



National Regulatory
Research Institute

How Performance Measures Can Improve Regulation

Ken Costello, Principal

The National Regulatory Research Institute

June 2010

10-09

Acknowledgments

The author wishes to thank his colleagues at NRRI, Ms. Evgenia Shumilkina and Mr. Scott Hempling; Commissioner Ric Campbell of the Public Service Commission of Utah; Dr. Mark Lowry of Pacific Economics Group Research; Professor Karl McDermott of the University of Illinois Springfield; Mr. Joseph W. Rogers of the Massachusetts Attorney General's Office; and Dr. Michael Schmidt of the Kansas Corporation Commission for their comments on an earlier draft of this paper. The author is responsible for any remaining errors in the document.

Online Access

The reader can find this paper on the Web at
http://www.nrri.org/pubs/multiutility/NRRI_utility_performance_measures_jun10-09.pdf.

Executive Summary

Regulation's central purpose is to induce high-quality performance from our utilities. To achieve that objective, regulators must measure and evaluate utility actions.

Performance depends on how well management uses the available resources. Also affecting performance are factors outside management's control.

Uses of Performance Measures

The challenge for regulators is to determine what constitutes a well-performing utility. What do they consider acceptable performance? These are questions that regulators need to address if they are to exploit fully the information contained in performance measures for regulatory actions such as prudence determination and rate setting. The measurement of performance trends in the absence of a standard, for example, might limit regulatory action to further review, not to a determination of cost recovery.

The National Regulatory Research Institute (NRRI) is writing a series of papers on performance. This particular paper helps regulators to form a context, rationale, and a general framework for initiating a strategy to measure and evaluate the performance of utilities in their states. It begins with a discussion on major questions that regulators should address before applying performance measures. The paper also provides guidance to regulators on how to better gauge utility performance in non-cost functional areas such as reliability and other dimensions of service quality. Such evaluation allows regulators to satisfy the objective of consumer protection.

This paper provides regulators with the following information:

1. The rationale for why regulators should measure and evaluate utility performance;
2. Guidance on how regulators can best apply performance measures in various areas of utility operations;
3. General interpretations of utility performance and alternative regulatory responses;
4. Different performance measures that regulators can use;
5. The uses and limitations of different performance measures and performance-measurement techniques;
6. The different regulatory venues for the application of performance measures, both within and outside a rate case; and

7. A general framework and sequence of steps that regulators can take to initiate performance measurement and evaluation tasks.

An Illustration of a Regulatory-Review Process

Figure ES-1 illustrates one way in which regulators can review a utility's performance and take appropriate action. The diagram shows four major things:

1. ***Regulation itself affects utility management behavior.*** Together with factors that fall outside the control of a utility, management behavior determines a utility's performance. Regulatory rules, policies and practices directly and indirectly affect utility performance. Utility performance, in turn, can influence regulatory actions. Poor utility performance, for example, might induce regulators to provide utilities with stronger incentives and disincentives or to establish standards for future performance.
2. ***Regulators should initially assess the utility's performance by comparing actual performance with a pre-specified standard.*** Any substantial deviation can reflect exceptionally good or bad performance. The utility would then have the opportunity to respond to the evidence of bad performance, with subsequent evaluation by the regulator.
3. ***Based on its review, the regulator can then take a particular action.*** The action may affect cost recovery by the utility, lead to a more detailed investigation such as a retrospective management audit or induce the regulator to institute a mechanism that would reward or penalize the utility for exceptional performance. The regulator can take other actions or no action in response to its assessment. One such action might include rewarding the utility for above-average performance that the regulator judged to reflect exceptional management behavior.
4. ***Performance measures can help regulators determine "just and reasonable rates."*** The objective of the proposed regulatory approach is to enhance the ability of state commissions to make informed decisions. Accountability requires regulatory assurance that utility costs incorporated in rates reflect prudent, efficient, effective and customer-responsive management behavior. Accountability also demands that regulators recognize the financial interests of utilities; namely, to permit prudent and efficient utilities a reasonable opportunity to earn a fair rate of return and attract capital to serve the long-term interest of their customers. Performance measures can provide regulators with a tool to achieve these outcomes.

Organization of the Paper

This paper contains six parts. Part I defines “performance.” Part II gives reasons for why regulators should measure utility performance. Part III identifies the challenges that regulators face in interpreting performance measures for various applications. In Part IV, the paper provides an overview of the different techniques for performance measurement. Part V discusses specific applications of performance measurement in different regulatory venues, including rate cases, the development of incentive mechanisms and periodic oversight. The final part lists six steps for executing a regulatory “performance” initiative.

Figure ES-1. A Regulatory Process for Reviewing and Responding to a Utility's Performance

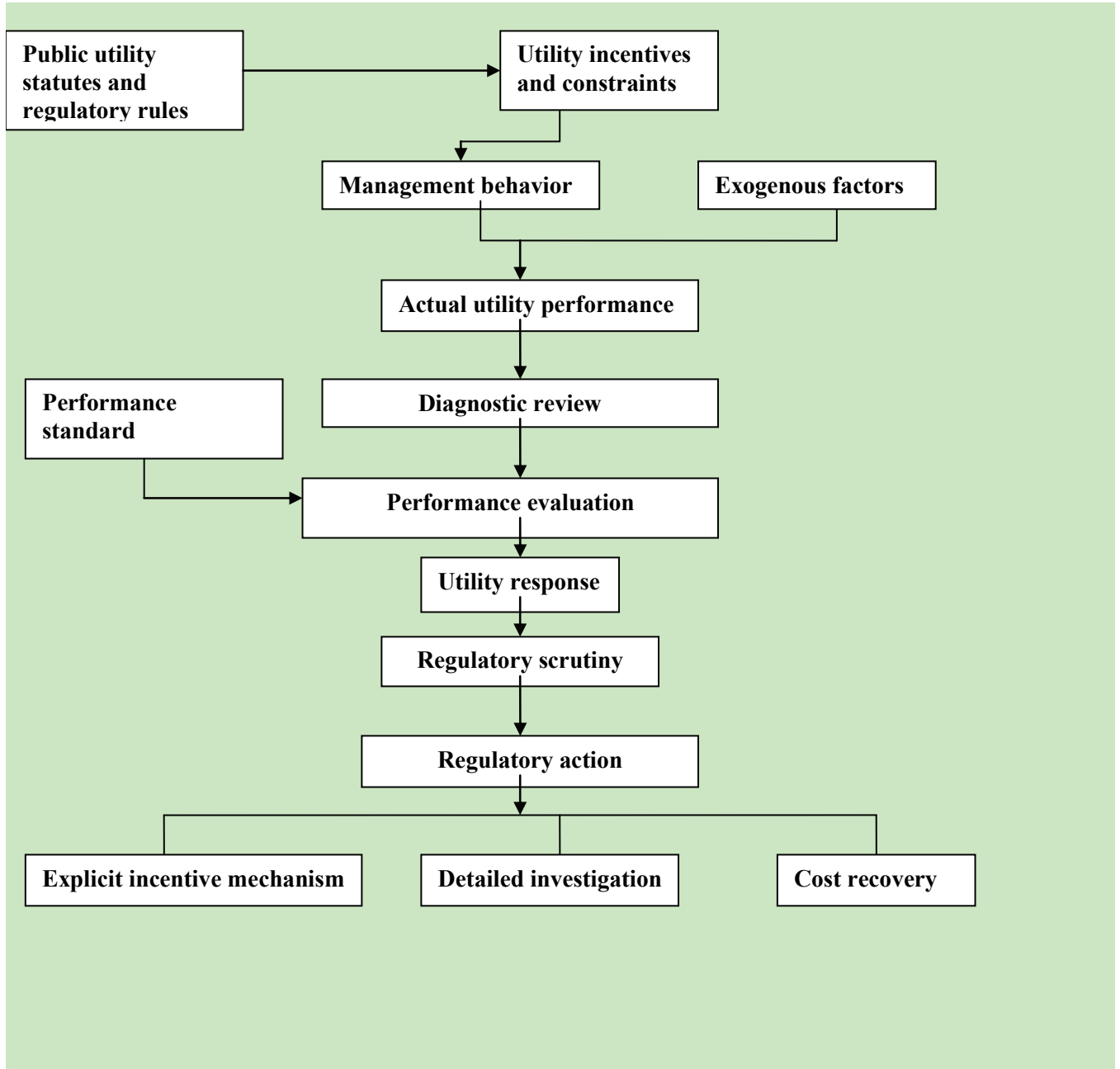


Table of Contents

I.	What Do We Mean by “Performance”?	3
A.	Multi-dimensional nature of performance	3
B.	Performance standards	3
II.	Why Should Regulators Measure Utility Performance?	5
A.	Performance problems under regulation	5
B.	Regulators have an information disadvantage	5
III.	The Challenge Facing Regulators in Measuring and Evaluating Utility Performance	7
A.	Factors affecting utility performance	7
B.	Sports metaphors	8
C.	Regulatory considerations for applying performance measures	8
D.	How might regulators interpret and use the results?	11
IV.	An Overview of Different Techniques for Measuring Performance	14
A.	Attributes of good performance measures	14
B.	Econometric methods, indexing, and data envelopment analysis	15
C.	Additional ways to measure utility performance	16
1.	Management audits	16
2.	Accounting ratios for individual functional areas	16
3.	Uses and limitations of performance measurement	17

V.	Applications of Performance Measures in Different Regulatory Venues.....	19
A.	General uses of performance measures and examples.....	19
1.	Illustration of service quality	19
2.	Illustration of energy-efficiency activities	20
3.	Prudence review.....	20
4.	Evaluation of a regulatory action.....	20
B.	Specific applications.....	21
1.	Regulatory incentive mechanisms	21
2.	Periodic monitoring of utility performance outside a rate case.....	22
3.	Comparison of a utility’s actual performance with a benchmark, both in rate cases and other regulatory forums	23
4.	Evaluation of the reasonableness of “cost-of-service” components, adjustment of the rate of return on equity (ROE), and use of total factor productivity	24
5.	Preliminary review of a utility’s performance to determine further action.....	25
6.	Examination of the reasons for performance differences across utilities.....	26
7.	Publicity of a utility’s performance on a periodic basis	26
VI.	A Six-Step Approach for a “Performance” Initiative.....	27
A.	Identify the uses of performances measures	27
B.	Select utility functional areas for regulatory review.....	27
C.	Calculate the performance measures	27
D.	Compare a utility’s performance with a predetermined benchmark.....	27
E.	Assess a utility’s performance	27

F. Take action.....	27
Appendix A: Price Differentials across Utilities: The Challenge of Detecting Causes.....	29
Appendix B: Real Unit Cost Indices	32
Appendix C: An Illustration of the Use of Performance Measures in a Cost-Sharing Incentive Mechanism.....	34
Appendix D: The Relationship between Total Factor Productivity and Average Cost.....	37

How Performance Measures Can Improve Regulation

Regulation's central purpose is to induce high-quality performance from our utilities. To achieve this objective, regulators must measure and evaluate utility actions, then inject the evaluation's results into regulatory decisions. Measurement can cause better regulatory incentives and improved utility performance. Improved performance, in turn, can lead to lower rates over time, higher quality of service, fewer rate cases, and avoidance of excessive utility costs. Performance measurement can detect subpar utility management that could lead to further investigation, cost disallowances, or a change in regulatory incentives. It can also help regulators determine whether utilities are satisfying stated objectives or targets. Performance measurement can also help regulators reward utilities for superior performance that benefits customers through lower rates or higher quality of service.

Compared to their foreign counterparts (especially European countries),¹ U.S. regulators have relied less on performance measures as a benchmarking tool to set rates and evaluate utility performance. In most U.S. applications, benchmarking has focused on operation and maintenance expenses rather than total cost performance.

In the absence of quantifiable performance measures, it becomes difficult for regulators to know if utilities are falling short of, meeting, or surpassing predetermined objectives or targets. Performance measures can empower regulators to grade utilities, mindful of the limitations of the particular measures for appropriate regulatory actions. Performance measurement can accompany special incentive mechanisms, management audits and other detailed investigations, and specific actions on cost recovery.

This paper addresses several questions. First, it provides reasons for why state public utility commissions (or "regulators") would want to measure utility performance. Next, it identifies the challenge that regulators face in interpreting performance measures for various applications. The paper then provides an overview of the different techniques for performance measurement. A previous NRRI paper detailed some of these techniques. Finally, it identifies specific applications of performance measurement in different regulatory venues, including rate cases, the development of incentive mechanisms, and periodic oversight.

This paper helps regulators by providing them with the following information:

1. The rationale for why regulator should measure and evaluate utility performance;

¹ See, for example, Cambridge Economic Policy Associates, *Background to Work on Assessing Efficiency for the 2005 Distribution Price Control Review*, prepared for Ofgem, September 2003; Per Agrell and Peter Bogetoft, *Benchmarking for Regulation*, Final Report, prepared for the Norwegian Water Resources and Energy Directorate, July 2003; and Jeff D. Makhholm, *Benchmarking, Rate Cases and Regulatory Commitment*, prepared for the Australian Competition and Consumer Commission, November 15, 1999.

2. Caveats on how regulators can best apply performance measures in various areas of utility operations;
3. General categories of utility performance and alternative regulatory responses;
4. Different performance measures that regulators can use;
5. The uses and limitations of different performance measures and performance-measurement techniques;
6. The different regulatory venues for the application of performance measures, both within and outside a rate case; and
7. A general framework and sequence of steps that regulators can take to initiate performance measurement and evaluation tasks.

I. What Do We Mean by “Performance”?

A. Multi-dimensional nature of performance

“Performance” refers to the outcomes of one or more utility actions resulting from management decisions. These actions affect the various dimensions of a utility’s operations and services, including cost performance, reliability, and service quality, all of which affect consumer welfare. Performance is the “proof of the pudding,” determining how a utility’s actions affect its customers and the public.

This paper focuses on quantifying with objective information (e.g., actual numerical “performance” outcomes based on accounting data) how well a single utility or a group of utilities address these multiple dimensions. Performance measures rely on historical data or on estimates derived from economic models and statistical techniques. The latter metrics contain an element of error in measuring actual performance that regulators need to interpret carefully.

B. Performance standards

Regulators can consider performance from different perspectives. One perspective is efficiency. From an engineering perspective, efficiency takes on a strictly physical form. The ratio of person-hours of labor to kilowatt-hours of output is an example. This perspective disregards costs and assumes that a lower input-to-output ratio is desirable. This perspective, by itself, is limiting: A utility can increase its labor productivity by simply reducing its employees and substituting inputs such as capital or outsourcing; these alternatives, however, might be expensive enough to increase the utility’s overall costs.

From an economic standpoint, efficiency reflects management behavior in minimizing costs over the long term. Management, for example, can affect a utility’s cost performance by: (1) adjusting inputs to reflect the relative input prices, (2) exerting the optimal amounts of managerial effort to control costs, (3) constraining costly managerial expenditures (e.g., on expensive art and furniture) and other sources of waste (i.e., X-inefficiency), and (4) adopting new innovations and technologies when cost-beneficial.

Another way to consider performance is by means of comparison. If the regulator’s standard for power plant equivalent availability² is 80 percent and the utility performs at 70 percent, the efficiency ratio is 0.875 (70/80). Efficiency is a relative term whose measurement requires a benchmark or standard of performance. The standard might be the average performance of other utilities or the maximum efficiency that the regulator feels the utility under review can achieve.

The evaluation of utility performance often relates to “prudence.” One widely applied definition of prudence is decisions consistent with what a “reasonable person” would do, based

² Equivalent availability is a measure of power plant reliability.

on information available to the utility at the time of those decisions. The prudence standard focuses on actions, not outcomes.³ One criticism of the prudence standard is that a utility can satisfy it without performing at an above-average level. It establishes a threshold of minimum acceptable performance; it does not distinguish acceptable performance from exceptional performance. Grading and evaluation are done dichotomously: the utility's behavior is either acceptable or unacceptable; there are no intermediary levels of utility-management behavior.⁴

While performance evaluations often focus on cost, management also affects the non-cost aspect of utility performance. The effects of outages and service interruptions to customers depend on the response of utilities in restoring service and in isolating these incidents to selected areas to minimize the overall effect on customers. Utility performance also reflects the responsiveness of utility personnel to customer complaints and overall service quality.

³ For a detailed discussion of "prudence," including the many ways in which state and federal commissions and courts apply the term, see Hempling, *The Fundamentals of Electricity Law* (2006), available from NRRI.

⁴ According to this interpretation, a prudent decision resembles a utility receiving a "passing" grade when it performs between C to A; a C grade connotes mediocre utility performance for which the utility recovers all of its costs but it could have reduced its costs with more effort and competence; if a utility improves its grade from C to B, it exerts more effort but it might gain nothing in the long term under conventional ROR regulation; the incentive is akin to college students taking a course on a pass/fail basis.

II. Why Should Regulators Measure Utility Performance?

A. Performance problems under regulation

Regulation has an obligation to induce high-quality utility performance, whether it is customer service, physical operation of the utility system, service reliability, cost controls, or the adoption of new technologies. The economics literature shows that public utilities left unregulated, or regulated ineffectively, would perform suboptimally. They would set prices too high, price discriminate among customers, provide inferior-quality service, deploy a nonoptimal mix of inputs, and expend too little effort to control costs and innovate.⁵

Further, economic theory predicts that regulated utilities subject to rate of return regulation would perform at less than the highest possible allocative or productive efficiency.⁶ Traditional regulation tends to give utilities weak incentives to minimize their costs. To the extent a utility can pass on to customers additional costs and also pass on any cost savings it achieves, it has diluted any economic incentive to perform efficiently. Since rate-of-return regulation, by itself, will not produce the desired performance, some form of performance standards, including measurement, evaluation, and consequences, becomes more essential.

B. Regulators have an information disadvantage

In traditional regulation, the regulator is at a disadvantage relative to the utility in interpreting the utility's performance. Do the actual costs reflect competent utility management, or do they include wasteful costs that the utility could have avoided? The utility generally would defend these costs as reflecting their best effort under the circumstances. Some utilities would, therefore, be inclined to provide misleading information on their managerial efforts and cost opportunities. They may portray themselves as high-cost providers because of an unfavorable business environment. Under existing incentives, utilities may act rationally by exerting less-

⁵ See, for example, Harvey Averch and Leland L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *American Economic Review* 52 (December 1962): 1052-69; Harvey Leibenstein, "Allocative Efficiency vs. 'X-Efficiency,'" *American Economic Review* 56 (June 1966): 392-412; and Paul L. Joskow and Nancy L. Rose, "The Effects of Economic Regulation," in *Handbook of Industrial Organization, Volume II*, Richard Schmalensee and Robert D. Willig, eds., 1449-1506 (New York: Elsevier Science Publishers, Inc., 1989).

⁶ What analysts call the Averch-Johnson (A-J) effect says that a utility would use excessive capital input relative to other inputs such as labor, fuel, and materials. This outcome occurs when a utility faces a binding rate-of-return constraint on its rate base and its allowed rate of return exceeds its actual cost of capital. X-inefficiency occurs when the utility wastes resources by operating above its cost frontier. Unlike the A-J effect, this source of inefficiency would tend to reduce the utility's profits, at least in the short run because of regulatory lag. The underlying cause of both inefficiencies is the lack of strong incentives for a utility to minimize costs.

than-desirable managerial effort to reduce costs. After all, the opportunity cost for managers to spend more time and effort at their job is lost leisure time and more discomfort. The regulator might deem extant incentives as inadequate for motivating exceptional utility performance. Performance measures and their various applications by regulators can help lessen the information asymmetry that they inherently face in their oversight of utilities.

If regulators had good information about how utilities should perform, they could readily set performance standards that the utility would have to meet or suffer the consequences. In the real world, however, the regulator faces the problem of less-than-perfect information on the efforts of utility management and on the utility's cost opportunities. Cost-saving opportunities differ across utilities, depending on the inherent features of their production technology, exogenous input costs, and other factors that cause costs to vary by location because of their attributes. Utilities serving rural areas, for example, tend to have higher average costs than urban utilities.

The regulator observes outcome (e.g., power plant reliability) but does not have a utility's expertise in assessing how management produced that outcome. Since regulators lack the required information to identify optimal performance, they have to resort to alternative actions, such as special incentives or judgment of a utility's performance based on the information provided to them by the utility and other sources.

III. The Challenge Facing Regulators in Measuring and Evaluating Utility Performance

The appropriate use of performance measures requires careful interpretation of what they represent. Some measures reflect a utility's short-term performance, mostly factors beyond utility management control. Other measures estimate performance in some functional area that is subject to statistical error.

A. Factors affecting utility performance

Utility performance depends on three general factors:

1. The resources used,
2. Management skills, which determine what resources a utility should use and how it should combine them to produce some "output," and
3. Market and business conditions over which the utility has little control.

Utility performance derives from two distinct factors: *internal efficiencies and external conditions*. The first factor encompasses resources used, and the management skills that determine how to combine and deploy those resources. The second factor accounts for market and business conditions over which an individual utility has little or no control. Events over which a utility has no control, such as abnormal weather or economic conditions, however, should not exonerate the utility for how it responds to these incidents. If a storm causes a utility to interrupt service, it should reestablish service with the shortest possible delay consistent with general safety and the public welfare; nor do external events eliminate the utility's responsibility to anticipate and cost-effectively mitigate the effects of those events.

The appropriate uses of performance measures depend on their ability to separate out the effects of external and internal factors on performance. As an illustration, the cost of providing electricity is a function of the level of labor, fuel, and capital; their costs; consumer demographics; size of different customers and their electricity usage over different periods of time; and geographical characteristics of the utility's service territory. Two distinct management teams in charge of operating the same utility would likely produce different outcomes. The one team may better economize on the use of labor; for different reasons it might elicit higher productivity from the employees than the other management team. It might also operate its power plants more efficiently, and adapt more optimally to changes in input prices. Overall, even though both management teams face the same outside factors and have access to the same resources, one team is more proficient, at least in controlling costs. We can then conclude that one management team is superior to the other team, at least in terms of cost efficiency.

Appendix A illustrates the challenges to regulators in interpreting differences in one broad performance measure across utilities, namely, retail price. Analysts face difficulty in isolating the effect of management behavior on the differences, even when they apply the most sophisticated techniques.

B. Sports metaphors

One analogy involves two golfers who play on the same golf courses week after week. If one golfer has an average score of 70 strokes per round and the other golfer averages 73 strokes per round, we can conclude that the first golfer is better. If both of these golfers play on different courses, however, the golfer who averages 70 strokes per round may not be the better golfer if he plays on easy courses while the other golfer plays on more difficult courses (e.g., courses with volatile weather, high rough, fast greens, and longer holes). The golfer who shoots lower scores might average 75 strokes per round if he played on the courses of the other golfer. It is assumed here that the two golfers use the same or similar equipment (e.g., clubs, balls, shoes), so score differentials result from either differences in the golfers' skills or the difficulty of the golf courses on which they play, or both.

The same difficulty arises when trying to evaluate the managers of different baseball teams, each with players of dissimilar abilities. Can we say that the teams with the best records have the best managers, or should we have to consider whether those teams just have better players? How can we control for the differences in players' ability in evaluating the managers? Are there other factors that we would need to consider before ranking the managers? What criteria do we use to evaluate the managers? Is it controlling for other all factors, to the extent possible, and then measuring the separate effect of the managers on increasing their team's wins?

C. Regulatory considerations for applying performance measures

Performance measures quantify the effect of both management behavior and outside factors on "outcome." Utility management makes decisions on what actions to take—for example, build a new power plant, procure natural gas under long-term contracts, hedge fuel costs, or purchase gas. The outcomes and their effects on consumers and society as a whole, however, depend to some degree on factors over which the utility has little or no control.

When not applied properly by regulators, performance measures can lead to wrong decisions and perverse outcomes. Regulators should understand the limitations of performance measures to avoid these problems and to use those measures most constructively.

The following list identifies several elements of performance measures and the methods of measurement that regulators need to understand before applying them in different venues.

1. *The first decision is to select the functional areas for measuring utility performance.* Major criteria for selection are: (a) the effect of a functional area on a utility's total cost or on consumer value from reliable and high-quality utility service, (b) the ease of measurement, (c) the effort required to interpret a performance measure, and (d) the influence of utility management in affecting performance. The ultimate goal is to

maximize the net benefits from society's perspective, which involves comparing the benefits from improved regulation with the costs of measuring and evaluating performance. Examples of performance measures that meet at least some of these criteria are power plant equivalent availability, operation and maintenance (O&M) expenditures, and service reliability levels. All of the measures are quantifiable, are important in terms of affecting consumer well-being, and are subject to utility-management discretion.

2. *Improved performance in one area can reduce performance in another.* An increase in power plant performance can reduce a utility's total factor productivity (TFP)⁷ or increase its total costs. A reduction in maintenance and other costs, as a second example, may jeopardize the utility's service quality. These outcomes call for a utility-wide cost-benefit test. When focused on a single component of utility operations—in our example, power plant performance—regulatory actions can create perverse incentives: The utility would tend to devote excessive resources to the targeted area, in the process jeopardizing performance in other areas. An emphasis on cost reductions can cause service quality to suffer by reducing reliability and customer service. As another example, a focus on improving power-plant capacity factors or equivalent availability could cause a utility to overspend on O&M and pass these costs onto its customers. These additional costs, conceivably, could more than offset the benefits to customers from increased power-plant performance.⁸ As a last caution, in recent years regulators have become involved in addressing non-traditional objectives such as the promotion of energy efficiency, renewable energy resources, and affordable energy.⁹ In achieving these objectives, regulators might have to compromise on the traditional objective of providing reliable utility service at a reasonable price.

⁷ Total factor productivity measures a utility's total quantity divided by total inputs. It reflects the firm's efficiency in combining inputs (e.g., labor materials, fuel, and capital) to produce and deliver utility services (e.g., kilowatt-hours, peak demand). With positive productivity growth, the utility is increasing output by more than inputs, which translates into a decline of real cost per unit of output. Productivity growth means improved efficiency in the use of society's resources.

⁸ The implication for regulators is that they might want the utility to report not only on its power plant performance but also on related functions such as O&M. The regulator could then see whether the utility's O&M costs substantially increase concurrently with improved power-plant performance. The regulator could require the utility to report on the O&M costs of comparable power plants owned by other utilities.

⁹ See Ken Costello, *How To Determine the Effectiveness of Energy Assistance, and Why It's Important*, NRRI 09-17, December 2009, found at http://www.nrri.org/pubs/gas/NRRI_energy_assistance_dec09-17.pdf.

3. *Improved performance in one area can increase performance in another.* In a complementary relationship, better performance in one area can directly lead to improved performance in one or more other areas. As an example, an increase in power-plant equivalent availability can reduce a utility's fuel costs. A complementary relationship between two or more areas of utility operations heightens the importance of performance improvement in those areas.
4. *The previous two items indicate an interrelationship between different performance areas of a utility that regulators should take into account.* For regulators, this association means that the cost-benefit effect of performance improvement in a single area has a spillover effect on other areas that requires consideration. When the association is negative, a seemingly attractive action to reduce purchased gas expenses, for example, might result in additional costs from hiring consultants and more in-house labor. The net effect might be to increase the utility's overall costs, although purchased gas costs would decline as intended. The implication for regulators is that to focus on improved performance in a single area can produce a counterproductive outcome in the form of higher rates to consumers without any corresponding increase in the value of service.
5. *Performance depends upon different factors, as mentioned above, some under a utility's control, others exogenous to a utility.* The challenge for regulators is to separate the effects of management from the effects of factors beyond a utility's control. Without separation, the proper applications of performance measures become greatly restricted. Specifically, it is unreasonable for regulators to then apply performance measures mechanically or as the sole source of information for evaluating a utility's performance.
6. *Performance measures are either estimates or actual accounting numbers.* Total factor productivity is an estimate of a utility's overall performance in using labor, capital, materials, and other inputs to produce and deliver a service. It is an estimate because it assumes certain production behavior by the utility and requires data that represent estimates rather than actual unadjusted accounting numbers (e.g., capital services). These performance measures require the use of statistical and econometric techniques that make certain, and sometimes restrictive, assumptions. Other performance measures derive directly from reported data □ for example, labor productivity, unit cost for customer service, and total operation and maintenance expense per customer.
7. *Varying degrees of difficulty exist in measuring performance.* The more sophisticated approaches, while in theory better suited for broader applications, are susceptible to measurement and data errors. These approaches include econometric and total factor productivity techniques. They require regulatory staff to have a good understanding of statistical techniques and other quantitative methods. If staff members don't have this understanding, the regulator would then have to rely on outside consultants, which can cost a non-minimal amount of money.

8. *Regulators can use either ex post or ex ante measures of performance, or both in a particular application.* Regulators can apply the former measure for prudence reviews or to compare a utility's actual performance with the expected outcome. In these applications, regulators can actually use both kinds of performance measures, with the *ex ante* measure acting as a prospective standard for benchmarking a utility's performance. Assume that the regulator sets a customer service standard for a utility. After observing the utility's actual performance, the regulator can compare this performance with the standard to help judge whether the utility acted prudently.
9. *Trade-offs can exist between short-term and long-term performance.* Additional capital expenditures have the effect of temporarily reducing a utility's total factor productivity while increasing long-term productivity. Tree trimming is a good example in which spending more today would likely lead to lower costs in the future because of fewer outages and lower maintenance costs. This kind of investment over time benefits both the utility and its customers. Higher O&M costs in general incurred today can lead to better utility performance in the long run.
10. *Benchmarking can use as a reference, "average," "exceptional," or "standard" performance.* In evaluating or measuring a utility's performance, the analyst often needs to specify a "reference" or "baseline" performance. Average performance can represent the "mean" performance for a sample of comparable utilities. Some regulators might interpret average performance as the costs incurred by an efficient utility. To other regulators, average performance might reflect subpar performance if they deem the "mean" utilities to be performing poorly, say, because of weak regulatory incentives. Exceptional performance might include the performance of the first quartile of utilities or, more stringently, those utilities lying on or close to the efficiency frontier measured by statistical or non-statistical approaches. Regulators can designate "standard performance" as a target for a utility to achieve or surpass. The standard itself can reflect the average performance of a sample of utilities or the performance of the top comparable utilities.

Regulators should consider whether they should view "standard" performance as a moving target, rather than as a static concept that remains constant over time. As technology improves and the utility adopts better management practices, regulators would expect the utility to improve its performance over time. Regulators might also press utilities to move in the direction of "frontier" performance in which they would adopt "best practice" technologies and management practices.

D. How might regulators interpret and use the results?

Regulators can interpret a utility's performance differently. Their interpretation affects what action they take with regard to cost recovery, prudence reviews, and a follow-up investigation. The different interpretations include:

1. The utility is performing prudently;

2. The utility is performing prudently but its performance can improve;
3. The utility is performing worse than peer utilities;
4. The utility is performing better than peer utilities; and
5. The utility is performing unsatisfactorily.

Each interpretation has different implications for regulatory action. The regulator would first need to have information before it can interpret utility performance. A performance metric would seem essential: The regulator would need to compute the utility's historical performance, the performance of a group of utilities, or a predetermined performance standard based on cost, engineering, and other information. In comparing performance across utilities, the regulator would have to select a peer group whose characteristics are similar to the utility under review. As an alternative, the utility could select a wider group of utilities and control for differences in characteristics through statistical techniques and other quantitative methods.

For each of the above five interpretations of utility performance, a different regulatory response would seem appropriate. The *first interpretation* can result in no incremental regulatory action. The regulator might perceive utility performance as satisfactory in reflecting prudent utility behavior; that is, the utility's performance coincides with acceptable management behavior.

In the *second interpretation*, the regulator perceives utility performance as acceptable but believes that it can improve. "Prudence" here refers to utility management behavior that meets some minimum threshold but is not necessarily "above average." The regulator might want to establish, for example, special incentives that would elicit "above average" performance or set a target that the utility would have to achieve by a specified future date. The regulator should first decide whether better performance for a specific area of operation is warranted (e.g., cost-beneficial) from the perspective of consumers and the general public. An improvement in system reliability, for example, can produce smaller benefits to consumers than the additional costs they will have to pay.

The *third interpretation* can result in a penalty for the utility or further regulatory action that would attempt to identify why the utility under review is performing below its peers. A comparison of a utility's performance with other utilities involves "benchmarking." Benchmarking means setting a standard that is a point of comparison or reference for performance appraisal. If, for example, the benchmark cost per customer is \$X and a utility has a cost per customer of \$1.2X, the utility is performing below the average level of its peers. The analyst can conduct a statistical test to determine whether the utility's cost is significantly different than the mean cost for the peer group. The test would calculate a confidence interval that would indicate the accuracy of benchmarking and allow for hypothesis testing of cost performance. Use of this information depends on what regulators judge it to represent. If the numbers adjust for those cost factors beyond a utility's control, then regulators might conclude

that any residual is attributable to utility-management behavior. In this instance, the regulator might be more inclined to penalize the utility or investigate further why the utility's performance falls below its peers.

The *fourth interpretation*, in which the utility is performing above its peers, can result in the regulator rewarding the utility for its performance. It can give the utility a higher allowed rate of return or at least signal to the utility that it won't be penalized for its performance. Analogous to the third interpretation, before rewarding the utility the regulator should further investigate to judge whether the utility's above-average performance is the product of exceptional management behavior or simply favorable conditions.

The *fifth interpretation* of performance can cause the regulator to penalize the utility or take some other response that intends to improve the utility's performance in the future. The regulator might require a management audit of the utility or set future targets for the utility to meet or else face penalties. In taking any action that directly affects a utility's financial condition, the regulator should have good evidence that the utility's poor performance reflects bad or imprudent management behavior. In other words, the regulator should clearly understand why the utility's performance is subpar before taking any action that affects the utility's financial condition.

Good regulatory decisions require a combination of quantifiable information and judgment. Performance metrics in conjunction with other information can empower regulators to take consequential actions. The action might involve cost and other adjustments in a rate case, a detailed investigation of the utility triggered by preliminary evidence of subpar utility performance, or penalties or rewards for exceptional performance.

IV. An Overview of Different Techniques for Measuring Performance

A. Attributes of good performance measures

Performance measures should be objective, quantifiable, and verifiable. One interpretation of these qualities is that good measures represent metrics with numerical values based on public data and sound analytical techniques that anyone can replicate. Benchmarking—that is, a comparison of a utility’s performance with some reference such as its past performance or the average performance of similar utilities—requires quantitative performance measures; otherwise, regulators would find it difficult to determine whether a utility has performed satisfactorily. Some measures are estimates derived from advanced mathematical and statistical techniques. Replication and proper interpretation of these measures requires a high level of skills. Other measures derived from actual accounting numbers are easier to calculate and replicate.

When establishing benchmarks, regulators should use performance measures that, as much as possible, reflect utility management behavior. One benchmark for regulators to consider is the performance of an “average utility.” If the regulator established a tighter or looser standard, a utility could face unfair penalties or enjoy windfall gains¹⁰ because of exogenous factors. Assume, for example, that the benchmark represents the performance of the most efficient utility and the regulator penalizes the utility for performing below this level. A utility can argue correctly that this outcome is incompatible with competitive markets where firms receive low returns when they perform below average, not if they perform less well than the highest performing firm; in competitive markets, firms receive above-normal returns when they perform above average.¹¹ When performance measures do not separate management behavior from other factors, a utility, on the other hand, could profit or assume a top ranking even if only because of the favorable environment under which it operates.

¹⁰ A “windfall gain” means that the utility’s profits increase without any benefits to customers.

¹¹ Some regulatory experts have argued that the primary objective of regulation should be to replicate the outcome of effective competition in achieving marginal-cost pricing and minimum cost of production. If regulators were to follow the second objective, they would not distinguish between outcomes that were beyond the control of the utility and outcomes largely influenced by management behavior. A competitive firm, for example, could have a good outcome even with bad management judgment if it has good fortune (e.g., good weather for a farmer). Conversely, it could have a bad outcome even if it performed superbly.

B. Econometric methods, indexing, and data envelopment analysis

A March 2010 NRRI paper identified various approaches for measuring utility performance.¹² The approaches include econometric methods, indexing, and data envelopment analysis (DEA).¹³ They differ in data requirements, ease of measurement, interpretation, and other ways. Their uses by different regulatory bodies vary. U.S. regulators have more experiences with the econometric and indexing approaches than with DEA.

In this country, the application of econometric methods for performance measurement has mostly involved the estimation of statistical cost functions for operation and maintenance. Performance for an individual utility relates to the difference between actual costs and predicted costs.¹⁴ This method defines standard performance or the benchmark as the average performance of utilities in the sample.¹⁵ In contrast, frontier cost functions define the standard as the best performing utility. The difference between the two definitions of a benchmark for setting rates can have large financial consequences for a utility trying to recover its costs.

A number of utilities have applied the statistical cost approach, most often to demonstrate to their regulators that they have performed above average in the operational area under review. As far as the author knows, no state public utility commission has taken the initiative in applying econometric methods or DEA to monitor and evaluate the performance of energy utilities.¹⁶

¹² See Evgenia Shumilkina, *Utility Performance: How Can Commissions Evaluate It Using Indexing, Econometrics, and Data Envelopment Analysis?* NRRI 10-05, March 2010, at http://nrri.org/pubs/multiutility/NRRI_utility_performance_mar10-05.pdf.

¹³ DEA is a method in which linear programming or other operations research methods calculate the efficient input-output relationships for individual utilities. A major shortcoming of this method, as well as other non-statistical ones, is that they are unable to separate the inefficiency effect from statistical noise or randomness because of poor quality data and data errors, omitted variables, and other problems. DEA defines the benchmark as the best performing utilities.

¹⁴ See, for example, Pacific Economics Group, *The Cost Performance of Boston Gas*, January 28, 2003; and Pacific Economics Group, *Benchmarking the Operating Performance of Portland General Electric*, February 10, 2010.

¹⁵ “Average performance” occurs when the predicted cost and the actual cost are equal. See the studies cited in footnote 14.

¹⁶ The Ontario Energy Board uses the econometric method to assist in evaluating utility performance. See, for example, Pacific Economics Group, *Benchmarking the Costs of Ontario Power Distributors*, March 20, 2008.

C. Additional ways to measure utility performance

1. Management audits

A management audit is a systematic assessment of the tools, processes, and policies of utility management in resource usage, planning, and organizational activities. Management audits can: (1) assess the current effectiveness of management, (2) recommend improvements, and (3) establish “best practices” standards for future use. U.S. regulators often use management audits to evaluate a utility’s performance. Overall, management audits can help both regulators and utilities understand current processes, evaluate those processes relative to “best practices,” and recommend changes.

The major positive feature of management audits is their scrutiny of utility processes and the detailed information they provide to regulators. Management audits can investigate specific utility operational areas or the utility as a whole. On the negative side, management audits are expensive and rarely provide a quantitative benchmark for evaluating a utility’s “output” performance. The most useful audits recommend improvements in management practices for a single component of a utility’s operation, such as work-force management or maintenance of power plants. Because they are expensive, management audits are most appropriate when there is evidence of a specific problem. That evidence can derive from narrow-based performance measures relating to specific functional areas.

2. Accounting ratios for individual functional areas

Examples of accounting ratios are labor expense per dollar of revenue, administrative and general expense per customer, and operation and maintenance expense per customer. These ratios are easy to calculate: They require no sophisticated estimation technique such as econometrics and linear programming.

Regulators must use caution, however, in applying these measures for benchmarking and evaluating a utility’s performance. Since ratios do not control or account for factors beyond a utility’s control, they reflect more than utility management behavior; when not used appropriately, ratios can lead to counterproductive outcomes. Appendix B illustrates accounting ratios adjusted for inflation, labeling them “real unit cost indices.”

Simple accounting ratios can assist regulators in “red flagging” operational concerns. They can also help regulators (a) identify historical trends—for example, the growth of labor costs per dollar of revenue over the past ten years; (b) determine today’s baseline performance—for example, the mean performance of a group of utilities; and (c) quantify relative performance across utilities—for example, the labor costs per dollar of revenue of a utility compared with the mean for other utilities in the same state. Regulators should refrain from using these ratios by themselves to adjust a utility’s rates, to determine cost recovery, or to make other decisions that directly affect the utility’s financial condition. It would be unfair to the utility or its customers: penalizing a utility for subpar performance or rewarding it for exceptional performance, both

explained by exogenous factors, would produce a zero-sum outcome; the regulators could deprive the utility of recovering prudent cost or the utility would enjoy a windfall gain with no apparent “performance” benefits to customers.

Accounting ratios could be useful in placing on the utility the “burden of going forward” to explain performance problems. Accounting ratios are a low-cost regulatory tool that has definite limitations but, when applied correctly, can improve the ability of regulators to evaluate a utility’s performance.

The usefulness of these ratios depends on the selection of the peer group whose average performance represents the benchmark for evaluating the performance of a single utility. No perfect benchmark exists, because no peer group operates in an environment identical to the utility under review. The selection of similar utilities can result in more meaningful benchmarking. Differences in performance between utilities would then reflect more management behavior than exogenous factors.

3. Uses and limitations of performance measurement

Table 1 shows different performance measures and measurement techniques. In addition to their uses, it lists the limitations that regulators should keep in mind when applying them in specific situations. This paper previously discussed these uses and limitations. Part V.B discusses how regulators can apply performance measures in different venues.

Table 1. The Uses and Limitation of Different Performance Measures and Measurement Techniques

Performance measurement	Use	Limitation
Statistical method	<ul style="list-style-type: none"> • Estimation of average performance as the predicted cost controlling for a utility’s exogenous conditions • Ranking of the performances of different utilities based on the deviation between a utility’s actual performance and average performance • Estimation of the effect of individual factors on cost • Application of statistical tests for performance evaluation 	<ul style="list-style-type: none"> • Predictions of average performance sensitive to different assumptions, model design, the data, and econometric errors • Requirement of substantial data • Demand for skills in sophisticated econometric and statistical techniques • Inclusion of only quantifiable factors
Accounting cost and non-cost ratios	<ul style="list-style-type: none"> • Provision of information that “red flags” or identifies potential problem areas at low cost • Provision of preliminary information for in-depth inquiry • Comparison of a utility’s performance over time or with other utilities 	<ul style="list-style-type: none"> • No separation of management effects and other factors on performance • Narrow-based measures that don’t account for interdependencies between utility functions • No definite benchmark
Management audits	<ul style="list-style-type: none"> • Evaluation of current processes, policies, and management practices for specific functional areas • Recommendations on improvements or prudence of past actions • Establishment of “process” standards for future performance 	<ul style="list-style-type: none"> • Expensive to conduct • No “outcome” metric or benchmark
Total factor productivity	<ul style="list-style-type: none"> • Quantification of the overall cost performance of a utility • Quantification of the effects of individual factors on performance • Comparison of a utility’s performance over time or with other utilities 	<ul style="list-style-type: none"> • Estimation of some required data • No separation of management effects and other factors on performance • No definite benchmark
Price	<ul style="list-style-type: none"> • Comparison of a utility’s average cost with other utilities 	<ul style="list-style-type: none"> • No separation of management effects and other factors on performance • No explicit benchmark

V. Applications of Performance Measures in Different Regulatory Venues

Performance measures offer regulators a tool that is useful for different purposes in different venues. This section will first identify three broad ways in which regulators can use performance measures. It will then discuss seven specific applications of performance measures.

Regulators first should recognize the shortcomings of the performance measures for benchmarking purposes. They need to exercise caution in interpreting and using the measures. It is not uncommon for rankings of utility performance to vary depending on the measurement and benchmarking methods used. A good approach is to use different benchmarking methods to compare and evaluate the results, rather than rely on a single method.

A. General uses of performance measures and examples

Regulators can judge a utility's actions in three general ways:¹⁷

1. Evaluate the information used by a utility prior to an action.
2. Observe and evaluate the utility's actual performance.
3. Retrospectively review the prudence of the utility in undertaking the action.

Regulators can use performance measures in each of these three ways. The first way requires evaluation prior to an action, while the second and third evaluate utility performance after the fact. One example is the regulator periodically reviewing a utility's construction performance in controlling cost and reaching scheduled milestones. Another example is a regulatory review of a utility's prospective and retrospective actions with regard to customer service.

1. Illustration of service quality

The regulator might want to assess in advance whether a utility's proposal to improve its service quality is cost-beneficial. It might judge, after the fact, whether the utility's actual service quality is satisfactory or requires additional review to determine whether the utility complied with the regulator's standard. The regulator might establish service quality targets to compare periodically with the utility's actual performance. The regulator might resort to an incentive mechanism that would reward a utility for surpassing a target and penalize it for performing below the target. Another option is for the regulator to penalize a utility for failing to meet pre-specified standards, but not reward it for superior performance. This option is premised

¹⁷ See, for example, William E. Encinosa, III and David E. M. Sappington, "Toward a Benchmark for Optimal Prudency Policy," *Journal of Regulatory Economics* 7 (1995): 111-130.

on the belief that a utility should not earn a reward for fulfilling a primary obligation, such as providing high-quality service.¹⁸

2. Illustration of energy-efficiency activities

In evaluating a utility's proposed action, the regulator can review other utilities' actions, in addition to the outcome of those actions, to compare with what the utility under review is proposing. If the utility, for example, proposes to invest in energy efficiency, the regulator can compare its estimated costs with the actual costs incurred by other utilities for comparable investments. The regulator can also compare the utility's estimated benefits with the actual benefits for similar initiatives undertaken by other utilities. These comparisons can help the regulator gain access to information that is presumably more reliable and objective than the information it receives from the utility under review. They can, consequently, enhance the regulator's ability to make an informed decision.

After the utility undertakes an action, the outcomes become measurable. Once the utility implements its energy-efficiency initiatives, the regulator or some other party can measure the actual benefits. The regulator can use the measurement to compare with the utility's estimates to judge whether individual initiatives should continue, expand, or terminate. Measured performance by itself does not imply prudence or management competence; it can, however, "red flag" a potential problem that needs correction or indicate that the utility's performance is exceptionally bad, warranting further investigation.

3. Prudence review

Performance measures by themselves cannot determine whether a utility acted prudently. If regulators use them in this capacity, the utility becomes highly susceptible to a whimsical evaluation based on outcomes rather than the prudence of the decisions themselves. A regulator who penalizes a utility for hedging its natural gas purchases when the spot market price turns out to be lower than the hedged price is an example. Could the utility not have hedged, and would it have resulted in lower cost? Yes, no question—the utility had the option to purchase all of its gas at the spot price and would have benefited from doing so. But was the utility imprudent in deciding to hedge? We don't know unless we do a detailed inquiry as to: (a) what the utility knew at the time it made the decision, and (b) how it used that information to conclude that hedging was a reasonable alternative. The ratio of the hedged price to the actual price over several years—a form of performance indicator—could suggest a problem requiring review.

4. Evaluation of a regulatory action

Another possible application of performance measures is to determine whether a particular regulatory action or change in policy produced the intended improvement. After establishing a new incentive mechanism for gas procurement, for example, the regulator should

¹⁸ For an excellent review of different regulatory options, see Pacific Economics Group, *Service Quality Regulation for Detroit Edison: A Critical Assessment*, March 2007.

want to know whether the mechanism improved the efficiency of a utility to purchase natural gas. A major challenge for the analyst is to attribute any improved performance to the incentive mechanism, *per se*, rather than to other factors: What would the utility's gas costs have been in the absence of the incentive mechanism?

Overall, performance measures can play an important, even if only a subordinate, role in the three general ways for regulators to evaluate a utility's performance. By themselves, the measures lack the capability to assess management performance. Performance measures, however, can supplement other information to assist regulators in assuring customers that utilities do not flow through excessive costs to their customers and underperform in other ways.

B. Specific applications

1. Regulatory incentive mechanisms

The core component of an incentive mechanism is the benchmark, which determines the specific costs and revenues applicable to the mechanism, the strength and nature of incentives, the relative likelihood of award or penalty, and the utility's exposure to risk as a result of the incentive mechanism. Appendix C describes one kind of incentive mechanism that highlights the importance of a benchmark in distributing the economic benefits between the utility's shareholders and consumers.

The rationale for an incentive mechanism is that it would motivate the utility to perform at a higher level than that at which the utility performed previously. It has this effect by decoupling revenues from a utility's actual costs when its performance falls in the "exceptional" category. Under one form of incentive mechanism, the utility earns no reward or receives no penalty if actual costs equal (or are within a tolerance band around) the benchmark, and the utility receives an incentive award if it beats the benchmark. In principle, then, the benchmark should measure performance that results from reasonable management behavior reflecting acceptable, but not superior, performance deserving of no award or penalty.¹⁹ The benchmark could represent average or non-exceptional performance. As illustrated in Appendix C, the wrong benchmark can have counterproductive results: They can cause higher rates for customers and a windfall gain to the utility. Incentive mechanisms require performance measures to calculate the magnitude of utility rewards or penalties (e.g., a prespecified percentage of the difference between actual performance and the benchmark).

Performance measures applied to past utility actions can help regulators determine whether an incentive mechanism had actually improved performance. Such a determination, however, is extremely difficult to conduct. The regulator would need to determine how the utility would have performed in the absence of the incentive mechanism. If the utility's performance substantially or even marginally improved with the mechanism, the regulator might

¹⁹ See, for example, Ken Costello and James F. Wilson, *A Hard Look at Incentive Mechanisms for Natural Gas Procurement*, NRRI Report 06-15, November 2006, at <http://www.nrri.org/pubs/gas/06-15.pdf>.

infer that the mechanism had a positive effect. But how much the mechanism improved performance depends on the collective effects of other factors that might have changed.

2. Periodic monitoring of utility performance outside a rate case

a. Performance for individual functional areas

Monitoring has four major purposes: (1) report and evaluate utility performance in one or more functional areas, (2) propose changes to regulatory policies and practices to improve utility performance, (3) determine utility compliance with rules, guidelines, and expectations, and (4) apply any mitigating actions when necessary. Performance measures offer regulators a tool in conjunction with other information to carry out monitoring activities. Regulators might want to quantify the performance of a utility in specific areas on an annual basis. If the measures suggest a potential problem, regulators might further investigate with more detailed information and analysis.

Periodic reviews can increase the regulator's understanding of a utility system, and its components, in addition to its actual performance. This understanding can assist regulators in determining whether to adjust rates or take other actions based on evidence of exceptional performance. Performance measures can help direct regulatory resources to those areas of utility operations that are most in need of improvement.

Regulators may establish performance targets to evaluate a utility's actual performance, at least in terms of deciding whether to pursue further inquiry. Monitoring of a utility's performance can lead to: (1) regulatory actions aimed at avoiding recurrence of past problems or (2) determining whether a utility has complied with a regulatory standard or obligation for a functional area of operation. Did the utility continue to have bad customer service that needed improvement? Did the utility meet the requirements established by the regulator for reliable service?

b. Utility-wide performance

Econometrics, data envelopment analysis, and total factor productivity are distinct approaches for measuring the overall cost performance of a utility. Although measuring a utility's cost performance in specific areas is important, it neglects the more substantial question of how these "component" performances add up to the utility's overall cost performance. After all, it is the utility's total cost that determines the rates it charges to different customers.

Appendix D shows how an improvement in total factor productivity reduces a utility's average costs and rates. By comparing a utility's past growth rate of total factor productivity with a peer group, the regulator is able to measure the effect of any differential on the utility's total cost. Assume that two utilities have different historical growth rates of TFP. This outcome should cause the utility with the higher growth rate to have a lower percentage change in cost over time, assuming other things remaining the same. An increase in TFP is equivalent to a decline in the real dollar cost of the aggregate input per unit of output. (See Appendix D for the mathematical relationship between TFP and average costs.) The regulator might want to know

the additional dollars expended by the utility with the lower historical TFP growth. TFP depends on several factors, including technical change, economies of scale, and the ability of utility management to combine inputs to maximize output (i.e., productive efficiency).²⁰ A comparison of TFP growth rates across utilities, therefore, reflects a mixture of internal efficiencies and external market conditions.

3. Comparison of a utility's actual performance with a benchmark, both in rate cases and other regulatory forums

The measurement of performance is the first step toward a preliminary evaluation of a utility's performance. The next step is to develop a standard, which can include selecting peer utilities and measuring their average performance. Regulators can then compare this average performance with the performance under review. A statistically significant difference can attract the regulator's attention and lead to further action.

Analysts have assigned different functions for benchmarking. They include:

1. Identify "best practices" in management processes and tools,
2. Monitor relative performance across utilities,
3. Identify areas of a utility's operations that require needed attention or further investigation,
4. Establish targets or standards for utility performance,
5. Mitigate the cost-plus nature of regulation, and
6. Place the focus on outcomes instead of inputs.

As one application, regulators can then use the "benchmarking" results, along with other information, to determine whether a utility should develop a plan to improve performance in a function the regulator deemed problematic. Regulators can apply this tool both within a rate case and in other regulatory venues.

²⁰ A major factor for the short-term movement of total factor productivity is output fluctuations as they affect a firm's capacity utilization.

4. Evaluation of the reasonableness of “cost-of-service” components, adjustment of the rate of return on equity (ROE), and use of total factor productivity

a. Rate-of-return regulation

Performance measures can assist regulators in a rate case. Regulators can adjust a utility’s allowed rate of return on equity for past performances. They might reward a utility by adjusting upward the utility’s rate of return by 50 basis points for surpassing performance targets established by the regulator. On the other side, a utility might receive a lower allowed ROE for poor customer service or other subpar performance.

A major task of regulators in rate cases is to determine whether a utility overstated its revenue requirements to justify a higher rate increase.²¹ Assume that a regulator uses a future test year to determine new rates. The two broad factors affecting differences between historical and future costs are forecasted changes in productivity and input prices. The utility’s forecasts of costs are, therefore, dependent on expectations of its future productivity growth and input-prices escalation. By understating the productivity gains, other things constant, the tendency is for the utility to overstate its future revenue requirements and, therefore, the rate adjustment required for earning a fair rate of return.²² Regulators should ask: Do the cost forecasts incorporate a change in productivity that reflects good utility-management behavior and is comparable to the utility’s historical performance? Regulators can use performance measures to determine whether future test-year costs for specific functional areas reflect an appropriate baseline for setting new rates. The regulator can observe historical values over a number of years to judge whether the change in specific costs from current values to forecasted values is consistent with historical changes. Changes in performance can supplement other information to determine reasonable costs for setting new rates.

Appendix B illustrates a cost measure for individual utility functions adjusted for inflation. Regulators can discern whether the implicit productivity change for a specific function as projected by the utility is in line with historical changes. Assume, for example, that the utility is projecting advertising costs per dollar revenue (in constant dollars) to increase by 10 percent per annum over the next two years. If historically over the past five years, the per-annum increase was only 2 percent, the regulator might rightly conclude that the utility is inflating advertising costs in its future test-year filing, unless the utility provides a good explanation for the higher growth rate in the future.

²¹ If a regulator approves test-year costs that are excessive, other things being the same, the utility’s actual rate of return would exceed its cost of capital and rates would be too high.

²² A larger share of the cost savings from actual productivity gains over the effective period of new rates would then go to the utility. In adjusting rates, regulators implicitly determine the distribution of productivity benefits between the utility and its customers. Integral to good regulation, any rate adjustment should reflect a level of productivity, as well as input prices, that are compatible with “reasonable” performance by utility management.

b. Price-and-revenue cap regulation

In a number of foreign countries, regulators have used performance measures as a benchmark to set the parameters for a price-cap mechanism. They apply statistical benchmarking to help determine the base price²³ and the “stretch factor” component of the X-factor for an individual utility based on changes in the TFP for peer utilities.²⁴ One interpretation of the base rate is that it represents the cost of an efficient utility, rather than strictly the cost of the utility under review. Many of the price-cap plans include benchmarks for service quality. One concern is that under price caps a utility would be strongly motivated to control costs, even to the point of compromising service quality. A separate component using historical service-quality levels as a benchmark would penalize a utility for falling below those levels (or, below the lower bound of a pre-specified “band”).²⁵

c. Riders

Regulators can tie rate adjustments outside a formal rate case to a utility’s performance. These adjustments occur within the confines of cost or formula-rate riders.²⁶ Annual adjustments of base rates can depend, for example, on the utility’s performance in customer service relative to some predetermined standard. Performance is, therefore, a factor in the context of both what the regulator expects from the utility and what outcome the utility achieves.

5. Preliminary review of a utility’s performance to determine further action

Performance measures can act as indicators of potential problem areas. They can help regulators assess the benefits expected from a management audit or other thorough investigation. This use of performance measures involves detecting areas of a utility’s operation for which its

²³ In traditional regulation, the base rate would correspond to the actual costs of the utility under review. Under benchmarking the base rate would account for the efficiencies and costs of peer utilities. The reason is that setting the base rate on the basis of information only for the utility under review would invite gaming and perverse incentives,

²⁴ Under a price-cap mechanism, the maximum price that a utility can charge for period t equals the base price plus the accumulated changes since the base period, determined by the change in the selected price index (e.g., GDP Implicit Price Deflator) minus the X-factor, which commonly relates to a measure of total factor productivity. The “stretch factor” attempts to adjust the X-factor for differences in past TFP changes between utilities.

²⁵ Another benchmark can include the average service-quality level of a group of similar utilities.

²⁶ Under one definition of formula rates, rates are adjusted annually to reflect changes in a utility’s costs and revenues relative to test-year levels. The goal is to assure that the utility’s actual rate of return on equity (ROE) does not deviate far from what the regulator approved in the last rate case.

current performance compares unfavorably with other utilities, with the historical performance of the utility itself, or with the regulator's predetermined desired outcome. A utility with certain costs that are "outliers" should undergo more detailed review to determine the reasons for its exceptional performance.

One such detailed review that can uncover potential problem areas is a management audit, which this paper discussed in Part IV.C.1. An audit can evaluate past performance to determine cost recovery; or evaluate current management practices to recommend changes in these practices, such as work-force management and power-plant maintenance. These changes have the purpose of improving the utility's future performance.

Narrow-based performance measures can provide the initial information to justify a management audit. Management audits can help regulators better understand current utility processes and practices. They can lead to changed utility actions that are more in line with "best practices."

6. Examination of the reasons for performance differences across utilities

A statistical analysis can identify factors explaining why some utilities perform better in certain operational areas than other utilities. Why, for example, do some utilities have lower O&M expenses or higher equivalent availability for coal power plants? Regulators should want to know why some utilities under their jurisdiction are performing worse than other utilities. Effective regulation would include inquiries into these questions—how else can regulators know that the utilities under their jurisdiction are charging "just and reasonable" rates that reflect prudent and efficient utility management? This use of performance measures requires more than just calculating performance directly from accounting or other reported data; it also requires statistical analysis that measures the effects of individual factors on a utility's performance.

7. Publicity of a utility's performance on a periodic basis

The use of publicity to induce utility performance is uncommon in the U.S., but regulators in other countries have more frequently used these tactics, especially in instances in which a utility's performance was poor.²⁷ In Massachusetts, some utilities send in their customers' bills an annual report card on their performance. The regulator checks the accuracy of the report before the utility releases it to the public. The information includes a comparison of the utility's commitment to a targeted performance level with its actual performance. Performance areas for one utility, Western Massachusetts Electric Company, include: (a) the utility's response to customer calls, (b) average outages per customer, (c) the average number of minutes without power per customer, and (d) customer complaints per thousand customers.²⁸

²⁷ See, for example, Sanford V. Berg, *Survey of Benchmarking Methodologies*, prepared for the World Bank, March 1, 2006.

²⁸ I thank Joseph Rogers for this information.

VI. A Six-Step Approach for a “Performance” Initiative

Regulators can undertake six sequential tasks for developing a “performance” initiative. These are:

A. Identify the uses of performances measures

What purposes would they serve in improving utility performance? Regulators need to know how they can best apply performance measures and not use them inappropriately. Part V.B discusses seven possible applications of performance measures.

B. Select utility functional areas for regulatory review

Part III.C lists four criteria for selection: (a) the effect of a functional area on a utility’s total cost or consumer value from reliable and high-quality utility service, (b) the ease of measurement, (c) the effort needed to correlate performance measures with management behavior, and (d) the influence of utility management in affecting performance. It makes sense to select a functional area that has a substantial effect on a utility’s costs or other dimensions of performance over which the utility has discretion.

C. Calculate the performance measures

Performance measures can derive directly from accounting or other statistics periodically compiled and reported by utilities; or utilities or regulatory staff can estimate performance measures using sophisticated analytical techniques. These techniques have the ability to separate the effects of management behavior from other factors in determining overall utility performance. Their applications require proficiency in statistics and other numerical methods.

D. Compare a utility’s performance with a predetermined benchmark

The benchmark can be the performances of other utilities, the regulator's own standard, or the utility’s own historical performance. A comparison can help determine whether a utility’s performance is exceptionally good or bad or falls outside the range of “standard” performance.

E. Assess a utility’s performance

The regulator can perform an internal review to further examine the performance statistics to identify possible explanations for exceptionally good or bad performance. The regulator might also want the utility to respond to performance metrics showing its performance to lag behind the performances of other utilities.

F. Take action

An action might include allowing a utility to recover costs for a particular function, conducting a more detailed review of the utility’s behavior, or establishing stronger regulatory incentives for improved utility performance. With supplemental analysis, regulators can apply performance measures to disallow costs as well as penalize a utility in other ways. Symmetrical

regulation would also reward a utility if its performance is exceptionally good—for example, if its performance exceeds the standard predetermined by the regulator because of outstanding management behavior.

Appendix A: Price Differentials across Utilities: The Challenge of Detecting Causes

Theoretical problem

Assume that regulators want to compare the prices charges by different utilities. They can use this information in various ways. First, they can see how the prices of utilities in their state rank with those in other states. Second, they might conduct a statistical analysis to identify reasons for price differences. They might, for example, want to know whether demand conditions and other factors beyond the control of a utility explain most of the differences. This analysis would require specifying and estimating a conceptual model such as:

$$P_{ci} = f(D_{ci}, C_{ci}, R \dots Z),$$

where the price charged by utility *i* to customer class *c* relates to demand conditions (D_{ci}), costs (C_{ci}), regulatory practices (R), and other factors (Z). By estimating the relationships between price and the individual factors, regulators can assess the effect of each factor on price. They can then use this information to better understand why prices vary across utilities. Regulators can then interpret price differences that are unexplained by these factors as a residual. The residual can reflect model error in predicting price or variations in management competence, or a combination of both.

Some regulators might attempt to use price as a benchmark to penalize or reward a utility. Price is easy to measure and it compasses all of a utility's costs, avoiding the distortive incentives that could arise from using a partial measure of performance. But the problems associated with a "price" benchmark are potentially serious. Utilities might have different prices at a point in time because of the uneven treatment of certain costs (e.g., some states may allow construction work in progress in rate base while other states do not). One utility also could have higher growth in output, which because of economies of scale would cause its average costs to decrease relative to other utilities. Each of these factors could cause one utility to rank lower than other utilities even though management behavior was no factor.

An illustration: identifying price factors for a natural gas utility

One or more of the following general factors can explain the large differences in retail gas prices between natural gas utilities, both within a state and across states: (1) customer-demand characteristics (e.g., load factor, gas usage per customer, use of gas for space heating), (2) cost and supply conditions (e.g., proximity to gas fields, the number of pipelines serving the utility), and (3) management practices (e.g., hedging strategies, proficiency in cost control). A major component of gas prices to small retail customers is gas commodity costs. These costs, when added to pipeline costs and distribution margins, comprise the retail price charged to small customers. Thus, in examining price differences across utilities, an analysis should first disaggregate the differences by individual functions. For example, to what extent do higher-

priced gas utilities have higher pipeline rates, distribution margins, and commodity gas costs?

One can imagine several factors accounting for price differentials across gas utilities. They include:

1. *Levels of storage available to each gas utility:* Those utilities without storage capability would tend to have higher costs, assuming other things held constant. Some gas utilities tend to have higher rates partially because of their lower storage capability relative to other gas utilities.
2. *Rate legacy:* Cost allocation methods and the ratios of rates to different customer classes may vary across utilities. For various reasons, the residential rates of some gas utilities may reflect cost-of-service principles less than those of other gas utilities. Also, different accounting treatment of storage costs and other cost components can affect rates.
3. *Cycling issue with purchased gas adjustment (PGA) clauses:* The adjustment period might not be uniform across utilities; adjustments, for example, might be monthly, quarterly, or annually, depending upon the gas utility. The periods for which a utility adjusts its purchased gas costs can, therefore, distort a snapshot comparison of prices across utilities.
4. *Gas procurement and hedging practices:* Transaction arrangements and hedging activities are important factors in affecting purchased gas costs. Viewed from across the country and within individual states, one observes a wide discrepancy in physical and financial hedging by utilities. This discrepancy means that when wholesale gas prices change, up or down, there would be a lesser rate effect on those utilities that have hedged more. Differences in management philosophy may explain why some gas utilities hedge more or less than other utilities.
5. *Distribution margins (i.e., the portion of retail rates left over after subtracting gas commodity and pipeline costs):* Large differences exist across utilities; a major factor is the sales volumes or throughput per customer. Distribution margins are generally higher for rural utilities, for utilities in warmer climates, and for those utilities with recent capital expenditures recovered in rates. Prices in warm-weather states are generally higher because utilities have to recover their fixed costs over fewer sales, which drives up their average cost and prices.
6. *Pipeline rates:* Factors include the zonal area of a pipeline, as the Federal Energy Regulatory Commission (FERC) allows price differences between designated zones. The number of pipelines that move gas to a specific utility may affect rates (i.e., competitive conditions would tend to place a downward pressure on rates). The load factor of firm customers may also be important—for example, utilities with lower load factors would tend to have higher pipeline rates because of FERC's straight-fixed rate design.

7. *Different services offered under the base rate:* Some utilities may still provide maintenance and other services under base rates.
8. *Economies of scale:* Larger gas utilities may have lower average cost because of the economies from procuring gas and pipeline transportation at greater amounts.
9. *Economies of scope:* Some utilities like combination electric and gas utilities may perform more functions that offer synergies with other functions, which would lower costs.

One consulting firm, Pacific Economics Group (PEG), has conducted several studies that apply an econometric statistical cost model to explain differences in non-fuel O&M costs and other costs across energy utilities.²⁹ These studies show that the services provided, the scale of operations, the prices of inputs, and other business conditions explain some of the cost differences across utilities. Their studies have found, for example, that greater use of cast iron increases both maintenance and replacement costs. PEG also found that scope economies lower costs. They distinguish between the effects on cost from increased throughput per existing customer and from the addition of new customers. (The latter has a greater effect.) They also found that natural gas utilities serving urban areas have higher costs partially because of the greater difficulty of installing mains and service lines.

²⁹ See, for example, Pacific Economics Group, *The Cost Performance of Boston Gas*, January 28, 2003; Pacific Economics Group Research, *Benchmarking the Operating Performance of Portland General Electric*, February 10, 2010; and Mark Newton Lowry et al., "Econometric Benchmarking of Cost Performance: The Case of U.S. Power Distributors," *The Energy Journal* 26, 3 (2005): 75-92.

Appendix B: Real Unit Cost Indices

The real unit cost index equals:

$$UC_k = E_k^r / Q_k,$$

where UC_k is the unit cost in constant dollars for utility function k , E_k^r is total cost for function k deflated by a price index, and Q_k is the output measure for function k . The percentage change in real unit cost equals the difference between the percentage changes in total cost in real dollars and output. An increase in UC_k over time reflects a decline in productivity, since mathematically the relationship between real unit cost and productivity is reciprocal. If E_k^r equals total cost for a utility and output is the total kilowatt-hours or therms provided by a utility, then UC_k represents the inverse of the total factor productivity for the utility.

Assume we want to calculate the change in a utility's real unit cost for operation and maintenance during the period 2007-2009. We define output as total sales. We have the following statistics as reported by the utility:

	2007	2008	2009
Total O&M expenses (10 ³ dollars)	\$223,063	\$242,789	\$266,519
Expense index	1.000	1.088	1.195
Price index	1.000	1.084	1.141
Sales (Gwh)	33,440	34,271	34,789
Sales index	1.000	1.025	1.040

The indices measure the values of O&M expense, price, or sales for a particular year relative to their values for the base year (2007). The expense index for 2008 (1.088), for example, equals the ratio of total O&M expenses in 2008 and 2007 (i.e., \$242,789/\$223,063); the sales index for 2009 (1.040) equals the ratio of sales in 2009 and 2007 (i.e., 34,789/33,440).

We can calculate the percentage change in real unit cost as

$$\ln (UC_{k,t}/UC_{k,t-1}) \cdot 100 = [\ln (E_{k,t}/E_{k,t-1}) - \ln (P_{k,t}/P_{k,t-1}) - \ln (Q_{k,t}/Q_{k,t-1})],$$

where k is the function under review, t and $t-1$ are time periods, and P is the price index used to convert expenses of different periods into constant dollars; UC and E , as defined above, are unit cost and total cost. Applying the numbers in the table, the growth rates for real unit cost percent during 2007-2008 and 2008-2009 are -2.1 percent and 2.7 percent, respectively.

A regulator can acquire this information for the utility under review as well as for other utilities with similar characteristics. Differences in growth rates can reveal whether the utility under review is an outlier or an average performer, as determined by the mean growth rate of the other utilities compared with the utility's growth rate. Regulators can calculate real unit cost indices with time series, cross-sectional, or panel data. With time series data, regulators can compare the performance of an individual utility over time with itself or a peer group of utilities. Cross-sectional data can compare a utility's performance with other utilities at specific points in time. Panel data can provide comparisons of performance both over time and at specific points in time.

Appendix C: An Illustration of the Use of Performance Measures in a Cost-Sharing Incentive Mechanism

Example of a cost-sharing mechanism

Assume that a regulator has approved an incentive mechanism for purchased gas. The mechanism has a cost-sharing arrangement, expressed as the following:

$$C_f = C_a + s \cdot (C_b - C_a), \text{ or}$$

$$C_a \cdot (1-s) + C_b \cdot s,$$

where C_f is the costs flowed through to consumers, C_a equals actual costs incurred by the utility, s is the sharing parameter, and C_b equals benchmark costs. A regulator might want to modify the above plan to include a “dead band.” This provision would account for the likelihood that small deviations of a utility’s performance from the benchmark do not reflect management behavior. These deviations may represent “white noise” explained by factors beyond utility-management control.

Applying the previous formula, assume that C_a equals \$100 million, C_b equals \$120 million and s is 0.2. Then, C_f equals \$100 million + 0.2(\$120 million - \$100 million) = \$104 million. The results seem positive: the utility earns \$4 million in rewards and consumers ostensibly receive benefits of \$16 million from lower gas purchasing costs, after adjusting for the utility reward. (The assumption is that actual costs would equal \$120 million without the incentive mechanism.) Consumers pay the *actual costs* plus the *reward* to the utility (when $C_b > C_a$) or the actual costs minus the penalty to the utility (when $C_b < C_a$).

Consumers benefit only when the reduction in actual costs exceeds the reward to the utility. So for consumers to benefit from an incentive mechanism, $(C_b - C_a)$ must be greater than $s \cdot (C_b - C_a)$. Thus, it seems, at least mathematically, that consumers always benefit when the utility beats the benchmark, since s is less than one. But this assumes that $(C_b - C_a)$ represents the real cost savings from the incentive mechanism. This presumption may distort reality if C_b , in fact, does not reflect what the utility’s costs would have been in the absence of the incentive mechanism. (The subsection below, “The effects of a biased benchmark,” examines this problem.)

When contemplating incentive mechanisms, regulators need to consider the tradeoff between: (1) creating strong incentives for superior performance and (2) achieving a balanced distribution of economic gains between the utility and its customers. Cost-sharing mechanisms such as the one presented above represent a compromise that provides better incentives for cost efficiency than cost-plus arrangements but mitigates the likelihood that utility customers would earn an unreasonably small share of the total economic gain from improved utility performance. Under a typical incentive mechanism, a utility receives additional revenues from improved

performance. A relevant question for “equity” purposes is: What benefits do consumers receive when utility performance improves? Do these benefits at least cover the additional revenues that consumers have to pay? To say it differently, do the benefits of improved performance to consumers coincide with the additional revenues to the utility? Although in many instances the benefits to consumers may be non-quantifiable, regulators should have the ability to make an informed decision on whether the benefits to consumers from improved performance correspond to the additional revenues that a utility receives. The significance of consumer benefits falling short of additional revenues is that the utility receives a windfall gain at the expense of consumers.

The “benchmark” cost is clearly pivotal for dividing up the gains between the utility and consumers. One tough challenge for regulators is to set the correct benchmark. The wrong benchmark can be derived from: (1) gamesmanship by utilities and consumer groups (e.g., biased cost revelation by the utility), and (2) incomplete information. The utility will argue for a benchmark that will make it easy to earn a reward and avoid a penalty; consumer groups will attempt to make it hard on the utility to earn a reward. The utility might reveal its cost opportunities to be lower than what they really are (e.g., the utility would argue that it has certain constraints in reducing costs when, in fact, it has no such constraints.) The regulator finds it difficult to know the “true benchmark.” What costs should the utility have under reasonable management? What costs would the utility incur in the absence of an incentive mechanism? What are reasonable utility actions deserving of neither a reward nor a penalty?

A good benchmark also is beyond the control of a utility. If the utility, through its actions, is able to affect the “benchmark” value, distortion can readily occur. A utility, for example, might be able to strategically manipulate the benchmark to improve its profits at the expense of consumers. The “benchmark” value should also change over time in response to changed market and other conditions.³⁰ In other words, it should adapt to changes in outside conditions. The intent of an incentive mechanism is to direct the incentives at only those activities over which the utility has some control.

The effects of a biased benchmark

The cost effect on consumers when a utility is able to manipulate the benchmark, and assuming no change in actual costs, is as follows: let $\Delta C_f = \Delta C_a \cdot (1-s) + \Delta C_b \cdot s$; with $\Delta C_a = 0$, $\Delta C_f = s \cdot \Delta C_b = \Delta R$ (rewards). The result is a zero-sum game, in which the additional reward to the utility is a dollar-for-dollar payment from consumers.

Assume that C_b equals the calculated benchmark and C_b^* is the true (“unbiased”) benchmark, with $C_b > C_b^*$. One defensible measure of the true benchmark is the cost that the

³⁰ See, for example, Ken Costello and James F. Wilson, *A Hard Look at Incentive Mechanisms for Natural Gas Procurement*, NRRI Report 06-15, November 2006, at <http://www.nrri.org/pubs/gas/06-15.pdf>.

utility would have incurred in the absence of the incentive mechanism. The utility receives a higher reward, equal to $s \cdot (C_b - C_b^*)$. What is the effect on consumers? It depends, but here we assume an alternative world without an incentive mechanism. The following calculates the effect on consumers (i.e. the change in the costs flowed through to consumers) from a benchmark cost that is set too high:

$$\Delta C_f = \Delta C_a + \text{Reward to the Utility}$$

$$\text{Let } C_b = C_b^* + \eta \text{ and } C_b^* = C_{a0}$$

$$\text{Then, } \Delta C_f = (C_{a1} - C_{a0}) + s \cdot (C_b^* + \eta - C_{a1})$$

$$\Delta C_f > 0, \text{ when } s \cdot (C_{a0} + \eta - C_{a1}) + (C_{a1} - C_{a0}) > 0, \text{ or}$$

$$\Delta C_f > 0, \text{ when } (1 - s) \cdot (C_{a0} - C_{a1}) < s\eta$$

The actual benchmark (C_b) exceeds the true benchmark by η . The true or unbiased benchmark (C_b^*) equals the actual costs incurred in the absence of the incentive mechanism (C_{a0}). One term not yet defined is C_{a1} , which equals the actual cost with the incentive mechanism in place. The incentive mechanism should reduce the actual cost (i.e., $C_{a1} < C_{a0}$).

Taking a numerical example, assume that C_b^* (i.e., C_{a0}) is \$50 million, C_b is \$54 million, s is 0.5, and C_{a1} is \$49 million. With no incentive mechanism, consumers pay \$50 million. With the incentive mechanism, consumers pay \$49 million (C_{a1}) + 0.5(\$54 million - \$49 million), which equal \$51.5 million. In this example, consumers become worse off even when the utility lowers its cost. The reason is that consumers pay an excessive reward to the utility because the benchmark cost was set too high. Performance assessment can help regulators set an appropriate benchmark that would mitigate the chances of a utility earning a disproportionate share of the economic gains from improved performance.

Appendix D: The Relationship between Total Factor Productivity and Average Cost

- Average cost = Total cost / Output level
- Average cost = (Price of inputs · Input level) / Output level
- Average cost = Price of inputs / (Output level / Input level)
- Average cost = Price of inputs / *Total factor productivity*

Assume that when the binding regulatory condition holds in which total cost (or the total revenue requirements) equals operating revenues, an increase in total factor productivity causes a decline in average cost, rates and revenue requirements. Growth in total factor productivity can originate from different sources—for example, technology improvements, economies of scale, higher output, less waste of internal resources, and more efficient mix of inputs. Some of these factors fall within the control of utility management, while others fall outside.

Assume a hypothetical firm that uses only one input whose price is \$5 per unit and that its total factor productivity equals 2. The average cost of the utility is then \$2.50; that is, for each unit of output the utility uses one-half input. Since one input costs \$5, one-half input is \$2.50. Assume that over time the input price increases by 5 percent and that total factor productivity increases by 2 percent. Average cost would then increase to $5(1.05)/2(1.02)$ or \$2.57.

As a general condition, when input prices increase faster than total factor productivity, prices would tend to rise. Prices would tend to fall when total factor productivity rises faster than input prices.