under scrutiny are compared to benchmark values that reflect performance standards. Given information on the cost of a utility and a certain cost benchmark we might, for instance, measure its cost performance by taking the ratio of the two values.

$$\text{Cost Performance} = \text{Cost}^{\text{Actual}}/\text{Cost}^{\text{Benchmark}}$$

Benchmarks are often developed using data on the operations of agents that are involved in the activity under study. Statistical methods are useful in both the calculation of benchmarks and in the comparison process. An approach to benchmarking that prominently features statistical methods is called statistical benchmarking.

Various performance standards can be used in statistical benchmarking. These standards often reflect statistical concepts. For example, one sensible standard is the average performance of the utilities in the sample. Alternative standards include the apparent best or frontier performance in the sample and the performance that would define the margin of the top quartile of performers.

2.2 External Business Conditions

For costs and many other kinds of performance variables, it is widely recognized that differences in the values of the variables that companies achieve depend partly on differences in operating efficiency and partly on differences in external business conditions. In cost research, these conditions are called cost drivers. An external business condition is a condition of the operating environment that a firm cannot control. In the electric utility industry, examples include the number of customers served and the market prices of labour, capital equipment, and other production inputs. Customer service expenses will, for example, vary with the number of customers served.

The cost performance of a company depends on the cost that it achieves given the external business conditions that it faces. Benchmarks must therefore reflect external business conditions if they are to reflect a chosen performance standard faithfully. This helps to explain why the identification of relevant business conditions and consideration of their impact on performance variables are important tasks in a responsible benchmarking study.

2.3 Contributions from Cost Theory

Economic theory is useful in identifying cost drivers and controlling for their influence in benchmarking. We begin by positing that the actual cost incurred by a company is the product of the *minimum achievable* cost and an efficiency factor.[1] The goal of cost benchmarking is then to accurately estimate the efficiency factor. Consider next that, under certain reasonable assumptions, cost functions exist that relate the minimum cost of an enterprise to external business conditions in its service territory. Two kinds of cost functions yielded by this theory are useful in benchmarking.

One is the *total* cost function in which the minimum total cost of an enterprise is a function of the prices of production inputs, output quantities, and variables representing

---

[1] Minimum achievable cost is a hypothetical notion and cannot be precisely calculated for specific utilities.

miscellaneous other business conditions. The latter group of variables is sometimes conveniently called “Z” variables.

The theory allows for the existence of *multiple* output variables. This is important because it is often impossible to accurately measure the workload of a utility using only one output variable. The cost of power transmission, for instance, depends as much or more on peak demand as it does on the volume delivered. It is also noteworthy that the theory allows for the possibility that numerous business conditions other than input prices and output quantities can affect the minimum cost of service.

Regulators considering the appropriate revenue requirement of a company often have special interest in certain subsets of the total cost of service. Examples include operation, maintenance, and administration (“OM&A”) expenses (sometimes called “opex”) and even more “micro” categories such as distribution labour expenses. The interest in these expenses is due in part to the fact that they are subject to greater control by utilities in the short run than are capital costs.

When the focus of benchmarking is a subset of total cost, *restricted* cost functions are useful for identifying the full range of relevant cost drivers.[2] In such functions the minimum cost of a group of inputs depends on the general run of prices of those inputs and on output quantities and other business conditions. It depends, additionally, on the amounts of *other inputs* that the company uses. The existence of the other input variables in restricted cost functions means that a fair appraisal of the efficiency with which a utility uses a certain class of inputs must consider the amounts of *other* inputs that it uses.

This result is important for several reasons. One is that there are inconsistencies in the manner in which utilities classify costs. Utilities may, for example, differ in the way that they categorize certain expenditures between administrative and direct operating expenses.

Another reason that the result matters is that opportunities exist for the substitution of certain inputs in the production process. Suppose, for example, that the focus of inquiry is opex. It is then germane that the minimum level of opex depends on

---

[2] For an example of the use of restricted cost functions in empirical research see Robert Halvorsen and Tim R. Smith, “Substitution Possibilities for Unpriced Natural Resources: Restricted Cost Functions for the Canadian Metal Mining Industry”, *The Review of Economics and Statistics*, Vol. 68, No 3 (Aug. 1986), pp. 398-405.

the capital inputs that the company uses. A firm may, for example, have unusually high opex because its facilities are in an advanced stage of depreciation so that it is using comparatively little capital. Suppose, alternatively, that the focus of benchmarking is the efficient use of labour. Economic theory suggests that the minimum amount of labour that a company uses depends on its use of other, non-labour OM&A inputs as well as the amount of capital it uses. A utility may have an unusually small labour force not because it is especially efficient in its use of labour but because it has relatively new facilities and/or outsources a lot of its OM&A activities. By the same token, a company with high labour costs might do very little outsourcing.

One complication that benchmarkers encounter in trying to control for the usage of capital inputs is the measurement of that usage. As a practical matter, it isn't always possible to measure capital quantities accurately. However, variables can sometimes be computed that represent important characteristics of the capital stock that influence opex. For example, one might employ an indicator of the age of a system such as the number of customers added in the last ten years.

2.4 Benchmarking Methods

In this section we discuss at some length the two approaches to benchmarking that are widely used in North America: econometric modeling and indexing. The econometric approach is discussed first to establish a context for the appraisal of the index approach.

2.4.1 Econometric Modeling

Econometric Cost Research

Relationships between the costs of utilities and the business conditions that they face can be estimated using econometric methods. In such an exercise, a specific mathematical form must be chosen for the cost function. The impact of business conditions on cost depends on the form chosen and on the values of model parameters.[3]

---

[3] Here is a simple example of a cost function for power distribution that conforms to cost theory:

$$C_{h,t} = a_0 + a_1 \cdot YN_{h,t} + a_2 \cdot W_{h,t} \qquad [1]$$

The various alternative forms include the linear, the double log, and the translog. These forms vary in the flexibility with which they capture relationships between costs and cost drivers. Flexible functional forms are generally preferable. Suppose, for example, that economies of scale are exhausted at a certain level of output. We would then desire a functional form that permits the elasticity of cost with respect to output to increase with the level of output.

A branch of statistics called econometrics has developed procedures for estimating the parameters of economic models using historical data.[4] For example, cost model parameters can be estimated econometrically using historical data on the costs incurred by a group of utilities and the business conditions they faced.[5] The sample used in model estimation can be a time series consisting of data over several years for a single firm, a cross section consisting of one observation for each of several firms, or a panel data set that pools time series data for several companies.

Econometric research involves certain critical assumptions. The most important assumption, perhaps, is that the values of some economic variables (called dependent or left-hand side variables) are functions of certain other variables (called explanatory or right hand side variables) and error terms. In an econometric cost model, cost is the dependent variable and the cost drivers are the explanatory variables. The explanatory variables are generally assumed to be independent in the sense that their values are not influenced by the values of dependent variables.[6]

The error term in an econometric cost model is the difference between actual cost and the cost that is predicted by the model. It reflects imperfections in the development of the model. The imperfections may include any or all of the following: the mismeasurement of cost and the external business conditions, the exclusion from the

---

Here for any firm h in year t, the variable $YN_{h,t}$ is the number of customers that the company serves. It quantifies one dimension of the work that it performs. The variable $W_{h,t}$ is a measure of the general run of wages in the service territory. The wage rate and number of customers are the measured business conditions in this cost function. The terms a_0, a_1, and a_2 are model parameters. The function in relationship [1] has a linear form.

[4] The act of estimating model parameters is sometimes called regression.

[5] A positive estimate for parameter a_1 in equation [1], for instance, would reflect the fact that the costs reported by sampled companies tended to be higher the greater were the number of customers that they served.

[6] In the simple cost model described in equation [1], for instance, we would assume that the number of customers that a utility serves and the price that it faces for labour are not influenced by its cost.

model of relevant business conditions, and the failure of the model to capture the true form of the functional relationship. Error terms are a formal acknowledgement of the fact that the cost model is unlikely to provide a full explanation of the variation in the costs of sampled utilities.

It is customary to assume that error terms are random variables with probability distributions that are determined by additional coefficients, such as mean and variance. This stochastic specification is useful in selecting business conditions for cost models. Specifically, tests 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 is rejected. Statistical significance is a sensible criterion for the inclusion of variables in cost models. Statistical significance is more likely to be confirmed when the econometric work involves a large and varied sample.

Estimation Procedures

A variety of estimation procedures are used in econometric research. The appropriateness of each procedure depends on the assumptions that are made about the distribution of the error terms. The estimation procedure that is most widely known, ordinary least squares, is appropriate if the distribution of the error term is quite simple. Another class of procedures, called generalized least squares ("GLS"), are appropriate under assumptions of more complicated error specifications. One example is distributions in which the error terms for different companies in the sample have different variances. Still another approach to estimation, called stochastic frontier analysis ("SFA"), focuses on the fact that the error terms include inefficiency factors and therefore do not have symmetric distributions.

Estimation procedures that address *several* of the error term issues that are routinely encountered in utility benchmarking are not readily available in commercial econometric software packages such as Gauss and Stata. They require, instead, the development of customized estimation programs. While the cost of developing sophisticated estimation procedures that are tailored for benchmarking applications is sizable, the incremental cost of applying them to different utilities is typically small once they have been developed.

Cost Predictions

A cost model fitted with econometric parameter estimates may be called an econometric cost model. We can use such a model to predict a company's cost given local values for the business condition variables. These predictions are econometric benchmarks. They can be made for a historical period or a hypothetical test period.[7]

Performance Standards

The estimation procedure influences the performance standard that is embodied in the model predictions. Suppose, for example, that we choose a GLS procedure. It can then be shown that since these procedures do not explicitly account for the fact that the error terms are asymmetrically distributed, the predictions generated by the resultant cost model embody a *sample average* efficiency standard.[8] SFA procedures, on the other hand, generate benchmarks that reflect a frontier standard of operating efficiency.

The notion of minimum cost considered in SFA research is of a *short run* character. Firms can, in the short run, incur a cost that is considerably below the cost that is sustainable in the long run. An example from the business of power transmission is the deferral of tree trimming and other maintenance expenses. In the long run, utilities that defer maintenance will experience service quality deterioration. A benchmarking model of opex that is estimated using a frontier estimation procedure such as SFA might then effectively compare the opex efficiency of a subject utility to that of utilities that have deferred maintenance expenditures.

Capital cost provides another example of the short run/long run issue. Plant investments in the electric utility industry are commonly useful for 30-50 years. The

---

[7] Suppose, for example, that we wish to benchmark the cost of a certain hypothetical utility, Northern Electric. Returning to our example, we might predict the cost of Northern in period *t* using the following model.

$$\hat{C}_{Northern,t} = \hat{a}_0 + \hat{a}_1 \cdot YN_{Northern,t} + \hat{a}_2 \cdot W_{Northern,t}.$$

Here $\hat{C}_{Northern,t}$ denotes the predicted cost of the Company, $YN_{Northern,t}$ is the number of customers it serves, and $W_{Northern,t}$ measures its wage rate. 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

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

[8] See Appendix section A.1 for further discussion.

value of an investment in plant is commonly treated as depreciating over the service life. The growth patterns of utilities vary. In comparing two power distributors that serve 500,000 customers we might find, for example, that one of the companies had added 100,000 customers in the last ten years, whereas the other had added only 10,000. It is quite possible for this reason alone that utilities serving the same level of output have different levels of capital cost. A benchmarking model of capital cost that is estimated using SFA might then effectively compare the capital cost efficiency of any subject utility to the capital cost efficiency of utilities with highly depreciated rate bases.

Another problem with the use of a frontier performance standard is that it is unusually sensitive to irregularities in the data. As we discuss further in Section 3 below, such irregularities are frequently encountered in statistical benchmarking work. Efficiency comparisons using frontier cost performance standards are much more sensitive to data irregularities than are efficiency comparisons using a sample average performance standard.

Accuracy of Benchmarking Results

A cost prediction like that generated in the manner just described is our best *single* guess of the Company's cost given the business conditions it faces. This is an example of a point prediction. Such predictions are likely to differ from the true benchmark, which accurately embodies the desired standard and controls for the impact of external business conditions.

One potential source of inaccuracy is the values of the parameter estimates that measure the impact of external business conditions on cost. Another is the ability of the explanatory variables to accurately measure business conditions. A third is the extent to which the model captures the form of the relationship between business conditions and costs. Still another is a failure of the model to include all relevant business conditions.

Statistical theory provides useful guidance regarding the extent of inaccuracy. One important result is that an econometric cost model can yield *biased* predictions of the true benchmark if relevant business condition variables are excluded from the model. A model used to benchmark the opex of a rural power distributor might, for example, yield a value for the benchmark that is below its true value (and is thus excessively challenging) if it failed to include variables that properly represent the extensiveness of a distribution system and the magnitude of rural cost management challenges such as forestation. It is therefore desirable to include in an econometric benchmarking model all business conditions which are believed to be relevant, for which data are available at reasonable cost, and which have plausible and statistically significant parameter estimates. Even when an econometric benchmarking model is unbiased it can be imprecise, yielding values that are sometimes too high and on other occasions too low.

Statistical theory provides the foundation for the construction of confidence intervals that represent the full range of possible cost model predictions that are consistent with the data at a given level of confidence. These are readily constructed from the statistical results of an econometric run. A confidence interval is wider the greater is the uncertainty about the true benchmark level. In general, it can be shown that 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 sample companies are not varied
- 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 sample of good operating data. When the sample is small, it will be difficult to identify all of the relevant cost drivers and the appropriate functional form. It follows that it will generally be preferable to use panel data instead of a single cross section of data when these are available.

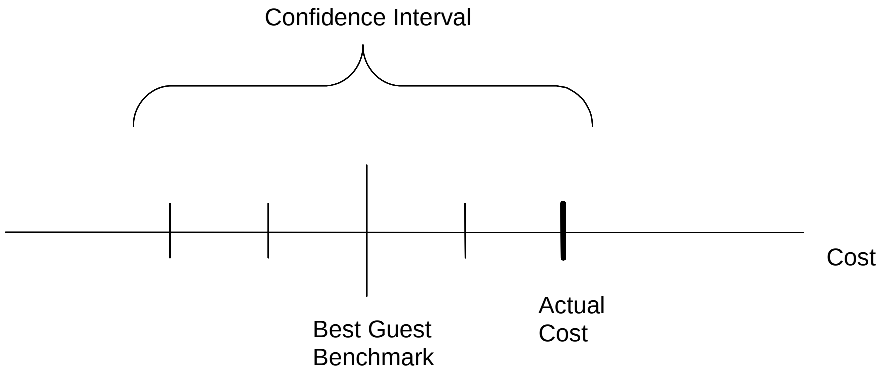
These findings are especially worthy of note in Canada, where large samples of quality data on utility operations are not readily available. The number of utilities in Canada is not large. Moreover, a lengthy series of data is unavailable for most individual utilities due to such circumstances as industry restructuring, utility mergers, and the prevalence of government enterprises.

To grasp the severity of the problem, suppose that we have available for benchmarking data on the operations of fifteen Canadian utilities over three years. In that event we would have forty-five observations. A sample of this size is less than a tenth the size of the samples that are readily available in the U.S. The confidence intervals generated by an econometric model based on available Canadian data are therefore likely to be wider than those based on U.S. data.

Notice also that the precision of an econometric benchmarking exercise is actually *enhanced* by using data from companies with diverse operating conditions. For example, we will obtain a better estimate of the impact of line length on cost if we include in the sample companies that, like Toronto Hydro Electric System (THES), have *high* customer density as well as data for companies that, like Sask Power, have low customer density.

Testing Efficiency Hypotheses

Confidence intervals developed from econometric results do much more than provide indications of the accuracy of a benchmarking exercise. In particular, they permit us to test hypotheses regarding cost efficiency. Suppose, for example, that we use a sample average cost standard and compute the confidence interval that corresponds to the 90% confidence level. It is then possible to test the hypothesis that the company is an average cost performer. If the company's actual cost exceeds the benchmark generated by the model but nonetheless lies within the confidence interval (as in the figure below), this hypothesis cannot be rejected. In other words, the company is not a *significantly* inferior cost performer. Suppose, alternatively, that the company's cost is below the cost predicted by the model by enough to be outside the confidence interval. We may then conclude that it is a *significantly superior* cost performer.



An important advantage of efficiency hypothesis tests is that they take into account the accuracy of the benchmarking exercise. As we have just discussed, there is uncertainty involved in the prediction of benchmarks. These uncertainties are reflected in the confidence interval that surrounds the point estimate (best single guess) of the benchmark value. The confidence interval will be greater the greater is the uncertainty regarding the true benchmark value. If uncertainty is great, our ability to draw conclusions about operating efficiency is hampered. For example, the small samples of data that are readily available in Canada would make it more difficult to draw conclusions about operating efficiency from an econometric model that is based on these data.

2.4.2 Index-Based Approaches to Benchmarking

The index-based approach to benchmarking is commonly employed by utilities in internal reviews of operating performance. Benchmarking indexes are also used in the regulatory arena. 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 in one respected dictionary 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)".[9] In benchmarking, indexing involves the

---

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

calculation of ratios of the values of performance variables for a subject utility to corresponding values of the variables among a sample of utilities. The group of companies represented in the sample is called, variously, a cohort or a peer group.[10]

These concepts are usefully illustrated by the process through which decisions are made to elect athletes to the NHL's Hockey Hall of Fame. Statistical benchmarking undoubtedly plays a major (albeit informal) role in player selection. Goalies, for example, are evaluated using multiple performance variables that include the goals-against average. The values achieved by Hall of Fame members like Ken Dryden of the Montreal Canadiens are useful benchmarks. These values reflect a Hall of Fame performance standard.

Economic indexes can be designed to summarize the results of multiple comparisons. Such summaries commonly involve the calculation of weighted averages of the comparisons. Consumer price indexes are familiar examples. These summarize the inflation (year to year comparisons) in the prices of hundreds of goods and services.

To better appreciate the advantages of complex indexes in benchmarking, recall from our discussion in Section 2 that economic theory allows for cost to depend on multiple output quantity variables and that multiple variables are often needed to accurately measure the workload of utilities. We might, then, wish to construct an output quantity index that is a weighted average of comparisons for several output measures. Suppose, by way of example, that we are benchmarking the power supply cost of a utility with a low load factor. It would be desirable in this case to consider its peak demand as well as its sales volume. If we separately calculate the company's cost per megawatt hour and per megawatt we would likely come up with two very different assessments. A final reckoning of performance then requires a sensible weighting of the assessments.

In a cost benchmarking application, it makes sense for the weights of an output quantity index to reflect the relative importance of the output measures as cost drivers. Econometric research is useful in this regard. We can, for example, use as the weight for

---

[10] The term cohort comes from the Latin word for one of the ten divisions of a Roman legion.

each measure its share in the sum of the econometric estimates of the output-related cost elasticities.[11]

Summary input price and quantity indexes can also be computed. We might, for example, compare the quantities of OM&A inputs used by a subject utility to those of a cohort using an index that involves weighted averages of the amounts of labour and non-labour OM&A inputs used. In the construction of input quantity indexes it is customary to use the corresponding cost shares to calculate weights. It can be shown that this approach to weighting best reflects the impact of input quantities on cost.

Unit Cost Indexes

Unit cost indexes are used to make unit cost comparisons. A simple example is the ratio of a company's cost per customer to the average cost per customer of a peer group. This can be stated, alternatively, as the ratio of a cost comparison to a comparison of the number of customers served.[12] In more sophisticated unit cost indexes, the workload comparison is a weighted average of several workload measures.

Unit cost indexes are, effectively, cost comparisons that control for differences between companies in one of the most important cost drivers: operating scale. This control permits us to use data for utilities with different workloads in evaluating cost performance. Unit cost is also of interest because it is the long run source of differences in the prices that utilities charge.

Despite these advantages, unit cost comparisons do not control for all of the cost drivers that are known to vary between utilities. Our discussion in Section 2 revealed that cost depends on input prices and miscellaneous other business conditions in addition to operating scale. The accuracy of unit cost benchmarking thus depends on the extent to which the cost pressures placed on the peer group by these excluded business conditions

---

[11] The elasticity of cost with respect to a certain business condition variable is the percentage change in cost that results from a one percent change in the value of the variable.

[12] Here is an example of a unit cost index for our hypothetical subject utility.

$$\frac{\left(\dfrac{Cost_{Northern}}{Customer_{Northern}}\right)}{\left(\dfrac{Cost_{Mean}}{Customers_{Mean}}\right)} = \frac{\left(\dfrac{Cost_{Northern}}{Cost_{Mean}}\right)}{\left(\dfrac{Customers_{Northern}}{Customers_{Mean}}\right)}$$

are similar on balance to those facing the subject utility. The appropriateness of the peer group is extremely important to the accuracy of a benchmarking effort that uses unit cost indexes. The ability to assemble a satisfactory peer group can be limited when the number of candidate peers with comparable data is small.

To better appreciate the peer group design challenge, consider a comparison of the unit costs of the power generation businesses of two utilities. This may control for differences in their operating scale. However, it would not control for differences in the prices that they paid for coal, natural gas, and other generation fuels. Prices of generation fuels vary considerably in different parts of Canada. The accuracy of a unit cost benchmark may thus depend on the extent to which the fuel prices paid by the two utilities are similar.

Excluded business conditions are even more problematic when the focus of unit cost indexing is a narrow cost category. In that event, we have seen that a good benchmark should take account of the amounts of other kinds of inputs that a company uses. Suppose, for example, that we compare the labour costs per customer of two utilities that have a markedly different reliance on outsourced services. In that event, the comparison is apt to be unfavourable to the company that doesn't do much outsourcing. It follows that in comparing unit labour costs, attention should be paid to differences in the extent to which candidate peers rely on outsourcing. This discussion suggests that, absent appropriate peer group controls, unit cost benchmarking will tend to be more accurate to the extent that the scope of costs under consideration is comprehensive. It will, for example, be easier to accurately benchmark *OM&A* expenses using unit cost indexes than it will be to accurately benchmark *labour* expenses.

Productivity Indexes

A productivity index is the ratio of an output quantity index to an input quantity index. It is used to make productivity comparisons. Many readers will think of productivity indexes as measures of *trends* in operating efficiency over time. However, they can also be designed to compare the efficiency *levels* of utilities at a point in time.

A simple example of a productivity index is the ratio of customers served per employee to the mean value of same for a peer group. This can be stated, alternatively, as the ratio of a customer comparison to an employee comparison.[13] In more sophisticated productivity indexes, the output comparison is a weighted average of several output measures.

A productivity comparison such as this can be shown to be the portion of a unit cost index comparison that is not due to differences in input prices. The unit cost of a utility will then compare more favourably to that of a peer group to the extent that its input prices are lower and its productivity is higher. This result helps to explain why productivity indexes are generally more accurate benchmarking tools than unit cost indexes. Productivity indexes are, effectively, comparisons of cost that control for differences in *two* sets of business conditions that vary between utilities and are major cost drivers: the amount of work performed and the prices paid for inputs. These controls make it possible to use data from a more diverse set of companies in choosing a peer group. Peer companies do not have to be of the same size and can, additionally, operate under different input price conditions.

Despite these advantages, productivity comparisons do not control for all of the important cost drivers that vary between utilities. For example, a comparison of the productivity of the power generation businesses of two utilities could control for differences in their operating scale and generation fuel prices. However, it would not control for differences in their access to sites that are suitable for low cost hydroelectric generation. It follows that the selection of a peer group is still important to the accuracy of a benchmarking study that is based on productivity indexes.

As we discussed above for unit cost indexes, excluded business conditions are apt to be a bigger complication to the extent that the focus of productivity indexing is a narrow input category. When the focus is narrow, we have seen that a good benchmark should take account of the amounts of other kinds of inputs that a company uses. Suppose, for example, that we compare the customers per employee of two utilities that

---

[13] Returning to our example, this can be expressed formulaically as

$$\frac{\left(\frac{Customers_{Northern}}{Employees_{Northern}}\right)}{\left(\frac{Customers_{Mean}}{Employees_{Mean}}\right)} = \frac{\left(\frac{Customers_{Northern}}{Customer_{Mean}}\right)}{\left(\frac{Employees_{Northern}}{Employees_{Mean}}\right)}$$

have a markedly different reliance on outsourced services. In that event, the comparison is apt to be unfair to the company that doesn't do much outsourcing.[14]

This problem can be finessed by considering a broader range of inputs in the productivity index. An index that compares productivity in the use of more than one input is called a multifactor productivity ("MFP") index. An MFP index that covers all inputs used by an enterprise is called a total factor productivity ("TFP") index.

Our discussion suggests that more comprehensive productivity indexes will generally yield more accurate benchmarking results. Consider, for example, the company that uses a lot of in-house labour and outsources very few tasks. Such a company is likely to have low labour productivity but will have high productivity in the use of other OM&A input. An MFP index covering all OM&A inputs can assess how things balance out.

Performance Standards

The cost performance indexes that we have discussed so far in this section embody a sample average standard of performance. Alternative standards can also be implemented. We can, for example, make calculations for each utility in the sample and then assess the apparent productivity shortfall between a subject utility and the utility with the best productivity ranking.

Frontier performance comparisons using indexes are, however, fraught with many of the same limitations as we discussed in the context of econometric modeling. The utilities with the best apparent productivity performance may, for instance, have achieved that status due to deferred maintenance. They may also have risen to the top due to data irregularities.

Statistical Tests of Efficiency Hypotheses

Statistical tests are generally not employed in index-based benchmarking but can be developed for regulatory applications. To better appreciate the possibilities, suppose that we are benchmarking the unit cost performance of a company using a cost per

---

[14] It follows that in comparing labor productivity, attention should be paid to differences in the extent to which candidate peers rely on outsourcing.

customer measure. The unit costs of the companies in the peer group may vary considerably due to either or both of variations between companies in the many excluded business conditions and the year-to-year volatility of the data for each company. We can then treat the data for the peer group as a sample drawn from a probability distribution that has an unknown mean and variance. The mean cost per customer is then an estimate and our best guess of the true mean of the population. A confidence interval can be constructed around the sample mean unit cost. A utility may be deemed to have an anomalous cost performance if its unit cost exceeds the upper bound of the confidence interval. The confidence interval will generally be wider --- making conclusions about efficiency more difficult to draw --- the larger and more varied are the data for the peer group.

2.4.3 Hybrid Approaches

Hybrid benchmarking approaches that employ elements of the econometric and indexing approaches are also possible. An example is the “comparators and cohorts” approach to benchmarking that the OEB is currently using to assess the operating efficiency of provincial power distributors. The methodology involves several steps.

1. A number of cost performance indicators were chosen.
2. For each such indicator, cost models were developed in which the variable (*e.g.* distribution OM&A expenses) was a function of certain business condition variables (*e.g.* the number of customers served). The parameters of the model were estimated statistically using confidential historical data on the costs incurred by Ontario distributors and the business conditions that they faced.
3. The parameter estimates obtained from the econometric work were employed in a statistical clustering analysis. This analysis identified, for each cost performance variable, cohorts of distributors with relatively similar values for the measured business conditions.
4. “Comparative diagnostic” variables of more micro character were calculated for each of the companies in each cohort.

5. The cost comparisons, together with the comparative diagnostics, are now being used by Board Staff to identify distributors with anomalously high costs.

2.5 Capital Cost

Capital inputs play important roles in utility operations. They are especially important in network businesses like power transmission and distribution. In these businesses, capital typically accounts for half or more of total cost. It follows that, in the long run, the success utilities have at holding down their costs depends greatly on their management of capital costs.

The cost of capital ownership has several components. One is the opportunity cost of having funds tied up in ownership. To the extent that the company borrows money, this is the interest that it must pay. To the extent that it secures financing in equity markets, this is the return on equity. Another important component of capital cost is depreciation. A third component of capital cost is taxes. The relevant taxes include income and property taxes and certain implicit taxes such as franchise fees.

The computation of depreciation and opportunity cost requires a valuation of utility plant. Two basic approaches to valuation can be used. One is book (historical cost) valuation. The other is current (replacement cost) valuation. Regulators must choose a method for calculating capital cost to establish revenue requirements. North American regulators commonly use book valuations of plant.[15]

Accurate benchmarking of the cost of any input generally requires a measure of the local input price. Accurate benchmarking of the cost of plant ownership requires, specifically, an estimate of the price of holding a unit of capital. These prices are sometimes called capital *service* (or rental) prices since prices for the rental of a unit of capital in competitive rental markets (*e.g.* those for real estate or automobiles) tend in theory to reflect the cost of owning a unit of capital. It can be shown that capital service prices reflect the cost of funds, depreciation and tax rates, and the cost of buying or building a unit of plant.

---

[15] Replacement valuations are used by regulators in some other countries, including Australia.

The benchmarking of capital cost involves special challenges. One is inconsistencies in the manner in which capital cost is reported. Companies differ most notably, perhaps, in the way that they calculate depreciation. Another problem is that the book valuation of plant used in regulatory accounts makes the reported net value of plant especially sensitive to the historical pattern of capital investment. Two utilities could thus own the same amount of plant, but one could have a lower net plant value because its plant is of older vintage.

A means of computing capital cost has been developed by scholars to help finesse these problems. This method is commonly employed in rigorous research on capital cost. The basic idea is to recompute the cost of capital using a standardized treatment of depreciation and historical data on net plant value in a certain benchmark year and on plant additions in subsequent years. The methodology involves the calculation of a capital quantity index using a perpetual inventory equation. The intent is to base capital cost calculation as much as possible on the plant *additions* data, which are less idiosyncratic.

The accuracy of this general approach to capital cost measurement is increased to the extent that the benchmark year is far in the past. In the electric power research of PEG that uses U.S. data, for instance, we use *1964* as the benchmark year. Computing past values of capital quantity indexes is complicated by past mergers and acquisitions involving sampled firms.

When this methodology is employed, data on capital cost and the amount of capital that utilities use is still sensitive to their patterns of plant additions over the years. For example, two utilities with the same operating scale and level of capital cost efficiency can still have different capital costs (and quantities) if one system has an average asset age of 20 years while the other has an average asset age of 30 years. This problem is just beginning to receive the attention that it deserves from benchmarking experts.

3. Application: Power Distribution

The challenge of accurate benchmarking is better appreciated by considering its application to a specific sector of the electric power industry. In this section we take an in-depth look at power distribution. We consider in turn the challenges encountered in benchmarking local delivery cost, distribution cost, and service quality.

3.1 Benchmarking Local Delivery

3.1.1 The Local Delivery Business

The typical LDC receives power in bulk from points on a high-voltage transmission grid and delivers it to consumers. Receipt commonly occurs at substations, where voltage is reduced from transmission to distribution levels. Power is in most cases delivered to end users at the voltage at which it is consumed.[16]

Continuous use of electric power is essential to the functioning of modern homes and businesses. Power storage and self-delivery are, additionally, generally not cost competitive with power produced in bulk and delivered by utilities. It follows from these circumstances that customers want local delivery capability to be continuous. The technology for providing continuous service requires a network in the sense of a system that is physically connected to end user premises.

Power flows to the customer through wire conductors. Other capital inputs used in local delivery include poles, conduits, station equipment, meters, vehicles, storage yards, office buildings, and information technology ("IT") inputs such as computer hardware and software. Local distribution companies ("LDCs") commonly operate and maintain such facilities and are also frequently involved in the construction of distribution plant. These activities require labour, materials, and services. Local delivery also typically requires a certain amount of power in the form of line losses. Opportunities are available to outsource many OM&A and construction activities. LDCs vary greatly in the extent of their outsourcing.

---

[16] However, some large volume customers perform their own voltage stepdowns. At the extreme, they may take delivery of power from the grid and bypass the distribution system entirely.

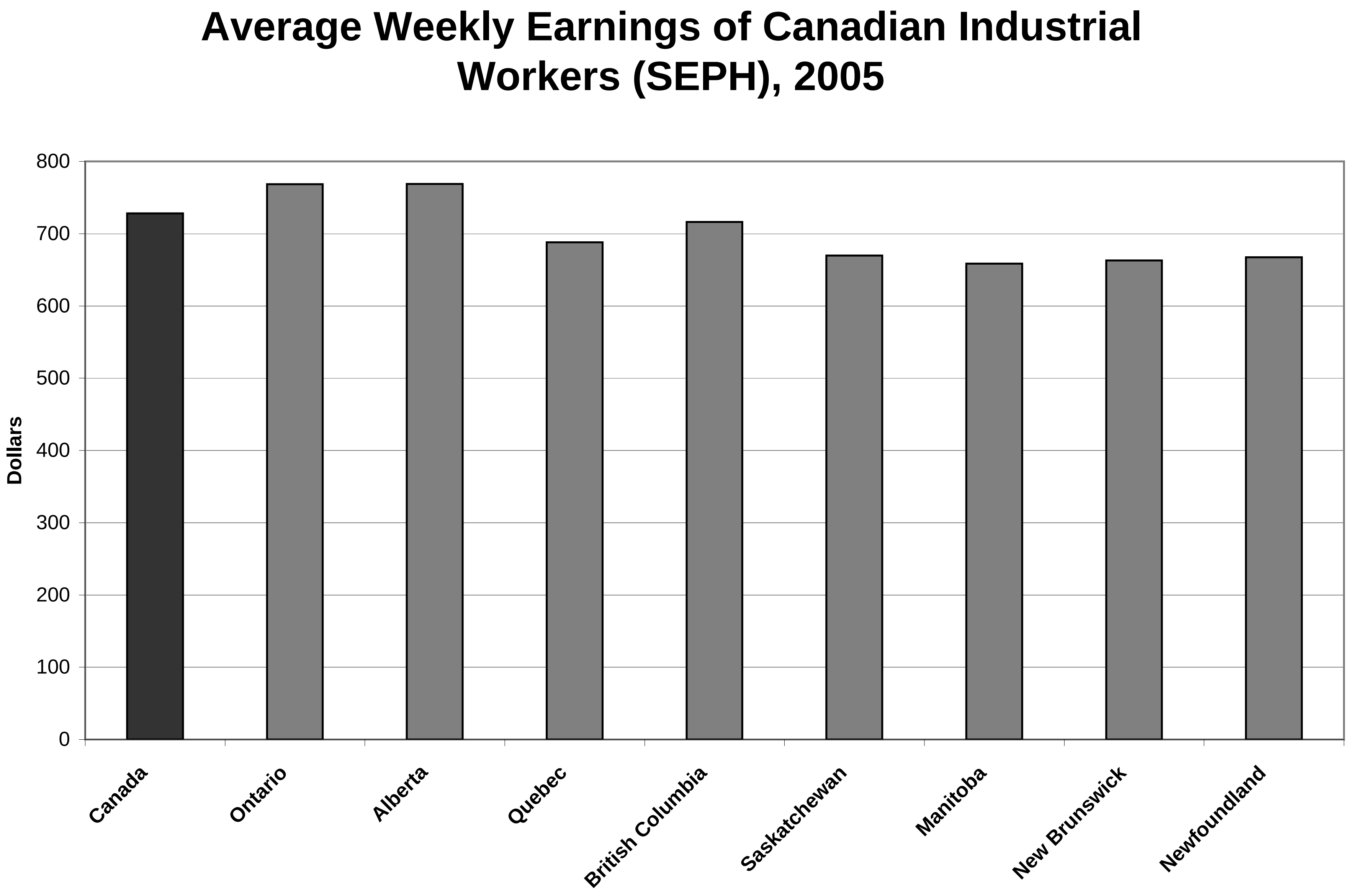
Local Delivery Cost

The total cost of local delivery service comprises OM&A expenses and the costs of plant ownership. At current input prices, capital inputs typically account for between 45 and 60 percent of the total cost of local power delivery and constitute the single most important input group. The exact cost share of capital depends on the age of a system and the manner in which plant is valued. The relative shares of labour and other OM&A inputs vary greatly.

This discussion suggests that prices for labour, capital, and other inputs are important drivers of power distribution cost. Input price differences are especially important in international benchmarking. Figure 1 displays information from Statistics Canada on the salaries and wages of industrial workers in selected provinces. Prices are highest in Ontario and Alberta and lowest in the prairie and maritime provinces.

Certain expenditures by LDCs have a periodic character. As one example, overhead line maintenance activities such as tree trimming do not have to be undertaken at the same level each year. As another, LDCs make capital investments in response to expected output growth. These investments, once made, may not require replacement for 30-50 years. The amount and cost of capital in a particular year therefore depend greatly on the historic pattern of output growth. For example, a distributor serving a region that grew much more rapidly in the 1960s than in recent years may today have a highly depreciated system and an unusually large need to make replacement investments.

Figure 1



SOURCE: The Survey of Employment Earnings and Hours (SEPH) is produced by Statistics Canada

Distribution Outputs

Cost theory suggests that the operating scale of a utility is an important cost driver. The outputs of a power distributor may be narrowly defined as measures of its operating scale that also serve as billing determinants. Three such measures are salient: the delivery volume, the peak load, and the number of customers served.

Service Packages

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

- One of the most important differences between LDC service packages concerns in their involvement in transformation of voltage from the transmission to the distribution level. Where transmission and distribution services are provided by separate companies, as is common overseas, policymakers have typically decided which kind of company provides this service. Where transmission and distribution services are provided by the same company, as is commonly the case in North America, the issue is how these services are *categorized*. In Canada, these substations are commonly classified as *transmission* facilities. In the United States, on the other hand, they are commonly classified as *distribution* facilities.
- Many power systems have lines with voltages that are intermediate between the extra high voltage lines used for long distance transmission and the low voltage lines used to deliver power locally. These lines are sometimes counted as transmission and sometimes as distribution facilities. For example, most British LDCs own and operate systems of 100+ kilovolt (“kV”) lines. Such lines would typically be counted as transmission facilities in North America.
- The United States USA has the further eccentricity of creating an ambiguous dividing line between transmission and distribution. The distribution system begins at the “entrance” to a consuming area. By this definition, subtransmission lines and substations that reduce voltage to

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 placed in the transmission category because they carry power to the entrance of multiple towns.

- The vagueness of accounting guidelines has, additionally, encouraged utilities to transfer assets between the transmission and distribution categories. Transfers have also been encouraged by FERC efforts to restructure the U.S. power transmission industry. These transfers have materially reduced the sample of quality benchmarking done in recent years.
- LDCs in some jurisdictions do not own all meters or service lines connected to end user premises. Victoria, Australia is an example. In such jurisdictions, data on the cost of owning and operating such facilities may be handled separately and may not be readily available.

Other Network Characteristics

Power distribution networks vary in a number of other respects that affect their cost.

- Systems vary widely in customer density. Density is highest in urban areas and is lowest in sparsely populated rural areas. All else equal, distribution cost is typically higher the lower is customer density. Areas of unusually low density in Canada include Manitoba and Saskatchewan. In cost research, system extensiveness is commonly measured by the number of line miles. This cost driver is sometimes treated as an output variable in benchmarking work due to its importance and its relevance to operating scale.
- There is marked diversity in the extent of distribution system undergrounding. Undergrounding generally raises the *total* cost of local delivery service but can lower local delivery *opex* due to the reduced need for line maintenance. Undergrounding is most common in the central

cities of major urban areas such as Calgary, Edmonton, Montreal and Toronto. Its prevalence in other areas depends greatly on public policy.

- The shape of distribution systems must conform to special features of the landscape. For example, distribution lines will typically go around sizable hills as well as lakes and other large water bodies. Distribution cost can be raised by such complications.
- The reliability of distribution services provided by utilities varies widely. Better reliability generally comes at a higher cost. The cost impact of quality is thus a valid issue in distribution benchmarking. There are special challenges in the estimation of the cost impact of quality. Despite its importance, empirical research on this topic is not well advanced.

Other Cost Drivers

Cost research by PEG and others has identified a range of additional business conditions that are drivers of local delivery costs.

- Distribution *opex* is generally *lower* the younger is the system. *Capital* cost is typically *higher* in a young system. The net effect of system age then depends on the relative magnitudes of the opex and capital cost effects. Our research to date has suggested that the *total* cost of power distribution is on balance *lower* in a younger power distribution system.
- Distribution cost is typically higher the greater is the degree of forestation in a service territory. An obvious reason is the greater need for tree-trimming and other maintenance expenses. Another is the greater difficulty in creating and accessing power line corridors.
- The rockiness of soil affects the cost of distribution pole installation.
- Another condition that affects the cost of power distribution is the number of gas customers that the company provides with distribution service. This presents opportunities for the realization of scope economies. Only one of the larger Canadian electric utilities is a gas distributor. However, combined gas and electric companies are fairly common in the U.S.

Examples from just across the border include Rochester Gas & Electric, Consumers Energy, and Montana Power.

3.1.2 Data Problems

Reporting Inconsistencies

Research has identified numerous inconsistencies in the manner in which LDCs report operating data. These problems tend to be especially marked where utilities have some discretion in cost reporting due to lax reporting guidelines and/or the inherent arbitrariness of cost allocations. We have already mentioned inconsistencies in the classification of power delivery operations between transmission and distribution. Another area of reporting inconsistency is the capitalization of OM&A expenses. An example of OM&A expenses that are capitalized by most utilities is those for plant construction labour. Areas where practices are more varied include work on software.

A third area where reporting inconsistencies tend to develop is the categorization of OM&A expenses. One issue is the breakdown between direct expenses and administrative and general ("A&G") expenses. The latter category of expenses, sometimes called corporate service expenses, is those that cannot be directly attributed to specific lines of business. Inconsistencies are also encountered in the allocation of direct expenses. An example from the United States is the grey area between billings and collections and customer service and information expenses.

Companies also differ in their practices for categorizing power delivery volumes. Some companies will break down the volumes of power delivered to business customers into "large" and a "small" category irrespective of the nature of the business. Others will break down volumes into a "commercial" and an "industrial" category irrespective of the volumes individual companies receive. Several U.S. companies have changed their approach to categorization over time.

Missing Data

Benchmarking is also complicated by the unavailability of important data. One major problem is the unavailability of good capital data. Adequate data for the calculation of standardized capital costs and quantities are not available for Canada or most other countries of the world. The United States is a prominent exception to this rule since detailed capital cost data have been reported there by major investor-owned utilities for decades.

While this is a cardinal advantage of U.S. data, other important data on U.S distribution operations are not available in the public domain. These include:

- Line miles[17]
- Line losses
- Reliability
- Peak demand
- Itemized data on distribution intangible plant
- Itemized data on demand-side management ("DSM") activities

3.1.3 International Benchmarking Challenges

Special challenges are encountered in the benchmarking of power distribution cost using international data. For example, certain complications discussed above are especially common in international benchmarking. These include, especially

- Differences in the services that LDCs provide
- Differences in accounting

Input price comparisons are especially important in international benchmarking due to the use of different currencies in different countries. They are also especially difficult to make accurately. A comparison of labour prices in the U.S. and Canada provides an instructive example. Here is a sampling of the problems that PEG has encountered in benchmarking Canadian utilities using U.S. data.

1. U.S. electric utilities no longer report labour costs per employee. Thus, price comparisons must be made using data from the U.S. and Canadian governments.
2. U.S. and Canadian utilities typically offer different benefit packages. Health care costs are lower in Canada, for instance, due to national health insurance. It follows that accurate U.S.-Canadian benchmarking comparisons require either the exclusion of pensions

---

[17] Line miles data were available in the past but may not be available prospectively.

and other benefits from the research or a time-consuming comparison of the total labour cost per employee.

3. The assessment of total labour cost per employee is complicated by the fact that a sizable share of the total benefit expenses reported by utilities is actually intended for retirees rather than current employees.
4. Good data on the salaries and wages of electric utility employees are available in the United States only for *production* workers. In Canada, on the other hand, these are available only for *all* workers.
5. It is also difficult to accurately compare prices for other O&M expenses and plant construction

Since extensive power distribution operating data are available in the U.S., it is also worthwhile to review some of the problems with this data that limit its usefulness in benchmarking.

1. Standardized, quality data are unavailable in the public domain for a number of important performance indicators and cost drivers.
2. Guidelines for transmission/distribution classification are vague.
3. Some utilities made sizable transfers in recent years of transfers of transmission and distribution assets and associated opex.

It may also be noted that the highly detailed micro cost indicators available on the FERC Form 1 are generally difficult to benchmark accurately, for reasons discussed above.

Notwithstanding these limitations, the U.S. data have some important advantages in power distribution cost benchmarking. Most data are of good quality, and are available for a large number of companies over many years. This facilitates the construction of peer groups with more than a handful of members. It also facilitates the development of econometric cost models of considerable sophistication. These models can shed light on the complexity of the benchmarking challenge.

3.1.4 Econometric Research on Distribution Cost Drivers

To better demonstrate the importance of various power distribution cost drivers, we present the results of a recent power distribution cost model developed by PEG using U.S. data. The principle source of data for the research was the Federal Energy

Regulatory Commission (FERC) Form 1. The services covered included customer services as well as local delivery services.

A model was developed that explains the total cost of power distribution and customer services.[18] Data for 77 companies were used in the estimation of model parameters. A total of 979 observations were used. The large size of the sample permitted the identification of numerous business conditions and experimentation with alternative functional forms.

Econometric results for the total cost function are reported in Table 1. Due to the form of the function the first order terms of the input prices and output quantities, as well as the parameters of the other business condition variables, are elasticities of the total distribution cost of the sample mean firm with respect to the basic variable. The first order terms are the terms that do not involve squared values of business condition variables or interactions between different variables. The table shades the parameter estimates of these terms for reader convenience.

Inspecting these results, it is first of all noteworthy that all of these parameter estimates are statistically significant and plausible. The model also has a high explanatory power. However, this is fairly common in utility cost research because the dominant source of variation in the sampled cost data is operating scale and this can be adequately if not perfectly measured.

The data supported the use of the number of customers served, the retail delivery volume, and the miles of distribution lines as output quantity variables. Cost was higher the higher were the values of all three variables. The positive sign for distribution line miles implies that the total cost of distribution is typically higher with the low customer density that is typical in more rural service territories.

---

[18] The full model consists of a total distribution cost function and cost share equations for labour, other O&M expenses, and capital.

Table 1

ECONOMETRIC COST MODEL FOR POWER DISTRIBUTION

VARIABLE KEY

L= Labor Price
K= Capital Price
N= Number Retail Customers
V = Retail Deliveries
M = Distribution Line Miles
OH = Percent of Distribution Plant that is Overhead
G= Number of Gas Distribution Customers
GN = 10 year customer growth
F = Forestation
RC = Percent Retail Deliveries that are Residential and Commercial
LF = Monthly Load Factor

Total Distribution Cost

EXPLANATORY VARIABLE	ESTIMATED COEFFICIENT	T-STATISTIC	EXPLANATORY VARIABLE	ESTIMATED COEFFICIENT	T-STATISTIC
L	**0.166**	**42.71**	**V**	**0.313**	**13.13**
LL	0.087	4.53	VV	1.298	14.23
LK	-0.066	-5.78	VM	-0.265	-4.68
LN	0.034	5.77			
LV	-0.042	-8.92	**M**	**0.218**	**12.22**
LM	-0.008	-1.91	MM	0.105	1.70
LOH	0.058	4.54	MF	0.021	3.87
LG	-0.001	-5.17			
LGN	-0.024	-8.02	**OH**	**-0.131**	**-2.74**
LFM	0.000	0.03			
LRC	-0.005	-0.56	**G**	**-0.006**	**-7.13**
LLF	0.051	3.60			
LTREND	-0.004	-9.39	**GN**	**-0.068**	**-5.66**
K	**0.585**	**101.47**	**RC**	**0.569**	**14.24**
KK	0.142	7.74			
KN	-0.148	-17.39	**LF**	**-0.140**	**-2.56**
KV	0.120	16.46			
KM	0.027	4.73	Trend	-0.015	-12.44
KOH	-0.137	-7.92			
KG	0.001	1.77	Constant	15.045	1019.27
KGN	0.041	10.03			
KFM	-0.003	-1.94			
KRC	0.069	4.88			
KLF	-0.099	-4.96			
KTREND	0.000	-0.81			
N	**0.427**	**15.64**			
NN	1.096	11.88			
NV	-1.114	-12.68			
NM	0.109	2.00			

Results for the other business condition variables are also informative.

- The positive sign for the system overheading variable suggests that systems with extensive undergrounding involve higher total cost.[19]
- The system age variable employed in the research was an estimate of the percentage of customers served that were added in the last twenty years. The negative sign of its parameter estimate suggests that the total cost of distribution should be lower the younger is the system.
- The positive sign for the forestation variable suggests that cost is generally higher the greater is the forestation of the service territory. Results from the cost share equations suggest, additionally, that forestation raises OM&A expenses more than capital costs.
- The negative sign for the percentage of customers served that are residential and commercial suggests that distribution service to these categories of customers involves higher cost.
- The positive sign for the non-contiguous service territory variable suggests that companies that serve scattered, non-contiguous areas can incur higher costs.
- The positive sign for generation and transmission OM&A expenses suggests that distributors that are not diversified into these businesses will incur higher costs.
- The positive sign for the competition variable suggests that service to retail access customers can involve higher costs.
- The data strongly supported the use of a flexible functional form for a distribution cost model.

These findings have important implications for the benchmarking of power distribution in Canada. Most notably, benchmarking studies that do not consciously control for differences between utilities in these business conditions may not be very accurate. Distribution cost models should have flexible functional forms. Benchmarking studies prepared using only Canadian data may nonetheless not confirm these results. In

---

[19] They may, however, involve lower OM&A expenses.

such cases, a failure to find statistical significance may be due to a sample that lacks sufficient size, variation, and data quality.

3.2 Benchmarking Customer Services

3.2.1 The Customer Care Business

The customer care unit of an LDC is responsible for revenue cycle and other customer contact responsibilities. Revenue cycle services include meter reading, billing, collection, and payment processing. Other customer contact responsibilities of distributors include the handling of calls and other contacts, arrangements to start and end services, and demand-side management.

The provision of customer care services requires capital, labour, and other operating inputs. Technological change has been rapid in the business in recent years. For example, software systems are now extensively used to manage customer information and prepare bills. With the advent of the Internet, the technology exists for customers to access account information, pay bills, and change service requests electronically. Automated meter reading makes possible more sophisticated rate structures such as hourly pricing. Because of these changes, customer care technology has become more capital intensive and software has become an important class of capital inputs. This also means that the cost of customer services is more prone than in the past to occasional “bumps” when major new automated systems are introduced.

The cost effectiveness of software is generally greater the larger is the scale of a distributor’s operations. That is because the chief cost in the use of an information system is its initial purchase and/or development. The cost incurred to serve an additional customer once a system is up and running is relatively modest. Major changes in the package of customer care services, such as those occasioned by the introduction of retail competition, can involve sizable short run cost growth due to investments in new systems.

There are many opportunities today to outsource calling centers and other customer care tasks. Customer service specialists can achieve scale economies by serving multiple utilities. Some utilities in the U.S. and Canada have outsourced the major portion of their customer service activities.

Customer Service Cost Drivers

The outputs of a customer service business can be narrowly defined as measures of its operating scale that also serve as billing determinants. One such measure is salient: the number of customers served. Our research on customer service expenses over the years has revealed some additional drivers of customer service cost. These include the following.

- The cost of local delivery services was noted above to be influenced by customer density. Customer density is likely to have an impact on the cost of customer service as well. One reason is that meter reading is a customer service. System extensiveness can once again be measured by the length of distribution lines.
- Customer service cost is quite sensitive to the scale of demand-side management activities. These activities, which can include the development of initiatives, equipment merchandizing, and extensive communications, can be quite expensive,
- Customer service cost will generally be raised by the introduction of retail competition. The experience of Ontario is illustrative in this respect. Retail competition led to more complex customer bills and more frequent rate changes. Relationships had to be established with independent power suppliers that included an extensive exchange of information. Distributors were required to have the capability to perform transactions with these suppliers electronically. The many changes in customer service responsibilities prompted larger distributors to make substantial and costly upgrades to their information systems.
- Cost is generally higher the greater is the number of languages spoken in the service territory. The service territories of several Canadian utilities have sizeable francophone populations that require bilingual services.
- Cost is generally higher in areas that involve high customer migration or turnover. An example of the former might be rapidly growing areas such as Calgary or Alberta’s tar sands region. An example of the latter might be a college town such as Guelph, Ontario.

- The quality of customer service matters to customers and some quality measures are used in service quality incentive plans. Important measures of customer service quality include billing accuracy, call response time, and the time required to resolve customer queries. The handling of sophisticated rate offerings such as real time pricing should be viewed as a premium quality service. Higher quality services are, in general, more costly.[20] Service quality expectations are generally highest in urban areas.

3.2.2 Data Problems and International Benchmarking Challenges

The data categorization problems discussed above for local power delivery apply with equal or greater force to customer services.

- Companies are inconsistent in their capitalization of OM&A expenses. A good example is the treatment of software maintenance expenses. Companies that outsource customer care tasks will report more of their IT costs as OM&A expenses.[21]
- Companies are inconsistent in their allocation of certain expenses between the customer care and A&G functions. For example, some companies assign most IT costs to A&G, whereas others allocate a sizeable share of the cost to customer care.

Missing data problems are, if anything, more severe for customer service benchmarking than for local delivery benchmarking. Data are not readily available in the public domain for important drivers of customer care cost such as service quality, language diversity, and customer turnover. Another salient problem is the poor quality of data on software costs. Data on the costs of intangible "plant" are typically not reported with the same care as data on the costs of tangible plant. In the United States, the FERC Form 1 contains no itemized data on the cost of software plant whatsoever, much less a breakdown into software used for distribution and customer service. This is also a problem in local delivery cost research, as noted above, but is more of a problem for customer services because of the greater prominence of IT in customer service costs.

---

[20] This implies that requests for better service by regulators can involve material cost increases.
[21] Outsourcing companies will, furthermore, be less able to detail customer care expenses.

International benchmarking of customer service costs involves special challenges. One formidable problem is the fact that customer services are subject to competition in a number of jurisdictions, including Britain and some Australian states. Among other problems, this limits the availability of customer service data. In addition, the usual problems are encountered in making international input price comparisons.

For all of these reasons, customer service costs have in our experience been more difficult to benchmark accurately than power delivery costs. Econometric research on customer service cost is much less advanced than in the power delivery sector. Benchmarking of detailed customer care cost items can be especially problematic due to the cost allocation inconsistencies we have discussed.

3.3 Benchmarking Service Quality

Most of our discussion in this section has pertained to the costs incurred by power distributors. Some regulators also have an interest in benchmarking service quality. Here are some comments about the challenges encountered in this benchmarking application.

1. Economic theory does not provide as much guidance on the development of service quality benchmarks as it does on the development of cost benchmarks.
2. The general approach to benchmarking that we have traced for utility costs is nonetheless applicable to service quality.
    a. The actual quality indicators of utilities depend on a wide range of external business conditions that vary between service territories. In the case of reliability, for instance, the potentially relevant conditions include customer density, the extent of system undergrounding, forestation, and the prevalence of severe weather (*e.g.* high winds and ice storms). Variations in these conditions across service territories cause the cost of achieving a given level of reliability to vary across service territories. Generally speaking, a utility is less likely to achieve a given level of quality to the extent that it is costly. The demand for quality service also varies

across service territories. For these reasons, it is natural for the level of quality to vary as well. For example, quality is often lower in rural areas than in urban ones.

b. Accurate benchmarking of service quality therefore requires controls for differences in the quality drivers facing utilities.
c. Indexing and econometric methods are both potentially useful in quality benchmarking. When indexing methods are used, accurate benchmarking requires the peer group to face quality drivers similar to those of the subject utility. When econometric benchmarking is used, accurate benchmarking requires that the model include as business condition variables the relevant quality drivers.
d. Data on service quality is at least as problematic in a benchmarking application as the data on utility cost. For example, U.S. utilities vary in the manner in which they report reliability and customer care quality metrics.
e. Methodologies for statistical benchmarking of service quality are far less advanced than methodologies for cost benchmarking. The paucity of good data is an important reason for this.

4. Role of Benchmarking in Regulation

In this section we consider some important issues that are encountered when statistical benchmarking is used in regulation. We first discuss potential uses of benchmarking in regulation. We then develop a framework for analyzing the desirability of benchmarking.

Regulators who are seriously interested in benchmarking must, of course, address a number of implementation issues as well as the broader issue of whether benchmarking is basically worthwhile. We conclude this section by discussing three important implementation issues: the choice of a benchmarking method, the choice of a performance standard, and the use of benchmarking results in ratemaking.

4.1 Potential Uses in Regulation

Benchmarking is potentially useful in utility regulation. Results of benchmarking studies can be a component of periodic rate reviews and/or an input to the design of automatic adjustment mechanisms of multi-year regulatory plans. For example, the allowed pace of rate escalation may be slower if the utility's performance is below the benchmarking standard.

There are many possible ways to link rates to performance assessments. For example, mechanistic formulas can be used. In that event, the design of the formula becomes an important policy issue. Benchmarking can, alternatively, play the more limited role of alerting regulators to the need for a particularly careful, traditional review of operating prudence.

Benchmarking studies can be used to reward utilities for superior performance as well as to penalize them for substandard performance. For example, a utility with a demonstratively superior performance can be permitted to keep a portion of the cost savings that it has achieved under a multi-year plan. If it operates under an earnings sharing mechanism, the return on equity threshold at which it is obliged to return all incremental earnings to customers may be raised.

An interesting example of the use of benchmarking to establish rewards as well as penalties comes from a recent New Zealand regulatory initiative. The Ministry of

Commerce developed a "thresholds regime" in which power distributors could avoid regulation by keeping their rate increases below a cap established by a CPI-X index. Companies were benchmarked using productivity level indexes, and grouped into three performance categories on the basis of the results. Bottom third performers saw their X factors *raised* by 1% (100 basis points). Top third performers saw their X factors *lowered* by a like amount.

4.2 International Benchmarking Precedents

Statistical benchmarking has to date played its most prominent role in regulation in Western Europe. It has been used there by regulators in several countries. Best known perhaps has been its use by Britain's Office of Gas and Electricity Markets ("OFGEM"). Most European benchmarking studies used in ratemaking are based on national rather than international data. Power distribution has been the chief focus of research.

Regulators in North America have rarely commissioned statistical benchmarking studies. However, studies are occasionally filed by utilities here to support rate filings. Indexing and econometric methods are both used.

Benchmarking studies for jurisdictional power distributors have been initiated by regulators in three Australian states. Various methods have been used. Satisfaction with the studies was decidedly mixed. None of these commissions have used statistical benchmarking in subsequent rate reviews. Australian gas and electric power distributors have voluntarily filed studies on several occasions.

In Latin America, statistical benchmarking has been used in regulation in only a few countries. Bolivia and Colombia have been leaders. Various benchmarking methods have been used. Power distribution has been the most common application.

4.3 Best Role in Regulation

4.3.1 Analytical Framework

In considering the right role for benchmarking in regulation it is useful to establish a framework for analysis. We begin by characterizing utility regulation as a process for making decisions about the allowed terms of utility service.[22] The effectiveness of this decision process is limited, like other processes in our economy, by the evolving state of technology.

In economics, technology is characterized as the amounts of outputs that can be produced by given amounts of inputs. The inputs to the technology of regulation include engineers, accountants, attorneys, and other workers, as well as the office space and equipment that are needed by regulators and stakeholders to undertake rate proceedings.[23] The output of regulatory technology is decisions on the terms of service that utilities are permitted to offer.[24] This workload can be measured. One useful measure is the number of terms of service that are approved. This will depend on the number of jurisdictional utilities and on the complexity of their rate and service offerings. It will also depend on the extent to which the terms of service change over time. For example, changes will be needed more frequently when the unit cost of utilities is rising. Conditions under which this is especially likely include rapid change in input prices, demand, and investment.

The *quality* of the decisions reached is another important technological consideration. Society wants the terms of utility services to be just and reasonable.[25] Rates, for example, should reflect the cost of service that is prudently incurred. The notion of prudence is somewhat vague in this context but boils down to a good standard of operating performance. An important dimension of the quality of regulation is thus the extent to which terms of service deviate from the just and reasonable standard. For

---

[22] We assume for simplicity here that utilities will offer the terms of service that regulators approve.

[23] The labour cost includes the opportunity cost of diverting the attention of senior utility management away from the basic business.

[24] These terms include various dimensions of service quality as well as rates.

[25] Another quality dimension worth noting is the *timeliness* of regulatory decisions.

example, rates above or below the levels that reflect operating prudence are unfair to customers and shareholders, respectively.

Let us turn, next, to the determinants of regulatory *cost*. Most fundamentally, the cost of regulation increases with the amount of work performed. Cost will be higher, for instance, the greater are the number of decisions that must be made about the terms of service. The cost of regulation is also generally higher the higher is the quality of decisions. Regulators can, for instance, deal with volatility in the price of natural gas by permitting the rapid recovery of gas purchase expenses. The costly part of regulating these purchases is, of course, the review of the reasonableness of these expenses.

A useful distinction can be drawn between the variable and fixed costs of regulation. For example, the human capital investment required to conduct competent cost of service regulation ("COSR") is substantial. If regulators lack experience with COSR but wish to use it, they must therefore pay sizeable fees to outside consultants and/or incur substantial start-up costs to develop in-house expertise. Similar cost challenges confront regulators who are competent COSR practitioners but experience a major expansion in their regulatory duties.

In fashioning the best approach to regulation, we may conclude that containment of the cost of regulating a given number of firms at a given level of quality is an important consideration. The importance of cost containment is especially critical when budgets for regulation are constrained. But society is also concerned about the quality of regulatory decisions.

Our discussion of the impact of quality on cost leads naturally to the question of what level of quality is optimal in the field of regulation. A sensible answer is that the optimal level is that which maximizes the net benefits of regulation. The net benefits are the difference between the benefits of regulation and its cost. Regulators must routinely consider whether an incremental improvement in the quality of decisions is worth the incremental cost that is required. They effectively render such judgments every time that they end their deliberations and make a decision.

The system of jurisprudence in western civilization provides an interesting perspective on the balance between the quality of judicial decisions and the cost of arriving at them. The process is costly. However, society evidently has weighed these costs against the benefits of fair enforcement of the law that our court systems provide.

Benchmarking may be viewed within this framework as an innovation in the technology for establishing terms of utility service. Its “active ingredients” include, as we have seen, cost theory, samples of utility operating data, and statistical methods. Studies can, in principle, be performed by commission or utility staff or outsourced to specialized consultants that realize scale and scope economies by serving multiple clients.

Some aspects of the cost of benchmarking merit note. As in the case of COSR, a substantial human capital investment is needed to integrate benchmarking into the regulatory process competently. If regulators lack experience with benchmarking but wish to use it, they must therefore pay sizeable fees to outside consultants or incur substantial start-up costs to develop in-house expertise. Note, secondly, that once a benchmarking study is underway for one utility the incremental cost of making efficiency appraisals for other sampled utilities is relatively low. Suppose, for example, that we are developing an index that compares the productivity of a utility to the mean productivity of a sample of twenty utilities. The incremental work required to provide analogous calculations for each additional sampled utility is well below that of performing a study for the first utility.

A third aspect of benchmarking cost that merits note is that the incremental cost of making improvements in benchmarking methods is often modest. Suppose, once again, that a benchmarking study is already underway. In that event, the incremental work that is required by a competent researcher to control for additional business conditions can be small.

4.3.2 Benchmarking Pros and Cons

Let us now use this analytical framework to discuss, in broad brush terms, some of the pros and cons of using benchmarking in regulation. We first consider the issue of regulatory cost. There follows a discussion of the impact of benchmarking on the quality of regulatory decisions.

Regulatory Cost

One potential advantage of benchmarking is the ability to economize on certain costs of traditional regulation. The cost of a testimony-quality benchmarking study can be sizable. One study may, furthermore, give rise to counterstudies and lengthy hearings.

Our analysis helps us to identify some circumstances under which benchmarking are especially likely to yield cost advantages over traditional COSR.

1. Regulators have little experience with COSR.
2. Regulators have experience with COSR but face a large expansion in their responsibilities.
3. Regulators have jurisdiction over numerous utilities.
4. Traditional prudence reviews are inherently difficult to conduct.

Better Operating Performance

A second potential advantage of benchmarking is its ability to encourage better utility operating performance. Suppose, by way of example, that a utility's revenue requirement for power procurement is based partly on its actual power procurement cost and partly on the results of a statistical benchmarking study. Specifically, the allowed cost is adjusted for 20% of the difference between actual cost and the cost benchmark. In that event, the allowed cost is a weighted average of the actual cost and the benchmark, where the benchmark is assigned a 20% weight. This would make rates considerably less sensitive to the cost of service actually achieved than straight COSR.

This externalization of ratemaking strengthens performance incentives. If a utility is operating under a five year rate plan and knows that due to benchmarking its rates at the start of the next plan may not be trued up entirely to its cost of service, it will have greater incentives to boost performance. The incentive enhancement will be especially strong in the last years of the plan period. To the extent that stronger incentives elicit better performance, the terms of service are more likely to embody the just and reasonable standard.

On the other hand, traditional prudence reviews also help to externalize the terms of service. Moreover, the evolution of regulatory technology has in recent years yielded a number of tools other than benchmarking for strengthening performance incentives.

For example, the periods between rate cases have been extended, where necessary with the help of automatic rate adjustment mechanisms. Some regulators today are experimenting with efficiency carryover mechanisms that share the benefits of performance gains between the end of one rate plan and the start of another.[26]

The externalization of rates that benchmarking facilitates can also help to reduce concerns about cross-subsidization that prompt regulators to discourage utility operating practices that are in the public interest. Suppose, for example, that an electric utility wishes to provide retail communications service using broadband over power lines technology. This can raise issues of cross-subsidization under traditional regulation. Statistical benchmarking of local delivery and customer service OM&A expenses can help to address this issue.

Quality of Decisions

The principal disadvantage of benchmarking is its potential to reduce the quality of regulatory decisions if used as a substitute for traditional prudence reviews. The quality of decisions is affected *directly* to the extent that benchmarking leads to terms of service that violate the just and reasonable standard. Quality can also be affected *indirectly* since inaccurate benchmarking can raise utility operating risk and thereby raise the cost of obtaining funds in capital markets. The resultant increase in the cost of service can materially reduce the net benefits from benchmarking.

The extent of the quality problem depends on the imprecision of the benchmarking work. The analysis we have presented in this paper suggests that the extent of imprecision will be greater to the extent that:

- The sample of good quality, standardized data available for research is not large and varied.
- Numerous business conditions have a material impact on the performance variable.
- The subject utility is dissimilar from those in the benchmarking sample.

---

[26] PEG has done extensive research to quantify the incentive impacts of alternative regulatory systems. Efficiency carryover mechanisms have been an important focus of this work.

- There is less accumulated benchmarking experience with the chosen performance variables.

Striking a Balance

It is evident from our analysis that the balance between the cost and quality of regulatory decisions is a key consideration in deciding whether to use benchmarking in regulation. Responsible regulators are more likely to use benchmarking to the extent that they can lower the cost of regulation without unduly reducing its quality. The disadvantages of benchmarking can in principle outweigh the advantages so that it is optimal not to undertake benchmarking.

This analysis goes a long ways towards explaining where statistical benchmarking is used today in regulation. For example, we have noted above that benchmarking is most commonly used in the regulation of power distribution. This reflects in part the fact that power distributors are much more numerous in many jurisdictions than power transmission providers. Ontario, for example, has more than 100 distributors but only one sizable transmission provider. Western European regulators that use benchmarking typically regulate a dozen or more power distributors but just a handful of transmission providers. With numerous distributors to regulate even conscientious COSR is especially costly. Moreover, there is a larger set of standardized data available for benchmarking. It is also noteworthy that benchmarking is more prevalent in jurisdictions that lack a long history with COSR.

In the United States and many Canadian provinces, on the other hand, regulators have considerable experience with COSR. Furthermore, most regulatory agencies have jurisdiction over only a handful of energy utilities that are in the same business.[27] The data available for benchmarking are voluminous in the States but not in Canada.

Our analysis also helps to explain the common focus of regulators on the benchmarking of opex. As we explain in Section 2, accurate capital cost benchmarking requires complicated capital cost measurement methods that require years of detailed capital cost data. Such data are not available in most countries. Note also that regulators

---

[27] State regulators in Australian are in an intermediate situation. Their responsibilities expanded in the 1990s to include the regulation of power distributors. However, most states have jurisdiction over only 3-6 distributors.

are more likely to focus on total opex than on more micro opex subcategories such as distribution labour expenses where the issue of other inputs used complicates analysis.

A recent proceeding in Ontario is interesting for having expressly considered the use of benchmarking in regulation. Power distributors there have incurred special costs in the transition to retail competition. Recovery of these costs was considered by the OEB in a recent proceeding. Statistical benchmarking was advocated by an intervener group. The prospects for accurate benchmarking of these costs were limited for several reasons. Transition costs are, for example, only a small subset of the total cost of providing customer services. Methods for accurate benchmarking of highly partial cost categories are noted in Section 2 to be in their infancy. There is little or no experience around the world with the benchmarking of these costs. The Board elected not to benchmark this cost category.

More light may be shed on the decision problem facing regulators by considering an analogous decision from the electric utility industry. Suppose that the following conditions hold for a utility with power transmission operations:

- The company is weighing the possible use of a new transmission line maintenance procedure.
- Its transmission lines are essential for the fulfillment of several important bulk power sales contracts.
- The contracts call for the company to pay its customers penalties in the event of a failure to deliver specified amounts of power at specified levels of reliability and quality.

A key consideration in deciding whether to adopt the new technology is the reduction in line maintenance expenses that will ensue. This must, however, be weighed against the danger that the new techniques will trigger reductions in the reliability and quality of power deliveries. Such problems would reduce net benefits directly, through the penalty payments. They would also reduce the net benefits *indirectly* to the extent that the resultant increase in operating risk raises the cost of securing funds in capital markets.

These considerations suggest that the utility should proceed cautiously with deployment of the new technology. One cautious approach might be to try it first on one line and to carefully appraise the resultant net benefits before applying it systemwide. Another would be to delay implementation of the new technology until more evidence accumulates around the world regarding its successful use by other companies.

Regulators can also adjust for any risk/return imbalance created by the commencement of benchmarking by raising expected return. This can be achieved most directly by allowing rewards for superior performance as well as penalties for inferior returns. Regulators should also consider placing more emphasis on carrots than on sticks. Suppose, by way of example, that a regulator has benchmarked one or more utilities during the development of a multiyear (e.g. four year) rate plan. The benchmarking results can be used to identify achievable cost targets for the next rate case. If a utility's test year unit cost in the next rate case is at or below the target in real terms, it can obtain the associated initial rates for the next plan even if its actual cost is lower.[28]

4.4 Choosing the Right Benchmarking Method

Section 2 of our report examines several benchmarking methods that are available for use in regulation. Our analysis there suggests that the menu of methodological options is extensive. Here is a partial listing of some of the methodological choices encountered.

- Benchmarking focus
  - Total cost, opex, or other?
  - Reliability or other quality indicators?
- Basic method: econometric modeling, indexing, or other?
- If indexing,
  - Unit cost or productivity?
  - Single or multiple-category output measure?
- If econometric modeling,
  - What estimation procedure?
  - What functional form?

---

[28] The discrepancy between actual and target cost can, alternatively, be limited to say, 100 of 200 basis points of ROE.

- Choice of sample
  - Number of sample companies
  - Cross sectional data (one observation per company) or panel data (multiple observations per company)?
  - Provincial, national, or international data?
- Statistical tests of efficiency hypotheses?
- What performance standard?

In this section we provide some general criteria for selecting a method. Concepts from statistics are used to aid reasoning. Specifically, we treat alternative benchmarking methods as means to predict the true extent of inefficiency using a certain performance standard. We might, for example consider the ability of a method to generate the extent of a company's deviation from the sample average level of efficiency. The appraisal produced by each method is treated as a random variable that is drawn from a probability distribution.

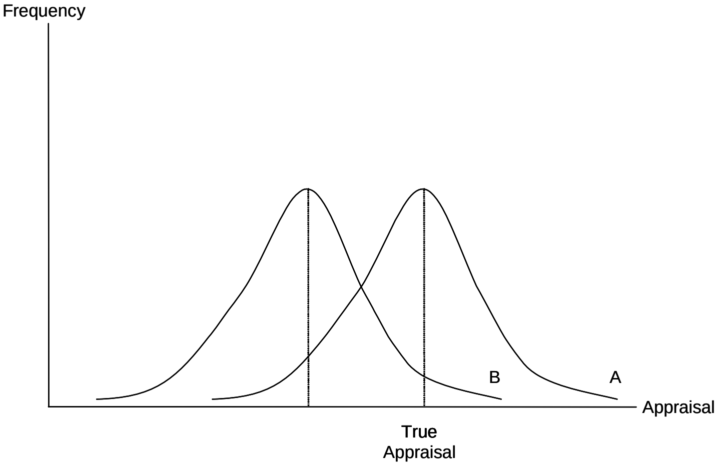
4.4.1 Accuracy

One criterion that is obviously important in choosing a benchmarking method is accuracy. A method is accurate to the extent that it generates an accurate performance appraisal. For example, an accurate benchmarking study that uses an industry average performance standard would correctly measure the deviation of efficiency from that standard.

The accuracy criterion can be usefully decomposed into two subcriteria: bias and variance. A method is unbiased to the extent that it is expected to generate the correct appraisal in repeated applications even if it does not do so in every instance. One example of a biased method is one that tends in repeated applications to make appraisals that are excessively harsh. Another is a method that tends to generate excessively harsh appraisals for companies facing certain business conditions, such as pronounced urbanisation. Amongst unbiased methods, we desire a method that generates appraisals that vary as little as possible around the correct appraisal.

The concepts of bias and variance are illustrated in Figures 1 and 2 below. In both figures, we treat the appraisals generated by alternative methods as random variables with values drawn from probability distributions that can be represented graphically as familiar “bell curves.” The high part of each curve indicates the range of appraisals that are likely to occur most frequently. The curves reach their peak at the appraisal that is expected in repeated applications. This is not necessarily the same as the correct appraisal.

Figure 1



In Figure 2 the curves for two benchmarking methods, A and B, are presented. The curves for each method center upon the expected appraisal for the method. The curves differ only in the expected appraisal value. It can be seen that in the case of method A the expected appraisal is the true appraisal. Thus, method A is unbiased. Method B, however, is biased since its expected appraisal value is lower (and less favourable) for the company than the true value.

Our discussion of benchmarking methods in Section 2 suggests several ways to reduce the bias of benchmarking studies. In econometric modeling, for example, researchers should at a minimum try to include all relevant business conditions in the model for which data are readily available. When doing index-based benchmarking, productivity indexes are generally preferable to unit cost indexes and peer groups must be carefully selected.

Figure 2

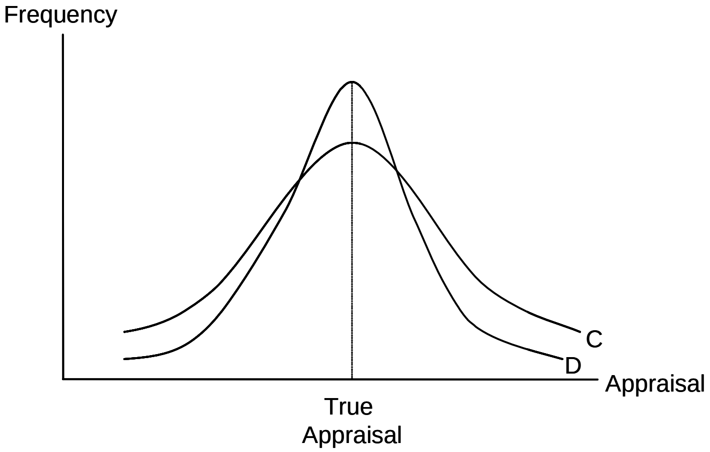


Figure 3 depicts the probability distributions for benchmarking methods C and D. It can be seen that both methods are unbiased. However, method D has a lower variance than method C and is therefore more accurate.

Our methodological discussions in Section 2 suggest ways to reduce the variance in the accuracy of benchmarking appraisals. One sensible step is to use statistical tests of efficiency hypotheses. With this approach, regulators will be less likely to conclude that a utility is an outlier when they have less confidence in the benchmarking study. Another sensible step is to use all readily available data of good quality. It does not make sense, for instance, to base benchmarking research on only one year of quality data when two are readily available. Rich econometric models and the use of productivity indexes should also reduce variance along with bias.

Whether or not confidence intervals are used mechanistically to draw benchmarking conclusions, it makes sense to use benchmarking practices that cause them to narrow. When undertaking econometric benchmarking, for instance, it is desirable to estimate model parameters using a large and varied sample of data.

4.4.2 Cost of Benchmarking

The cost of benchmarking is another relevant consideration in method selection. Methods are generally more costly to the extent that they are more accurate. Simply put, accurate benchmarking of complex phenomena requires complex methods. Complex methods are more costly to develop and implement than simple methods and are also more costly for stakeholders to review. The cost of achieving a given level of accuracy will be greater to the extent that readily available data are deficient in quantity or quality. It will also depend on the current state of the benchmarking art. That is to say that over time the cost of achieving a given level of accuracy should improve.

4.4.3 Striking a Balance

In choosing a benchmarking method a reasonable balance must be struck between accuracy and cost. Since traditional regulation is quite costly, even complex benchmarking methods have the potential to reduce net costs. Furthermore, the amounts of money at stake in the regulatory arena are generally far in excess of the cost of regulation. A bad benchmarking method that generates bad appraisals can violate the just and reasonable standard and, by raising operating risk, can materially offset any net cost savings in the regulatory process. We have also shown that the incremental cost of making improvements in benchmarking methods is often quite reasonable.

One sensible conclusion that may be drawn from this discussion is that it will generally make sense to use relatively sophisticated benchmarking methods in the regulatory arena. The cost of a benchmarking study that can attain such accuracy can be considerable. This conclusion should not be misinterpreted to suggest that there always exists some benchmarking method that is optimal. No benchmarking may be the desirable strategy when a reasonable degree of accuracy cannot be attained.

Despite the reasonableness of these conclusions, the benchmarking methods that are used by regulators are often quite far from the state of the art. The benchmarking studies of power distribution opex performed by OFGEM are a case in point. In OFGEM's first benchmarking study it used an econometric cost model that was based on only one year of data. Since fewer than twenty observations were available, the potential sophistication of the model was severely limited. The model ultimately chosen in fact had only one explanatory variable and an inflexible functional form. Results from such a study had little chance of being accurate. Yet, no statistical tests of efficiency hypotheses were undertaken.

Five years later OFGEM performed a follow-up study. Although it had the opportunity to gather a panel data set over the intervening years and to upgrade its benchmarking methods it did not to do so. The model ultimately selected was once again based on less than twenty observations and featured only one business condition variable and a restrictive functional form. No statistical tests of efficiency hypotheses were undertaken.

4.5 Benchmarking Standards

The choice of a performance standard is noted above to be important in the development of a benchmarking strategy. We consider here three alternative standards: the sample average, competitive market, and frontier standards.

4.5.1 Average

Under a sample average performance standard, benchmarking focuses on how a company's performance compares to the average amongst sampled utilities. Many of the benchmarking methods discussed in Section 2 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 be used to reward good cost performance as well as to call substandard performance into question. In a multiyear rate plan, for example, a company might be eligible for a zero stretch factor if benchmarking suggests a performance markedly superior to the norm. Incentive power research reveals that rewards like this can materially improve performance.

4.5.2 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 at least the desirable aspects of competition.[29]

In a competitive market, prices reflect the interaction of supply and demand conditions at the industry level and the industry as a whole earns a competitive rate of return. Since individual firms cannot 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 performances leave the industry.

These attributes of competitive industries encourage the view that *all* firms in such industries are efficient. The reality, however, is that it is the *industry as a whole* that earns a competitive rate of return and not each of the individual companies in the industry. At any point in time the efficiency of firms can vary considerably. More efficient firms achieve superior rates of return. 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 are especially relevant. Recall, however, that the current state of the benchmarking art only permits comparisons to the short-run unsustainable efficiency frontier. Furthermore, the distance from even this frontier is difficult to estimate accurately.

On behalf of two British power distributors, PEG has conducted surveys of frontier benchmarking studies in two competitive sectors: banking and farming. Results are reported in Tables 2 and 3. 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%. Note that the efficiency studies in the farming sector that we examine consider only technical efficiency and therefore do not consider all possible sources of efficiency.

---

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

Table 2. 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 Bank 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 3. Survey of Efficiency Studies of Farming Firms

Study	Data Coverage	Method	Result
Brummer, Glauben and Thijssen (2002)	German, Dutch and Polish Diary Farms 1991-1994		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. Diary 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 short-run non-sustainable frontier in either of these two competitive industries. Taking these results as representative of competitive industries as a whole, we may conclude that to simulate competitive markets the relevant cost performance standard is one at some substantial distance from the short-run unsustainable 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 non-sustainable frontier. Another possible approach is to determine that any company with a performance that is significantly superior to the sample average standard has equalled or exceeded the competitive market standard. A sensible approximation of the competitive market standard might well be a top quartile standard.

4.5.3 Frontier Standard

Some regulators have used a frontier standard for benchmarking. Under this standard, companies are judged not by their position relative to the sample norm but, rather, by their distance from the efficiency frontier. This paradigm can be criticized on several grounds. One is that the frontier that can be measured is the short-run unsustainable frontier and not the long run sustainable frontier. As we have seen, it is unreasonable to expect utilities to operate for sustained periods at this frontier. If anything, companies found to lie on this frontier are likely to need additional revenues to ensure the sustainability of quality service.

The difficulties in measuring distance from the short-run unsustainable frontier have also been noted. 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 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 labour unions to offset the potential monopsony power of employers is one of the major arguments ventured for legalizing their activities.

The frontier standard has been used in several European countries. OFGEM is perhaps the best known practitioner. However, familiarity with the disadvantages of a frontier standard recently prompted OFGEM to abandon it. In its latest power distribution price control review, OFGEM has opted instead for a *top quartile* standard. Companies with a measured performance exceeding this standard have a chance to earn superior returns.

4.6 Concluding Comments

The use of benchmarking results in regulation matters as much as the choice of a performance standard or a basic benchmarking method. To the extent that there are concerns about the accuracy of benchmarking, our analysis suggests that its use in ratemaking should be more limited. For example, a first generation benchmarking study could be entirely experimental and have no role in the ratemaking process. Some regulators may wish to use benchmarking only to help direct the discussions and discovery that are a normal feature of the rate case process. Others may wish to focus more on the trends in the cost and quality of utilities rather than on inter-utility cost comparisons.

This analysis helps to explain the cautious approach to benchmarking that the OEB has taken. Board staff recently completed their first generation benchmarking study. They used the results of the study only to identify companies that should be the focus of in-depth traditional reviews.

We conclude, finally, that the many challenges encountered in doing benchmarking accurately argue in favour of a collaborative approach to the development of regulatory benchmarking. Utilities should have the opportunity to share their expertise on the complex set of business conditions that affect their cost and service quality. They should also be encouraged to comment on the appropriate methods and the best role for benchmarking in the regulatory process.

5. Implications for Alberta Regulation

Statistical benchmarking in Alberta regulation has to date been limited. Performance indicators have been simple and statistical hypothesis tests have not been used. The Board is currently considering the implementation of USAs and MFRs for power transmission and distribution. The implementation cost of these measures is expected to be substantial.

Our analysis makes it possible to make some constructive comments concerning the use of benchmarking in Alberta regulation. Performance incentives may be strengthened through benchmarking, and unusually bad or good performances may be discovered that make it possible to make rates more just and reasonable. However, benchmarking has a limited ability to improve the efficiency of Alberta regulation since the Board does not, like its counterpart in Ontario, have jurisdiction over dozens of utilities.

Methods will have to be improved considerably before benchmarking can play an important role in rate-setting. For example, it is difficult to draw statistically significant conclusions concerning operating performance from simple unit cost comparisons that are based on the small amount of data that can be gathered from Alberta utilities. Better benchmarking will require better statistical methods and data drawn from other Canadian provinces or the United States. Good benchmarking studies are costly, as are the revisions in Alberta data reporting requirements that are needed to achieve the requisite standardisation for ex-provincial comparisons. It will be difficult to benchmark capital cost accurately without years of historical plant addition data.

The development of a Uniform System of Accounts would be a constructive step towards data standardisation. However, it should be recognized that the some of the micro level data typically gathered in such accounts are difficult to use in regulatory benchmarking. This finding may point the way towards a streamlining of reporting requirements.

It is also important to recognize that standardized, quality data on the operations of utilities in other Canadian provinces are not yet available in the public domain. While more abundant data of good quality are available in the U.S., these data have noteworthy limitations in international benchmarking work. The limitations include inconsistencies in the classification of transmission and distribution, the difficulty of making accurate international price comparisons, and the unavailability of important performance indicators and business condition variables.

If the Board intends to use benchmarking in rate-setting it should do so in a way that preserves the balance of operating risk and return. Balance is encouraged by good scientific methods that include statistical tests of efficiency hypotheses. Benchmarking results should be interpreted cautiously. Rate adjustments may not be desirable in early benchmarking efforts. If and when benchmarking is used in rate setting, utilities should have the opportunity to earn a superior return for superior performance.