



National Regulatory
Research Institute

**Utility Performance: How Can State
Commissions Evaluate It Using Indexing,
Econometrics, and Data Envelopment
Analysis?**

Evgenia Shumilkina

NRRI

**March 2010
(Revised April 26, 2010)**

10-05

Acknowledgements

The author would like to thank Scott Hempling and Ken Costello for their contributions to this report.

Online Access

This report can be found online at
www.nrri.org/pubs/multiutility/NRRI_utility_performance_mar10-05.pdf.

Executive Summary

All regulators want to see utilities under their jurisdiction delivering high levels of performance. But how can a regulator tell if a utility is a top performer? It seems easy to determine an Olympic winner: In most sports, an athlete's performance can be measured by a fraction of a second or a millimeter. Even in sports like figure skating, where performance is a combination of different subjective qualities, judges manage to rank athletes in a convenient fashion. It seems relatively straightforward, thanks to the set of well-defined evaluation procedures developed by the Olympic Committee.

This paper aims to inform regulators about statistical methods they can use to evaluate utility performance—to determine which utilities are top performers and which ones require additional training.

The paper covers the three statistical methods most commonly used in the evaluation of utility performance: indexing, econometrics, and data envelopment analysis. **Indexing** is a relatively simple method that allows the regulator to compare the performance of a targeted utility against the performance of its peers; the choice of a peer group is the main challenge in applying this method. **Econometric methods** build and estimate an econometric model—an equation—that describes a relationship between a specific aspect of utility performance and factors that affect it. The third method—**data envelopment analysis**—determines a set of most efficient solutions, then measures the performance of the utilities in the sample against them. For each of the three methods, the paper provides the basics of the technique, implementation steps, data requirements, advantages and limitations, and examples of application. The paper's goal is to remove the lay reader's technical fears so that regulators can use these techniques more frequently.

The methods vary in their **data requirements**, **ease of application**, and **reliability of results**. The paper compares the three methods across these three dimensions to help policymakers choose an optimal method or methods for measuring performance, based on their resources and goals.

After reading this paper regulators will be able to make an informed choice from among the techniques for evaluating utility performance and will be able to examine the evaluation studies submitted by utilities.

Table of Contents

Introduction	1
I. The Indexing Method	3
A. Basics of the indexing method	3
B. Types of indices	4
C. Implementation steps	6
D. Data requirements	7
E. Advantages and limitations	7
F. Application examples	7
II. The Econometric Method	9
A. Basic econometric concepts	9
B. The use of econometrics in the measurement of utility performance	14
1. Implementation steps	15
2. Implementation example	15
C. Advantages and limitations	20
D. Application examples	21
III. Frontier Methods: Data Envelopment Analysis (DEA)	22
A. Basics of the DEA method	22
B. Implementation steps	24
C. Extensions of the basic DEA model	25
D. Advantages and limitations	25
E. Application examples	27
IV. A Comparison of the Methods	28
Appendix I: Calculation of a Multifactor Cost Index	29
Appendix II: Recommended Texts for Further Reading	30

Utility Performance: How Can State Commissions Evaluate It Using Indexing, Econometrics, and Data Envelopment Analysis?

Introduction

The central goal of regulation is to induce high-quality performance in regulated utilities. Despite the centrality of this goal, there is no common practice by which commissions define and measure performance. Without definitions and measurements, it is difficult to create effective frameworks to induce improved performance. Regulators would benefit from a greater awareness of the existing methods for measuring utility performance and how they can integrate the results of these methods into their regulatory practice.

This paper provides regulators with a practical guide to the available methods of measuring utility performance. Our hope is that this guide will encourage greater use of these methods, leading to a base of data that will assist all states in making performance comparisons.

The paper covers the three benchmarking methods most commonly employed throughout the world in measuring utility performance: *econometrics*, *indexing*, and *data envelopment analysis*. Benchmarking—the theme common to all three methods—is a process of measuring a utility’s performance against a reference performance. For each of the three methods, the paper provides the basics of the technique, implementation steps, data requirements, advantages and limitations, and application examples.

Utility performance is a multidimensional issue. Utility performance dimensions include, but are not limited to, cost efficiency, reliability, customer satisfaction, energy efficiency progress, capacity utilization, power plant performance, labor productivity, total factor productivity, utility anticipation of future events, reaction to unexpected events, use of renewable resources, and innovation. The paper primarily discusses measuring operations and maintenance (O&M) costs that are often subject to benchmarking analysis, though the application of statistical performance metrics is not limited to O&M costs. One can measure other aspects of utility performance as long as quantifiable data is available. The analyst can adjust each method to serve the task at hand.

The methods vary in their *data requirements*, *ease of application*, and *reliability of results*. The paper compares the three methods across these three dimensions in order to help policymakers choose an optimal method or methods for measuring performance, based on their resources and goals.

The results of the statistical methods can provide objective, credible evidence about utility performance. Commissions can use this information in a variety of ways, both within and outside the context of rate cases.

In *rate cases*, a commission can use the methods to evaluate the reasonableness of the “cost-of-service” component of test year filings. Based on the information gleaned from statistical methods, a commission can assess the appropriateness of expense projections, which

may be high or low relative to peer groups. Commissions can also reduce the rate of return if the utility's performance is below industry standards. The data is useful whether a state is basing rates on embedded costs or on a performance-based ratemaking strategy (for example, where the commission sets benchmarks for a particular level of reliability and then designs corresponding incentive mechanisms to achieve or surpass those benchmarks).

We can find the application of statistical methods in the testimonies of a number of utilities. Most of the testimonies use these methods to demonstrate a particular utility's performance compared to other utilities in the industry. This paper will help regulators understand the methods employed by utilities in their studies, as well as the credibility of these studies.

Outside of rate cases, a commission can use the information to inform the public about utility performance, use rulemakings to induce improved performance, assess a utility's continued right to hold the monopoly franchise to serve, conduct investigations into accidents (for example, large outages), and compare a utility's expenditures and performance to its peers.

Understanding this paper does not require the reader to possess a statistical background. The engaged reader will, on completion, understand the basics of the measurement techniques and know how to assess the results of performance measurements.

I. The Indexing Method

Indexing is the most basic benchmarking method. It compares a utility's performance indicators against those of a **peer group**—a group of companies that operate under similar business conditions. The peer group's performance serves as the benchmark. We discuss how to construct a peer group, as well as other basics of the indexing method, below. The section on the basics is followed by discussions of the types of indexes, advantages and limitations of the method, data requirements, and application examples.

A. Basics of the indexing method

Utilities differ in their performance due to (a) operating efficiency, which is subject to the utility's discretion, and (b) business conditions that are outside of utility's control, at least in the short run. Benchmarking methods focus on **operating efficiency**—reflecting how well a utility manages the assets it already has.

Benchmarking methods, including the indexing method, are not well suited to examine the effects of long-term **strategic decisions** (regarding resource mix, operating scale, and percentage of underground lines, among others) on a utility's performance; these strategic decisions establish conditions that themselves are outside of the utility's control. This is one of the limitations of benchmarking. For example: A utility might choose to build a coal plant. Once it does, coal market prices are outside the utility's control. It would be difficult to assess the wisdom of the utility's decision to build a coal plant via benchmarking.

To compare utilities' performances we need to distinguish between two factors: operating efficiency and outside business environment. Indexing does this by comparing a utility's performance indicators with those of a peer group. A peer group allows this separation (between factors within and outside a utility's control) because we compare utilities that belong to the same cohort in terms of business conditions. The analyst compares the subject utility's performance against the average performance of peers, but one can also compare the utility against the group's best practices.

Comparing a utility's performance with the performance of its peers does not tell us about performance in an absolute sense, i.e., whether that performance is good or bad. It simply makes a comparison. If the entire peer group is underperforming relative to what is feasible or prudent, the only conclusion we can draw is that a particular utility is better or worse than a set of its peers. This approach does not tell us what constitutes "good" performance, or the performance level of which utilities are capable.

The accuracy¹ of the index analysis depends to a large extent on the choice of a peer group. When a peer group includes utilities operating under different business conditions, indexing results are misleading. If, for example, the analyst compares the cost efficiency of a small utility with the performance of a peer group consisting of large utilities, she might

¹ By "accuracy" in this context we mean how well the results reflect the differences in the level of operating efficiency across utilities.

mistakenly conclude that the small utility performs less well; error would flow from her failure to take into account economies of scale.² In addition to operating scale, the analyst can select peer utilities on the basis of geographical region, percentage of underground lines,³ lack or presence of generation assets, input prices, service offered to rural or urban areas, and other factors.

Two techniques that can assist in selecting an appropriate peer group are **econometric analysis** and **cluster analysis**. Econometric analysis (discussed in Section II) can help determine the most significant factors affecting utilities' performance. As an example, say the analyst determined, based on her industry expertise or according to the results of econometric analysis, that the most important factors determining the level of utilities' costs are geographical region, operating scale, and percentage of undergrounding lines. She should then select peer utilities that operate in the same region and have a similar percentage of underground lines and a similar volume of operations. Thus, econometric analysis plays two roles in this paper: first, as a way, within the context of indexing, to help refine the selection of a peer group and construct multifactor indices (as discussed below); additionally, as an independent (as well as indexing and data envelopment analysis) method of benchmarking.

Cluster analysis can help identify cohorts of utilities with similar business conditions. Cluster analysis is a technique used in various disciplines to assign a set of observations (utilities, in our case) to subsets (clusters), such that observations in the same cluster are similar in some characteristics (in our example, percent of underground miles and operating scale), but have different characteristics between clusters.⁴

It is advisable to include as many utilities that meet the criteria for a peer group as possible. The smaller the group, the greater the risk that its members will be unrepresentative of the industry practices, making the benchmarking results unreliable. But it is not always possible to find an appropriate peer group—meaning a group that (a) has the necessary homogeneity (similarity) in important characteristics and (b) is large enough to make comparisons useful. In this case, regulators should consider other benchmarking methods.

B. Types of indices

A **unit cost index** is the ratio of a company's unit cost to a peer group's unit cost. Unit cost is calculated as a ratio of a measure of costs (e.g., total costs, O&M costs, labor costs, or any

² "Economies of scale" are the increase in efficiency of production as the amount of output increases. We expect larger utilities to have lower costs due to the presence of scale economies.

³ Utilities with a larger share of underground lines typically have lower O&M costs.

⁴ Cluster analysis requires a statistical package such as STATA, SAS, or SPSS. For a discussion of the mathematical basis for clustering analysis, see Brian S. Everitt et al., *Cluster Analysis* (2001); Hartigan, J., *Clustering Algorithms (Probability & Mathematical Statistics)* (1975); and Charles Romesburg, *Cluster Analysis for Researchers* (2004).

other category of costs) to some measure of output (total number of customers served, retail delivery volume, miles of lines, total service area). The following unit cost index is based on the ratio of O&M costs to the number of customers.

$$\text{UnitCostIndex} = \frac{O \& M \text{Costs}_{\text{UtilityX}} / \text{NumberofCustomers}_{\text{UtilityX}}}{O \& M \text{Costs}_{\text{PeerGroupAverage}} / \text{NumberofCustomers}_{\text{PeerGroupAverage}}}$$

Suppose the ratio of O&M costs to the number of customers for Utility X is 1.5 and the same ratio for the peer group is 1.7. The index is then 1.5/1.7=0.88, which indicates that Utility X's costs are below those of the peer group.⁵

The narrower the category of costs, the more carefully the analyst must interpret results. If we look at the category of labor costs we might find that a company is very cost-efficient compared to its peers. The low level of labor costs, however, might be a result of the utility's outsourcing activity, where outsourced costs exceed internal labor costs. Indices based on a wider category of costs (for instance, total costs) take into account the variation in companies' reliance on outsourcing.

Costs tell us nothing about quality of service. Lower costs can be achieved at the expense of quality. Thus, it is advisable to compare utilities with similar quality levels. This is not always feasible, given the fact that data on utilities' quality of service is not always available.

Another popular type of indices are **performance indices**. These indices have a measure of output in the numerator and a measure of input in the denominator. The following performance index employs the *number of kWh delivered* as an output and the *number of employees* as an input.

$$\text{PerformanceIndex} = \frac{KWhdelivered_{\text{UtilityX}} / \text{NumberofEmployees}_{\text{UtilityX}}}{KWhdelivered_{\text{PeerGroupAverage}} / \text{NumberofEmployees}_{\text{PeerGroupAverage}}}$$

The weakness of a single indicator: Indices in the examples above are based on the **ratios of single measures**—the particular type of cost to a particular output in the case of a unit cost index; specific output to specific input in the case of a performance index. These indices are the easiest to calculate and are used most often in regulatory practice. A utility's performance, however, can hardly be captured by a single indicator. As an example, let us say that a utility has an index of 0.88 (better than average) for O&M costs per customer, 1.4 (worse than average) for O&M costs per kWh delivered, and 1.01 (about average) for O&M costs per mile of lines. Each

⁵ An index value of above 1 would indicate that a utility has problems with cost efficiency relative to its peers. The index value of 1 equals the average performance.

of these indices reflects the utility's performance only in part. It is difficult to make a general conclusion about this utility's O&M cost performance based on three single-measure indices.

To resolve this problem, instead of calculating a number of different statistics we can construct a **multifactor index** that takes into account different measures of output. Different output measures play different roles in cost formation. Econometric modeling (discussed in Section II below) helps to examine these roles. Say the econometric analysis indicates that O&M costs increase by 3 cents⁶ with the addition of a new customer, by 2 cents with an extra kWh of delivery, and by 1.6 cents with an extra mile of lines. The multifactor index should reflect the different effects of output factors on costs by assigning weights to each output factor. In our example, the analyst can assign the following weights to the output factors: number of customers = 46%, retail deliveries = 30%, total miles of lines = 24%⁷.

The multifactor unit cost index is then obtained by dividing O&M costs by the sum of weighted output measures. The technical calculation of the multifactor index is provided in Appendix I.

In addition to unit cost indices and performance indices, we can construct various other indices that reflect other aspects of the utility's performance—for example, indices that reflect rate level, reliability of the utility's services, and use of renewable sources—as long as the performance information is quantifiable.

C. Implementation steps

To apply the indexing approach the analyst should follow several steps:

Step 1. ***What do we want to measure?*** - Choose variables of interest. For instance, O&M costs as a measure of costs, and a total number of customers as a measure of output.

Step 2. ***Who is in the peer group?*** - Select a peer group. For example, small utilities that operate in the Northwest and serve urban areas.

Step 3. ***Construct the index.*** - The construction of multifactor indices requires two **additional steps**: (a) assign weights; and (b) construct the index using the assigned weights.

⁶ The values are used for illustration purposes only.

⁷ These weights represent shares of each output in the sum of the effects of all three outputs on costs. The sum of the effects of outputs on costs in this example is $3+2+1.6=6.6$. The share of the number of customers in this sum is $(3/6.6)*100=46\%$, the share of kWh delivered is $(2/6.6)*100=30\%$, and the share of total miles of lines is $(1.6/6.6)*100=24\%$.

D. Data requirements

Since all benchmarking methods are data-driven, it is important to list data requirements for each method. For U.S. regulators, one of the attractive features of the benchmarking method is the existence of extensive publicly available data. Different categories of cost data are available from FERC Form 1. Output data (number of customers served, retail delivery volume, total miles of lines) are available from FERC Form 1 and EIA-861.

E. Advantages and limitations

Indexing is the method most often used by utilities and their regulators to evaluate utility performance. The popularity of this method can be explained by the following advantages: (a) it is simple, (b) it is easy to carry out—staff members do not need to possess extensive statistical training, (c) it requires less data than the econometric and data envelopment analysis methods, and (d) data is publicly available. The simplicity of the technique comes at a cost, however. The accuracy of the results is very sensitive to the choice of a peer group. The method does not allow accounting for numerous differences in utilities' circumstances. Thus, the method in most cases does not permit good separation of the differences in performance due to outside factors from differences due to utility efficiency (the latter being what we are trying to capture with this technique).

F. Application examples

Below are a couple of examples of the use of the indexing method in regulatory practice.

The Mississippi Public Service Commission adjusts Mississippi Power Company's return on investment according to the company's operating performance. The Commission examines the following three performance indicators: (1) comparison of the utility's retail price to its peer's prices, (2) measurement of customer satisfaction, and (3) customer service reliability.⁸ The last two factors are not based on comparisons to the peer group and thus are not relevant to our discussion. The first indicator serves as an example of the indexing method application. The Commission calculates this indicator as a ratio of the company's average retail price per kWh to the weighted average price of the utilities in the Southeastern Electric Exchange (peer group).

The indexing approach has a history of application in the proceedings of the Ontario Energy Board. For example, Hydro One Networks used the indexing approach in its 2006 testimony on power transmission costs. Enbridge Gas Distribution employed the indexing and econometric methods in its 2004 and 2005 testimonies on O&M expenses.

The OM&A costs-per-customer statistic informs the staff of the Ontario Energy Board in various analyses. For example, the level of OM&A (operation and management, and administrative) costs per customer informs the Board regarding the reasonableness of the OM&A

⁸ For more details see Mississippi Power Performance Evaluation Plan Rate Schedule "PEP-4A", January 9, 2009. Docket No. 2006-UN-0511. We express no opinion on the appropriateness of these three factors.

costs submitted for approval in a utility's cost-of-service application. This index is one of the pieces of evidence the Board can take into consideration.

The Ontario Energy Board uses statistical methods to determine the X-factor in the context of the 3rd Generation Incentive Regulation Plan. X-factors for individual electric distribution utilities consist of an industry productivity trend (productivity factor) and a stretch factor. The productivity factor is set based on total factor productivity, which is calculated with the help of the indexing approach. Two benchmarking methods (the indexing method and the econometric method) are used in assigning stretch factors.⁹

⁹ *Report to the Board on the 3rd Generation Incentive Regulation for Ontario's Electricity Distributors*, Ontario Energy Board, July 14, 2008.

II. The Econometric Method

This portion of the paper illustrates how regulators can use econometrics to measure utility performance. We begin with a brief overview of econometrics. As discussed in more detail below, econometrics is a statistical tool used by economists to quantify economic relationships. An examination of how to apply econometrics to utility financial and operating data to evaluate utility performance follows the discussion of the basic concepts.

We can find the application of the econometric method in a number of rate cases. Econometric results are typically used to compare a performance of a particular utility to performance of other utilities in the industry. Examples include the following: On behalf of The Illinois Citizens Utility Board and The People of the State of Illinois, Power System Engineering evaluated the electric cost performance of the Ameren Illinois Utilities and found that the utility was an inferior O&M cost performer¹⁰; Oklahoma Gas & Electric conducted a benchmarking study to show its above-average O&M cost performance in 2009¹¹; Ameren UE used econometrics to provide evidence of the superior cost performance of its Missouri operations in a 2003 testimony. Econometrics played a supporting role in San Diego Gas & Electric's study of total factor productivity (TFP), which was based on the indexing method¹².

A. Basic econometric concepts

We start with a definition. **Econometrics** is a combination of economic theory and statistics that helps to analyze and test economic relationships. If we are trying to understand what contributes to the variation in wages within a particular segment of the work force, we might examine the relationship between wages per employee and years of education. We begin with the **hypothesis** (i.e., the tentative statement that we can test by further investigation) that wages are higher at higher levels of education. Thus, we expect the relationship between the two variables to be positive. We can test the relationship with the help of an **econometric model**, which represents the relationship between economic variables as an equation. The following is a generic form of an econometric model.

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni} + u$$

Subscript i stands for an individual observation (an employee, in the case of a wage model).

¹⁰ *Statistical Evaluation of the O&M Cost Performance of the Ameren Illinois Utilities*, Power System Engineering, Inc., September 28, 2009.

¹¹ *O&M Cost Performance of Oklahoma Gas & Electric*, Pacific Economics Group, LLC, February 2009.

¹² *TFP Research for San Diego Gas & Electric*, Pacific Economics Group, LLC, December 2006.

Y is a **dependent variable**—the variable we are studying (for example, an employee’s wage). The **independent variables** X_1 through X_n (also called explanatory variables), which are the factors that determine the value of Y, are on the right-hand side of the equation. The independent variables in the wage model include education, experience, training, and all other variables that the researcher believes are determinants of—that is, influence—wages. The researcher should try to make the list of independent variables as complete as possible so that the model explains as much variation in the wages across different individuals as possible. Each independent variable has a coefficient (b_1 through b_n). We discuss the meaning of the coefficients below.

The right-hand side of the equation also contains a **constant term** (b_0) and an error term (u). The constant term represents the value that the model predicts for the dependent variable if *all* the independent variables are equal to zero. Referring to the wage example, the constant term would be the wage for a person with no education, work experience, or training, such as the legal minimum wage.

The **error term** contains all unobserved factors that affect the dependent variable and are not explicitly incorporated into the model because the analyst is unaware of the relationships or lacks data (for example, we do not include an individual’s innate ability, since we cannot measure it).

An econometric model that explains wage levels can take the following form:

$$Wage_i = b_0 + b_1 Education_i + b_2 Experience_i + b_3 Training_i + b_4 Female_i + u$$

Coefficients b_1 through b_4 reflect the weights associated with each independent variable. Put another way, they indicate how a change in one of the independent variables affects the magnitude of the dependent variable. For instance, the coefficient of b_1 shows how much an additional year of education changes the wage of an average individual while all other factors (experience, training) remain unchanged. Our goal is to estimate (i.e., determine) the value of the unknown coefficients of the model ($b_0 - b_4$). The sign of the estimated coefficient (plus or minus) determines whether the dependent and independent variables are positively or negatively related. The value of the estimated coefficient measures the strength of the relationship of each independent variable to the dependent variable (i.e., its influence on the dependent variable).

In order to carry out the estimation of the coefficients, we need data on dependent and independent variables. The estimates of the coefficients capture the relationships that exist in the data between different variables. For our wage model we need data on wages, education, experience, training, and gender for each individual in the data set. Having data on a larger number of individuals and for more years allows us to capture the relations between variables more precisely. After we compile the data, we need to validate it. Data validation is the process of identifying potential outliers or errors in the data. We might define the data for our example to be wages measured in thousands of dollars per year, years of education ranging from 1 to 18, number of years of training, and a variable for gender called Female that takes a value of 0 for males and 1 for females.

To estimate the unknown parameters of the model—that is, the coefficients—we apply one of several estimation techniques. The researcher can use different techniques to estimate econometric models. The choice of the estimation technique depends on the assumptions made by the researcher and on the available data.

Each of these techniques has the same purpose: to make an estimate of the relationship between the independent and dependent values. We can use this estimate to make predictions—in this example, the change in wages resulting from an additional year of education. To make the estimate useful, i.e., predictive, we want it to be as close as possible to the actual observations. We want to minimize the differences between the observed (in the data) values of the dependent variable Y and the values of this variable estimated by the econometric model (\hat{Y}).

One such technique is **ordinary least squares (OLS)**. This is the most basic and most commonly used estimation method.¹³ The following chart illustrates the difference between observed and estimated values for an equation that explains the variation of wages relative to years of education. Our estimate takes the form of a line on a graph (because a line is the result of an equation). The OLS technique helps us to draw a line that minimizes the distances between the observed data (the points on the graph) and the line we are drawing.

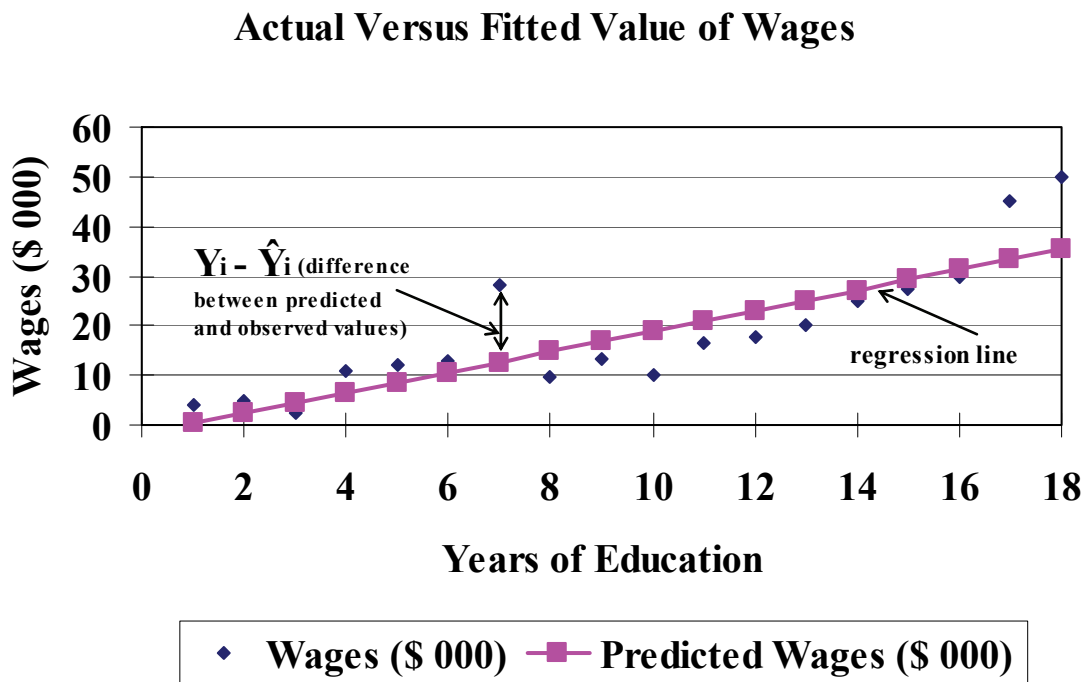


Figure 1. Observed values and estimated values of the dependent variable (wage).

¹³ For a more detailed discussion of the OLS method and a discussion of other available estimation methods, please refer to the econometrics reference volumes provided in Appendix II.

The vertical axis presents the values of the dependent variable, i.e., wage. The horizontal axis presents years of education. Other factors (experience, training, gender) are kept constant and thus do not appear on the figure. We keep all other factors constant to separate the effect of an additional year of education from all other factors. Thus, the graph shows the effect of an additional year of education for individuals of the same gender with the same level of experience and training. Values observed in the data are presented by the points located in different parts of the graph.

The econometric model cannot provide precise estimates of wages due to the presence of the unobserved factors that affect wages and are not captured in the model. The best we can do is make the gap between these two sets of values as small as possible. This is the minimization problem that the estimation technique solves. The differences between the estimated values and the observed values for each observation are vertical distances shown on the graph as $Y_i - \hat{Y}_i$. We want these vertical distances to be as small as possible. However, it is not possible to minimize *all* of the distances because there is no straight line that can go through all data points at the same time. Thus, we minimize some combination of the distances. OLS uses the sum of the *squares* of the distances.¹⁴ OLS uses squares of the distances to avoid the use of absolute values that would complicate the calculation process.

The line that minimizes this sum is called the **regression line**. This is the line that comes closest to passing through all of the points. The regression line reflects the relationship between wages and years of education. The estimates of the unknown parameters (b_0 - b_4) serve as the **solutions** to the minimization problem.

For illustrative purposes, the results of the estimated wage equation are as follows:

$$\text{Wage} = 200 + 0.5 \text{ Education} + 0.1 \text{ Experience} + 0.2 \text{ Training} - 0.00005 \text{ Female}$$

We can interpret the results in the following way. The 0.5 coefficient on education indicates that wage increases by \$500¹⁵ with an additional year of education, given that all other factors remain constant. We interpret the coefficients one at a time while all other factors remain constant, to separate the effect of a particular factor from all other factors. In other words, if two individuals have the same level of experience, training, and gender, but the first individual has one more year of education than the second individual, the model predicts that the wage of the first individual will be \$500 higher than the wage of the second individual.

Continuing through the equation: Wages increase by \$100 with one additional year of experience, given that all other factors remain constant; and by \$200 with an extra year of training, given that all other factors remain constant. The coefficient of negative 0.00005 on gender indicates that females earn \$0.05 ($0.00005 * 1000$) less than males with the same level of education, experience, and training.

¹⁴ Mathematically, OLS minimizes $\Sigma(Y_i - \hat{Y}_i)^2$. For more details and calculations, please see the econometrics books provided in Appendix II.

¹⁵ We express wages in thousands of dollars.

When interpreting the results of the OLS estimation, in addition to examining the signs and magnitudes of the coefficients, we should also look at the **statistical significance** of these coefficients. There is a danger that the results obtained by the estimation might have been caused by accident, i.e., be random events. When a result is statistically significant, it means that the researcher is sure that the result is reliable, i.e., not caused by accident—not explained by a variable. The statistical packages routinely report statistics that help to assess the statistical significance of the coefficients along with the coefficients themselves. P value is one of these statistics.¹⁶ P value is an estimate of the probability that the result has occurred by chance, i.e., that the independent variable played no role in the outcome. In order to assess the statistical significance of our results, first we should determine the statistical level threshold, i.e., the level at which we are ready to tolerate the possibility of chance. The researcher has to choose an appropriate level of statistical significance. The conventional values most commonly used in practice are 0.01 (1 percent), 0.05 (5 percent), and 0.1 (10 percent). The value of 0.01 means that there is a 1 percent chance that the result was accidental; the value of 0.05 means that there is a one-in-twenty chance that the result was due to accident. Then we compare the p value on the coefficient provided by the statistical package against the threshold. For example, we set the statistical level at 0.05. Let us say that the statistical package reports that the coefficient on gender has a corresponding p value of 0.02. In this case the test of statistical significance indicates that $0.02 < 0.05$; thus we conclude that the result is statistically significant at the 5 percent level. However, if our threshold is 0.01, then $0.02 < 0.01$ is not true and we conclude that the result is not statistically significant at the 1 percent level.

Significance is a statistical term that indicates how sure the researcher is that the difference or the relationship exists. The existence of a statistically significant relationship tells only half the story. For example, we might find that the coefficient on the gender is statistically significant, but the coefficient itself is only 0.00005. This means that females on average earn \$0.05 less than males a year; this result is not very important from a practical point of view. On the opposite end, let us say that we find that the coefficient is large (for example, an additional year of education increases wage by \$5,000 a year), but not statistically significant at any conventional levels. In this case the finding is not meaningful, since it is unlikely to be true. Thus, it is important to look both at the magnitude of the coefficients and their statistical significance.

In addition, when examining the results of the OLS estimation, the researcher should look at a measure of how well her model (i.e., the right-hand side of the equation, consisting of the variables and their weights) explains the dependent variables. For the OLS estimate, the statistic R^2 (R-squared, also called the coefficient of determination) serves this purpose. R^2 is the **proportion of the total sample variation in the dependent variable** that is explained by the independent variables (as they are weighted in the model). Total sample variation is a measure of how spread out the observations on the dependent variable are in the sample, i.e., how far individual observations are from the average value of the dependent variable. The value of R^2 is

¹⁶ Other statistics routinely reported by a statistical package along with p-values for each OLS coefficient are t-statistics and confidence intervals; these statistics also convey information about the statistical significance of the coefficients. For more information, please refer to the econometrics books listed in Appendix II.

always between 0 and 1; when interpreting R^2 we usually change it into a percentage by multiplying it by 100. For example, if the R^2 that we obtain is 0.2, it means that the explanatory variables we employ (education, training, experience, gender) explain only 20 percent of the variation in the observed levels of wages. The other 80 percent is left unexplainable due to the presence of unobserved factors. An R^2 of 1 means that all changes of the wage level are completely explained by changes in the explanatory variables that we have. In this case the unobserved factors do not have any effect. An R^2 of 0 indicates that our model did not find any relationship between the dependent variable and independent variables. Statistical packages report R^2 along with the results of the OLS estimation.

Our econometric model assumes **linear relationships** between wage and explanatory variables. In a linear relationship, changes in an explanatory variable produce a corresponding change in the dependent variable. For instance, a linear relationship between wage and experience means that a one-year increase in experience results in a \$1,000 increase in wage per year. Graphically, a linear relationship between a dependent and an independent variable is represented by a straight line. (Figure 1 represents a linear relationship between wage and years of education.)

The researcher can choose another form of relationship if she believes that it fits the theory and the data better. For example, she might believe that experience has a **diminishing effect** on wage, i.e., each additional year of experience yields a smaller and smaller increase in wage (the second year of experience is worth less than the first year of experience, the third year of experience is worth less than the second year, and so on). In this case, she will use a **quadratic relationship** instead of a linear relationship between wage and experience to show the effect. The researcher can do this by including a quadratic term (i.e., a term with an exponent) of the variable on experience into the equation, as in the following equation:

$$Wage_i = b_0 + b_1 Education_i + b_2 Experience_i + b_3 Experience_i^2 + b_4 Training_i + b_5 Female_i + u$$

For the discussion of other possible forms of the relationships between variables in econometric models, please refer to the books provided in Appendix II.

This subsection, admittedly, has provided only an overview of basic econometric concepts in order to facilitate understanding of the use of econometrics in measuring utility productivity. There are many econometric issues that lie outside the scope of this paper. We encourage the reader who wants additional information to refer to econometrics texts, including those listed in Appendix II.

B. The use of econometrics in the measurement of utility performance

The **indexing** method discussed in Part I of the paper employs the **performance of the peer group** as the benchmark with which to evaluate a target utility's performance. The **econometric** method employs a different sort of benchmark. It estimates the performance level that an **average utility** would have if subject to the business conditions of the target utility. The estimated performance level of an average utility then serves as a benchmark against which the researcher compares the target utility's actual performance.

In this section of the paper we will rely on the example that evaluates utility O&M cost performance. Below is the list of the steps taken to apply the econometric method in measuring utility performance. The first six steps cover the estimation of the econometric model, and the last two steps (steps 7 and 8) cover the performance analysis that employs the results of econometric estimation from the previous six steps. After listing all the steps we will talk about each step in more detail.

1. Implementation steps

a. Estimation of the econometric model

Step 1. What do we want to measure? Choose a dependent variable—for instance, O&M costs per customer.

Step 2. Which factors affect the variable that we measure? Choose independent variables—for example, share of underground lines, percentage of self-generation, natural gas operations, composition of load by customer class, input prices (for fuel, labor), weather, and facilities' age.

Step 3. Specify the econometric model. Describe the relationship between the dependent variable and explanatory variables with the help of an equation.

Step 4. Collect data. Obtain information on the dependent and all the independent variables used in the model.

Step 5. Estimate the model. Find values for each variable (the coefficients, the constant, and the error term) with the help of a statistical package.

Step 6. Examine the results of the estimation. Evaluate the signs and magnitudes of the obtained coefficients and the coefficients' statistical significance to understand what role each independent variable plays in determining the level of the dependent variable, and the measure of how well the model explains the variations in the dependent variable (R^2 for OLS).

b. Performance analysis

Step 7. What is the estimated level of performance of an average utility?

Step 8. Calculate the performance measure.

2. Implementation example

The example below illustrates how to apply the above steps to measure utility O&M cost performance.

Step 1. What do we want to measure?

The first step of the process is to decide what aspect of utility performance to measure. As in the case of the indexing method, it is possible to evaluate different performance factors (cost efficiency, productivity, or reliability) as long as the information is quantifiable. In each case the analyst should build an econometric model designed specifically to estimate a particular performance factor. To illustrate the application of the econometric technique in this paper, we build an econometric model that examines operations and maintenance (O&M) costs. Utility O&M costs per customer is a dependent variable in our econometric model.

Step 2. Which factors affect the variable that we measure?

In our example, we are building an econometric model that quantifies the relationships between O&M costs per customer and the variables that explain the level of these costs. The researcher might consider the following explanatory variables for inclusion in her O&M costs model:

a) *UNDERGROUND SHARE*¹⁷. Ratio of underground lines to total lines. The literature¹⁸ indicates that it is less costly for a utility to operate and maintain underground lines compared to above-ground lines. Underground lines are less susceptible to damage from factors such as weather and animal activity.

(b) *SELF-GENERATION*. Percentage of utility's self-generation—the ratio of electricity produced by the utility itself to the utility's total kWh sales. The variable reflects the potential economies from **vertical integration**. Vertical integration is the common ownership of businesses at different stages of production (generation, distribution, and transmission in case of an electric utility).

(c) *NAT GAS*. Percentage of revenues generated from sales to natural gas customers of a utility that provides both electric and natural gas service. The variable captures **economies of scope**. Economies of scope is an economic concept that states that the total cost of production decreases as the number of different goods produced increases. As an example, the total cost of producing both hamburgers and French fries by McDonalds is lower than the costs of two firms producing these two goods separately. The explanation is that the same food storage, preparation facilities, employees, and other assets are shared in the production of both goods. In the case of an electric utility, electric and natural gas divisions can share some of the costs, such as load forecasting, customer billing, and customer service expenses, thus reducing the total costs of the utility.

¹⁷ The notation in capital letters in front of each variable shows how the variable enters the econometric model that we set in the next step.

¹⁸ *Benchmarking the Costs of Ontario Power Distributors*, Pacific Economics Group, 2008, p. 50.

(d) *CUSTOMER CLASS*. Composition of load by customer class—the ratio of kWh sales to residential and commercial customers. The percentage of sales to residential and commercial customers captures differences in the cost of servicing these customers compared to the cost of servicing industrial customers.

(e) *INPUT PRICES*. The variable reflects how changes in input prices (wages, prices of materials and services) alter the cost of production.

(f) *WEATHER*. Unusual weather conditions in a utility’s service territory, such as hurricanes and snowstorms, require more maintenance efforts, thus increasing costs. The researcher can use damages to the territory of the utility’s operations (measured in thousands of dollars) from major weather events (storms, lightning, tornadoes, floods) as a measure of the weather effect.

(g) *AGE GEN, AGE DISTR, AGE TRANSM*. We include separate variables that reflect the age of different types of facilities (generation facilities, distribution facilities, and transmission), since the younger systems are typically less expensive to operate and maintain (they require fewer repairs). The researcher can use the remaining years of life of different facilities as a measure of the facilities’ age.

The above list of explanatory variables illustrates the variables the researcher might consider including in her model. She might add other variables if she believes that they affect the level of O&M costs, or she might exclude some variables from the above list. In practice, it is impossible to include all the factors that affect a utility’s costs. Some of the factors are non-measurable, and data on some of the variables might not be available. The above list is applicable to the O&M costs model only. Models that estimate other aspects of utility performance (for example, reliability) require the use of different explanatory variables.

Step 3. Specify the model.

Model specification is the derivation of the mathematical form of the economic relationships quantified using econometrics. We can express the O&M model by the following equation:

$$O\&M\ Costs_i = \beta_0 + \beta_1 UNDERGROUND\ SHARE_i + \beta_2 SELF-GENERATION_i + \beta_3 NAT\ GAS_i + \beta_4 CUSTOMER\ CLASS_i + \beta_5 INPUT\ PRICES_i + \beta_6 WEATHER_i + \beta_7 AGE\ GEN_i + \beta_8 AGE\ TRANSM_i + \beta_9 AGE\ DISTR_i + u_i$$

This econometric model reflects the relationships between utilities’ costs and the business conditions they face. The model in our example assumes linear relationships between O&M costs and explanatory variables. (If one or more factors had an exponent, the relationship would not be linear; it would be quadratic).

Step 4. Collect data.

For an econometric analysis, it is advisable to have data for as many years and as many utilities as possible. This is one of the differences between the indexing approach and the

econometric approach. As Part I mentions, the indexing approach requires the use of peer utilities that have very similar business conditions. In the case of econometrics, having more utilities with diverse business conditions in the data set only improves the precision of the model's estimates.

Data sets come in different forms. The data sets employed in performance measurement usually come in the following two forms: **cross-sectional data** and **panel data**. Cross-sectional data consists of a sample of utilities (individuals, countries) taken at a given point in time (e.g., the cross-sectional data that consists of information on O&M costs, kWh sales, and other variables for 100 utilities for the year 2005). Panel data follows the same utilities over a certain period of time (e.g., panel data that includes information on O&M costs, kWh sales, and other variables for 100 utilities for each year from 1990 to 2005). It is important to distinguish between different data structures because their features affect the estimation process.

We need data for the dependent and all the independent variables employed in the econometric model for all the utilities in the sample. The FERC Form 1 reports cost data on a functional basis. Other data (amount of own generation, natural gas customers, number of customers of different classes, underground lines percentage) are available from FERC Form 1, Form EIA - 861, Form EIA-923, and Form EIA-860. The Bureau of Economic Analysis Regional Economic Information System offers a source of regional, state, and county-level wage and employment data at the regional, state, and county level by industry (i.e., the North American Industry Classification System). The National Climatic Data Center maintains localized weather data, including damage from unusual weather events, number of storms, heating degree days, cooling degree days, maximum degree days, and precipitation.

Step 5. Estimate a model with the help of a statistical package.

Once the researcher sets up the model (meaning the equation with all its variables but without any numbers yet) and collects the necessary data, she can use a statistical package such as STATA, SPSS, or SAS to “estimate” the model, i.e., to determine the values of the unknown coefficients, the constant, and the error term.

Step 6. Examine the results of the estimation.

The statistical package will produce values for each unknown term, along with p values and a single R squared value for the equation. Using this information, the analyst can then present the results as an equation, such as the following:

$$\begin{aligned}
 O\&M\ Costs &= 15 - 0.5\ UNDERGROUND\ SHARE - 0.3\ SELF-GENERATION + \\
 &\qquad\qquad\qquad (p\text{-value} = .043) \qquad\qquad\qquad (p\text{-value} = .061) \\
 0.6\ NAT\ GAS &+ 0.8\ RESIDENTIAL\&COMMERCIAL + 1.5\ INPUT\ PRICES + \\
 (p\text{-value} = .005) &\qquad\qquad\qquad (p\text{-value} = .612) \qquad\qquad\qquad (p\text{-value} = .003) \\
 0.2\ WEATHER &+ 0.2\ AGE\ GEN + 0.5\ AGE\ TRANSM + 0.1\ AGE\ DISTR \\
 (p\text{-value} = .501) &\qquad\qquad (p\text{-value} = .001) \qquad\qquad (p\text{-value} = .001) \qquad\qquad (p\text{-value} = .004) \\
 R^2 &= .832
 \end{aligned}$$

The constant term is estimated by the statistical package to be 15¹⁹. This means that if all the explanatory variables have a value of zero, the O&M cost per customer is \$15. This result can be interpreted to be an estimate of fixed costs that the utility continues to incur regardless of production level.

Coefficients $\beta_1 - \beta_9$ reflect the magnitude and direction of the effects of each explanatory variable on costs. The value of the coefficient on underground lines of 0.5 shows that O&M costs per customer decrease by \$0.5²⁰ in response to the 1-percent increase in the share of underground miles, given that all other factors remain constant. The coefficient is statistically significant at a 5 percent level.

We interpret other coefficients similarly. O&M costs per customer decrease by \$0.3 when the degree of the utility's self-generation increases by 1 percent, indicating the presence of economies of scope. The coefficient is statistically significant at 10 percent²¹. Costs increase by \$0.6 when the percentage of sales to natural gas customers increases by 1 percent. The result is statistically significant at 1 percent. The next coefficient that reflects the customer composition of the utility load indicates that costs increase in response to a 1 percent increase in sales to residential and commercial customers. The result is not statistically significant at any conventional level (i.e., 10 percent or less); thus, we do not consider the relationship between the level of costs and customer classes reliable. Costs also increase with an increase in the input prices; the result is statistically significant at 1 percent. The coefficient on the weather effects is not statistically significant at any conventional level. The coefficients on the age of different types of facilities indicate that it is more expensive to operate and maintain older facilities; all three coefficients are statistically significant at 1 percent.

The value of R^2 indicates that the model explains about 80 percent of the variation in O&M costs.

Step 7. What is the expected level of performance of an average utility?

Once we have estimated the model, we can obtain the costs that an average-performing utility would have if subject to the business conditions experienced by the target utility. Let us say that we want to assess the cost performance of Utility X in 2008. First, we find the values in the data of all explanatory variables for this utility for the year 2008. The values of these explanatory variables reflect the business conditions that Utility X was under in 2008.

¹⁹ All the numbers in this example are for illustrational purposes only.

²⁰ In our example, we measure O&M costs in dollars.

²¹ While 10 percent is one of the levels of statistical significance used in practice, the researcher should decide for herself whether or not to consider this level of significance acceptable (and the found relationship reliable).

<i>UNDERGROUND SHARE (%)</i>	20
<i>SELF-GENERATION (%)</i>	10
<i>NAT GAS (%)</i>	0
<i>RESIDENTIAL&COMMERCIAL (%)</i>	60
<i>INPUT PRICES (\$ per customer)</i>	5
<i>WEATHER (\$ in thousands)</i>	5
<i>AGE GEN (years)</i>	12
<i>AGE TRANSM (years)</i>	14
<i>AGE DISTR (years)</i>	15

Next we substitute these values into the estimated model and multiply them by the corresponding coefficients.

$$O\&M \text{ Costs}_{Utility X 2008} = 15 - 0.5 * 20 - 0.3 * 10 + 0.6 * 0 + 0 [RESIDENTIAL\&COMMERCIAL \text{ is not statistically significant thus we do not take this relationship into account}] + 1.5 * 5 + 0 [WEATHER \text{ is not statistically significant}] + 0.2 * 12 + 0.5 * 14 + 0.1 * 15 = 20.4$$

The result is the O&M costs per customer for an average-performing utility, given Utility X's 2008 business conditions.

Step 8. Calculate the performance measure.

Now we can compare the actual costs of Utility X and the costs of an average-performing utility as estimated by the model. Let us say that the actual costs that Utility X reported for 2008 were \$30 per customer. We know from the previous step that the average-performing utility's costs for the same year were \$20.4. We assess Utility X's performance as follows:

$$Performance = \text{Actual Costs}_{Utility X 2008} / \text{Estimated Costs}_{Average utility 2008}$$

The ratio is 30/20.4=1.47. The value of more than 1 indicates that the performance of Utility X is worse than the performance of an average-performing utility subject to the same circumstances. An average-performing utility would have 47 percent lower costs. A performance value of less than 1 would indicate above-average performance.

C. Advantages and limitations

One of the **advantages** of the econometric method is that it does not require the creation of a peer group. As we saw in Part 1 regarding the indexing approach, it is not always possible to find an adequate group of similar utilities. The econometric method employs information on a group of utilities with diverse operating and business conditions by controlling for many business conditions simultaneously. In general, the econometric approach provides more precise results than the indexing approach.

At the same time, the econometric modeling approach is more complicated than the indexing method. That fact leads to the following **limitations**. The method's implementation requires the analyst to have a statistical and econometric background. The complicated nature of the approach limits the transparency of both the results and the process of estimation.

The approach is also expensive. It is time-consuming to compile and validate a database from either primary or secondary sources. Modeling is also expensive: There is the one-time cost of obtaining software, and, if there are multiple users, site license requirements can increase procurement costs. The initial process of modeling requires time and effort. Once the model is in place, however, it can be modified with less effort (for example, when it is re-estimated with more recent data).

The results and accuracy of the econometric model can suffer from a number of problems. For example, the prediction of a utility's cost from a statistical cost model may not accurately reflect the benchmark (i.e., the expected cost of an average-performing utility) for the following reasons: a) errors in the estimated parameters measuring the effect of outside factors on cost; b) errors in capturing the functional form of the relationship between costs and business conditions; c) failure of the model to include all relevant business conditions; d) data errors. The model might not include all of the factors that affect performance, as some of them cannot be measured. For example, if there is no quantifiable information on quality of service, then we cannot include this factor in the model.

Finally, econometric analysis requires a larger amount of data than the indexing approach—the researcher has to obtain data for all the variables for each utility in the sample.

D. Application examples

In the example, we used econometric analysis to measure O&M costs. We can use the method to measure other aspects of utility performance as well. For example, in an investigation of the causes of a large outage, a commission can use the method to evaluate the particular utility's level of reliability. In this case, our dependent variables would be indicators of the reliability of the utility's services (for example, duration of outages or frequency of outages). The list of explanatory variables would include factors affecting the reliability level of electric service (such as weather conditions in the territory of the utility's operations, composition of customer load, or composition of utility lines). As a result, we would be able to compare the particular utility's reliability with the level expected from an average-performing utility. We can also use the method to measure various categories of costs. For example, in the same outage investigation we can evaluate the utility's spending on tree trimming and other reliability-related expenses (dependent variable).

III. Frontier Methods: Data Envelopment Analysis (DEA)

We can divide benchmarking methods into two broad categories: **average methods** and **frontier methods**. Average methods compare the performance of a target utility against the *average performance* of other utilities. The indexing and econometric methods described in the previous two parts of this paper belong to this category. Frontier methods use as their basis for comparison the *most efficient performance level* among the utilities in the sample. This section focuses on data envelopment analysis (DEA), which is the most commonly used frontier method.

A. Basics of the DEA method

The DEA method is based on the idea that if a particular firm can produce X units of output by using Z units of input, then other firms should be able to do the same; otherwise, they do not operate efficiently. Inefficient producers can improve their performance by either producing more output with the same amount of input or using fewer input units to produce the same amount of output.

The best way to describe the results of the DEA approach is through a graphical illustration. The graph that follows provides an example of DEA.

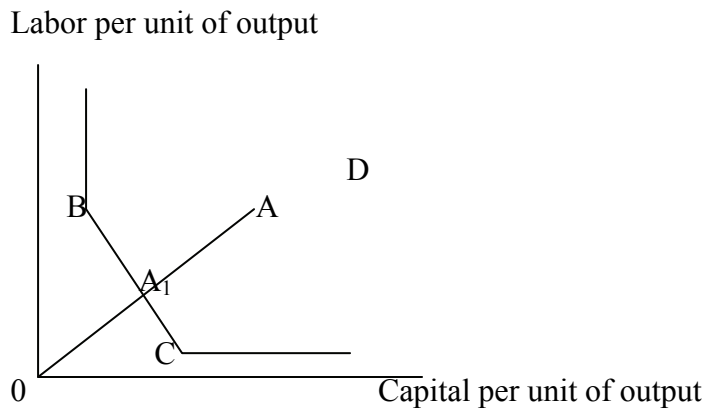


Figure 2. DEA output.

We use an example of only two inputs and one output for illustration purposes, since it is easier to demonstrate such a simple arrangement on a two-dimensional graph. However, DEA allows the use of more types of inputs and outputs. We are looking at four utilities in our example: Utility A, Utility B, Utility C, and Utility D. Each utility combines two inputs (capital and labor) to produce the same amount of output (kWh). The vertical axis represents the amount of labor used to produce a unit of output (one kWh). The horizontal axis represents the amount of capital used to produce one unit of output. Points A, B, C, and D represent the actual combination of capital and labor used by each utility.

DEA examines all the data points in the sample to find the points with the lowest total cost (capital and labor) to produce one unit of output and then connects those points to form **the frontier line** (also called the efficiency frontier). The frontier line represents combinations of inputs (labor and capital) that minimize the cost of producing one unit of output. In our example, the line segment BC defines the frontier line. The graph shows that Utility B and Utility C use different mixes of capital and labor, but in combinations that minimize the total cost to produce one kWh. Hence, both are efficient. Any utility operating with a combination of capital and labor that places it on the frontier line is efficient. Utilities located on the line represent the benchmark for measuring performance. A firm whose combination of unit inputs and unit outputs places it to the right of the frontier line is inefficient. Firm A uses the same amount of labor as B, but requires more capital; thus, it is inefficient. D is inefficient because it uses more of both labor and capital than B and C.

The degree to which a utility is inefficient is measured by its distance from the frontier line. For example, in the graph, the distance between point A and the frontier represents a measure of Utility A's inefficiency (AA_1 on the graph). The graph indicates room for improvement for Utility A. Utility A can become as efficient as the benchmark, which is defined by the frontier, by reducing labor and capital inputs, and therefore the total cost of production, to reach point A_1 on the frontier.

We assign the value of one to the efficiency of the best performers—those that are located on the frontier (B and C in our example). Less efficient utilities have scores below one. The efficiency of Utility A is measured as OA_1/OA . Assume that Utility A has an efficiency measure of 0.40. This means that the utility's efficiency is 60 percent below the benchmark. To be efficient relative to its peers, the utility needs to reduce both its capital costs and its labor costs by 60 percent.

The DEA scores measure *relative* efficiency. A score of one does not indicate the best possible performance; rather, it represents the experience of the most efficient utilities in our data set in comparison to other utilities in the sample. Likewise, inefficient firms are inefficient only relative to their peer group from which the frontier is estimated.

The DEA method uses **linear programming** to calculate the efficiency level of each firm in the sample. Linear programming²² is a process for determining the best outcome (i.e., minimum costs or maximum output) by solving the problem of maximizing or minimizing a linear function²³ subject to one or more constraints. In the electric utility industry, the focus is normally on cost minimization, since output level is fixed by the utility's obligation to serve all load in its service territory. That is, an electric utility usually improves efficiency by minimizing its costs given the level of output it is required to serve (i.e., its statutory obligation to serve all load in its service territory) rather than by maximizing its output given the level of costs. In the

²² For details on the calculation process of the linear programming used in DEA, please refer to the texts listed in Appendix II.

²³ The functional expression of a linear equation is that of a straight line: $f(x) = mx + b$.

two-input example, linear programming would be used to identify whether the utility has minimized the total cost of capital and labor required to produce one kWh of electricity.

Various software programs are capable of performing linear programming. These include Excel, Matlab, SAS, STATA, and SPSS. Specialized DEA packages are also available, such as ONFront, IDEAS, Frontier Analyst, Warwick DEA, and DEAP.

In our generic example we use two types of inputs (labor and capital) and one type of output (kWh). In the actual analyses of electric utility performance the list of inputs used in DEA can include labor costs, capital costs, production expenses, transmission expenses, distribution expenses, administrative and general expenses, number of employees, installed capacity, and length of network lines. The analyst can use the following output variables: number of customers, retail and wholesale sales (kWh), and total service territory.

Factors that are outside of a utility's control, such as customer composition of utility load and customers per transmission or distribution line mile, can affect the DEA efficiency score. For example, serving residential customers might require more inputs per kWh; thus, utilities that serve many residential customers might have higher total costs. Higher customer density (a large number of customers per line mile) results in lower utility costs due to lower line losses per customer and fixed costs being more spread out across customers. DEA efficiency scores might be partially driven by differences in a utility's business conditions. DEA is capable of taking into account such factors to ensure that DEA efficiency scores reflect utilities' efforts rather than differences in business conditions across utilities²⁴.

B. Implementation steps

In general, we can represent the DEA application with the following **implementation steps**:

Step 1. Decide which input, output, and control variables to use in the model. For example, the analyst can use the number of workers and the amount of capital (lines and transformers) as inputs; number of customers, sales, and network length as output variables; and customer density and proportion of residential customers as control variables.

Step 2. Obtain data. The accuracy of the data is critical to the DEA results. The analyst can obtain most of the data on input, output, and control variables from FERC Form 1 and from Form EIA-861 and Form 923.

Step 3. Create an efficiency frontier with the help of a computer model. Use one of the software programs listed above.

²⁴ Please see *An Introduction to Efficiency and Productivity Analysis* by Timothy J. Coelli, Dodla Sai Prasada Rao, Christopher J. O'Donnell, and George Edward Battese for the discussion of how factors that are outside of a utility's control can be included in the DEA estimation process.

Step 4. Calculate the efficiency score for a target utility as the distance from the frontier. We assign the score of one to the efficient utilities located on the frontier line and scores below one to reflect the degrees of utilities' relative inefficiency.

C. Extensions of the basic DEA model

The basic DEA model assumes **constant returns to scale**. If increasing all inputs by some proportion results in output increasing by that same proportion, then the production is characterized by constant returns to scale. For instance, if increasing the use of capital and labor by 10 percent results in an increase in production of 10 percent, we have constant returns to scale.

The analyst might not believe the assumption about the constant returns to scale. For instance, the analyst might believe that the scale of production affects the changes in costs, i.e., as production increases, the cost of producing each additional unit falls (**increasing returns to scale**). This is the case when, for example, at the level of production of 1,000 units of output the increase in capital and labor by 10 percent results in an increase in production by 10 percent, whereas at the higher level of production of 2,000 units of output the increase by 10 percent in inputs corresponds to an increase in production by 12 percent. The extension of the DEA basic model allows the analyst to take different returns to scale into account.

The basic DEA model measures **technical efficiency**. Technical efficiency reflects the effectiveness with which a utility uses a given set of inputs to produce output. When the firm is producing the maximum output possible given the inputs it employs, as well as the best technology available, it is technically efficient. We can extend the basic DEA model by including data on input prices. This would allow us to assess a firm's **allocative efficiency** (i.e., how well the firm chooses the input mix given the input prices.)

For a discussion of other possible extensions of the basic DEA model, see *An Introduction to Efficiency and Productivity Analysis* by Coelli, Rao, O'Donnell, and Battese and other books listed in Appendix II.

D. Advantages and limitations

One of the advantages of data envelopment analysis compared to the econometric technique is that DEA does not require the analyst to specify a mathematical form for the relationship between inputs and outputs. Econometrics requires an understanding of the functional form of the relationship between the dependent variable and independent variables (i.e., linear, quadratic, or other). DEA requires only the selection of outputs and inputs for a linear programming model, but does not require specification of the relationships between these inputs and outputs.

Among the strengths of DEA is that DEA is less data-demanding than the econometric approach, as DEA works well with a small sample size. One rule of thumb commonly applied for DEA suggests that the number of observations (utilities, in our case) in the data set should be

at least three times the sum of the number of input variables and output variables²⁵. Another rule of thumb suggests that the number of observations should be at least twice the product of the number of inputs and outputs²⁶. Econometric models in general require a larger number of observations to obtain statistically significant results. The other advantage of DEA is that we can apply the method to data sets of smaller sizes than required under the econometric approach. Also, DEA allows the use of multiple inputs and outputs. In addition, DEA results provide policy makers with suggestions on how to improve performance. In the cost-minimization case, efficiency scores demonstrate the percentage by which a firm should decrease its inputs to reach the position on the frontier.

At the same time, DEA has a number of disadvantages that one must weigh against its positive aspects. First, DEA regards any utility with a score of less than one as inefficient. However, a low score may be the result of the DEA model's not capturing some factors that affect a utility's efficiency but are beyond its control. These factors might explain why a particular utility uses more inputs than the frontier utilities. In this case, a low score is not an indication of a utility's inefficiency but a result of a model's inaccuracy—its failure to capture efficiency-reducing factors outside the utility's control.

On the other hand, there is a limit to the number of variables that can be used in the DEA model. The number of input, output, and control variables included in the analysis should not be too large in comparison with the number of firms in the data set (see rules of thumb for the number of observations provided above). If the number of observations is too small, DEA would place too many utilities on or close to the frontier. A rule of thumb suggests that a sample is large enough if the number of efficient utilities—utilities placed on the frontier line—is no larger than one third of the sample²⁷. The limit on the number of variables increases the risk that the model will not capture all the factors affecting utilities' costs.

The other problem is that the analyst can erroneously regard outliers as a part of the frontier. Consider a utility that operates under unusually favorable conditions compared to other utilities in the data set (e.g., a utility that operates in a more temperate region may have fewer storm-related outages than other utilities that operate in a service area within a tornado belt). DEA can yield results suggesting that the utility operating in the temperate region is more efficient and thus place this utility on the frontier, when in fact it is weather-related uncontrollable extraordinary costs that make other utilities appear less efficient. DEA might also consider as efficient a utility that deferred some costs that it will inevitably have to incur at some future time.

²⁵ Cooper, W. W., Seiford, L. M., & Tone, K. (2000), *Data Envelopment Analysis*, Boston: Kluwer Academic Publishers.

²⁶ Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001), "Pitfalls and Protocols in DEA," *European Journal of Operational Research* 132: 245-259.

²⁷ Avkiran, N. K. (2002), *Productivity Analysis in the Service Sector with Data Envelopment Analysis*, Camira, Queensland: N. K. Avkiran.

Another disadvantage is that the analyst cannot use well-established statistical tests to assess the robustness of a DEA model. The available tests are elaborate and subject to ongoing research; thus, the reliability of a DEA model might be questionable. This shortcoming is the opposite of the results of an econometric model, which are subject to rigorous statistical tests of significance.

Also, DEA results are very sensitive to the underlying assumptions and data used in the analysis. An example from Dutch electricity regulation illustrates this problem. In 2000 the agency conducted a benchmarking study of Dutch electric distribution utilities and calculated DEA efficiency scores for these utilities. The agency then used the DEA scores to set the productivity factor X in the revenue-cap mechanism. Many utilities disagreed with the way the agency conducted DEA and contested the results of the analysis in court. One of the companies asked the agency to use a different data set in the DEA analysis. The use of another data set resulted in the utility's score increasing from 65 to 95 percent. This example illustrates the instability of DEA results²⁸.

At the outset of this Section III, we mentioned two types of benchmarking methods: average methods and frontier methods. Besides DEA, other frontier methods include correlated ordinary least squares (COLS) and stochastic frontier analysis (SFA). The COLS method estimates the regressions with the help of an econometric technique—ordinary least squares (OLS)—and then shifts this regression to create a frontier. Under this approach, only one firm is considered to be efficient. The method compares all other firms to this utility's efficiency level. SFA is similar to COLS but uses a different econometric technique (the random effects model) and does not rely on a single firm's efficiency level²⁹.

E. Application examples

The DEA approach is more popular in Europe than in North America. Examples of the DEA application can be found in the practice of regulatory agencies of the Netherlands, the UK, Norway, and Australia. The agencies often use the results of DEA to set the productivity factor X in revenue caps and price caps.

²⁸ Irastorza, Veronica, "Benchmarking for Distribution Utilities: A Problematic Approach to Defining Efficiency," *The Electricity Journal*, December 2003.

²⁹ For a more detailed description of these methods please see Coelli, Rao, O'Donnell, and Battese, *An Introduction to Efficiency and Productivity Analysis*; and Kumbhakar and Lovell, *Stochastic Frontier Analysis*.

IV. A Comparison of the Methods

Below is a matrix that compares the methods in terms of the following characteristics: ease of application, reliability of results, and data requirements. The text of the paper provides a more detailed discussion of the methods in terms of their characteristics (in the sections labeled “Advantages and Limitations”). The following matrix offers a way to summarize these characteristics. The matrix shows that if a commission needs to take a quick look at the performance of the utilities, it can apply the indexing approach, which does not require the analyst to possess extensive statistical training and does not require an extensive data set. If a commission is interested in a more thorough review of utility performance, it should consider employing the other two methods.

The following is a general observation and reflects the author’s opinion. In practice, the characteristics of the methods vary depending on a particular application.

Table 1. Comparison of the Methods

Characteristics Method	Ease of application	Reliability of results	Data requirements
Indexing	Easy to apply	Not reliable/Medium reliability*	Low data requirements
Econometrics	Medium/High difficulty*	Medium/High reliability*	High data requirements
Data Envelopment Analysis	Medium/High difficulty*	Medium/High reliability*	More data than for indexing, but less than for econometrics

Note:

** Depends on the situation*

Appendix I: Calculation of a Multifactor Cost Index

The numerator of the index is a cost index:

$$\text{Cost Index} = \ln O\&M \text{ Costs}_{\text{Utility X}} - \ln O\&M \text{ Costs}_{\text{Peer Group Average}}$$

The denominator of the multifactor index is calculated as follows:

$$\text{Output Quantity Index} = 0.46 (\ln \text{Number Customers}_{\text{Utility X}} - \ln \text{Number of Customers}_{\text{Peer Group Average}}) + 0.30 (\ln \text{KWh Deliveries}_{\text{Utility X}} - \ln \text{KWh Deliveries}_{\text{Peer Group Average}}) + 0.24 (\ln \text{km of lines}_{\text{Utility X}} - \ln \text{mean km of lines}_{\text{Peer Group Average}})$$

Now we can calculate the multifactor index:

$$\text{Multifactor Cost Index} = \text{Cost Index} / \text{Output Quantity Index}$$

Appendix II: Recommended Texts for Further Reading

Indexing

An Introduction to Efficiency and Productivity Analysis by Timothy J. Coelli, Dodla Sai Prasada Rao, Christopher J. O'Donnell, and George Edward Battese, Kluwer Academic Publishers.

The Benchmarking Book: A How-to Guide to Best Practice for Managers and Practitioners by Tim Stapenhurst, Elsevier.

Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance by Robert C. Camp, ASQC Quality Press.

Strategic Benchmarking Reloaded with Six Sigma: Improving Your Company's Performance Using Global Best Practice by Gregory H. Watson, Wiley.

The Basics of Benchmarking by Robert Damelio, Productivity Press.

Econometrics

Essentials of Econometrics by Damodar Gujarati, McGraw-Hill/Irwin.

Introductory Econometrics: A Modern Approach by Jeffrey M. Wooldridge, South-Western College Publishers.

Introduction to Econometrics by James H. Stock and Mark W. Watson, Addison-Wesley.

More advanced texts:

Econometrics Analysis by William H. Greene, Prentice Hall.

Mostly Harmless Econometrics: An Empiricist's Companion by Joshua D. Angrist and Jorn-Steffen Pischke, Princeton University Press.

A Guide to Econometrics by Peter Kennedy, Wiley-Blackwell.

Econometric Analysis of Cross Section and Panel Data by Jeffrey M. Wooldridge, MIT Press.

Data Envelopment Analysis

Introduction to Data Envelopment Analysis and Its Uses by William W. Cooper, Lawrence M. Seiford, Kaoru Tone, Springer.

An Introduction to Efficiency and Productivity Analysis by Timothy J. Coelli, Dodla Sai Prasada Rao, Christopher J. O'Donnell, and George Edward Battese, Kluwer Academic Publishers.

Data Envelopment Analysis: Theory, Methodology and Applications by Abraham Charnes, William W. Cooper, Arie Y. Lewin, Lawrence M. Seiford, Kluwer Academic Publishers.

An Introduction to Efficiency and Productivity Analysis by Timothy J. Coelli, Dodla Sai Prasada Rao, Christopher J. O'Donnell, and George Edward Battese, Kluwer Academic Publishers.

Handbook on Data Envelopment Analysis (International Series in Operations Research & Management Science) by William W. Cooper, Lawrence M. Seiford, and Joe Zhu, Kluwer Academic Publishers.