DEFINING, MEASURING AND EVALUATING THE PERFORMANCE OF ONTARIO ELECTRICITY NETWORKS: A CONCEPT PAPER

REPORT TO THE ONTARIO ENERGY BOARD

April 2011

Pacific Economics Group Research, LLC

DEFINING, MEASURING AND EVALUATING THE PERFORMANCE OF ONTARIO ELECTRICITY NETWORKS: A CONCEPT PAPER

REPORT TO THE ONTARIO ENERGY BOARD

April 2011

Lawrence Kaufmann, Ph.D

Senior Advisor

PACIFIC ECONOMICS GROUP RESEARCH LLC

22 East Mifflin, Suite 302
Madison, Wisconsin USA 53703
608.257.1522 608.257.1540 Fax

TABLE OF CONTENTS

1. INTRODUCTION AND EXECUTIVE SUMMARY 1

1.1 INTRODUCTION 1

1.2 EXECUTIVE SUMMARY 4

2. BASIC DEFINITIONS 15

2.1 OUTPUTS AND OUTCOMES 15

2.1.1 Traditional Network Services 16

2.1.2 "New" Network Services..17

2.1.3 Outputs and Output Quality..21

2.2 STANDARDS AND PERFORMANCE 22

2.3 PRODUCTIVITY AND EFFICIENCY 22

2.3.1 Productivity 22

2.3.2 Efficiency..24

2.3.3 The Relationship Between Productivity and Efficiency..................25

2.4 REGULATORY INCENTIVES 28

3. MEASUREMENT 30

3.1 OUTPUTS...30

3.1.1 Appropriate Output Measures in Setting Rates..............................30

3.1.2 Output Measures for Cost Analysis...33

3.1.3 Measuring Quality 35

3.2 INPUTS...41

3.2.1 O&M Cost and Inputs..41

3.2.2 Capital Cost and Inputs...42

3.2.2.1 Measuring Capital Stocks 42

3.2.2.2 Measuring Capital Flows.. 46

3.2.2.3 Capital Data for Ontario Distributors 47

3.3 PRODUCTIVITY 48

3.3.1 Index-Based Estimates of TFP Trends *48*
3.3.2 Econometric Estimation of TFP Trends *49*
3.4 PERFORMANCE 51
3.4.1 Index-Based Methods *51*
3.4.2 Econometric Methods *51*
3.4.3 Stochastic Frontier Analysis *53*
3.4.4 Data Envelope Analysis *53*

4. SETTING STANDARDS 56

4.1 Standards and Performance Benchmarking *56*
4.2 Principles for Establishing Appropriate Cost Standards *57*
4.3 Service Quality Standards *62*

5. CREATING INCENTIVES 66

5.1 INDEXED RATE CAPS 66
5.2 REVENUE CAPS 68
5.3 EARNINGS SHARING MECHANISMS 70
5.4 BENCHMARK REGULATION 72
5.4.1 Comprehensive Benchmark Plans *72*
5.4.2 Non-Comprehensive Benchmark Plans *74*
5.5 PLAN TERMINATION PROVISIONS 75
5.6 MENU APPROACHES 78
5.7 TARGETED INCENTIVES TO ACHIEVE REGULATORY OUTCOMES 80
5.8 REGULATORY JUDGMENT/DISCRETION 81

6. EVALUATING PERFORMANCE 82

6.1 RIIO 82
6.2 ILLUSTRATIVE FRAMEWORK 86

7. GOING FORWARD 91

7.1 FOUNDATIONS IN ONTARIO 91
7.2 FURTHER ISSUES TO ADDRESS 93

APPENDIX ONE: DECOMPOSITION OF TFP GROWTH 95

APPENDIX TWO: PHYSICAL AND MONETARY CAPITAL MEASURES ... 99

APPENDIX THREE: EVALUATION OF BENCHMARKING OPTIONS....... 111

A.3.1 Index Based Benchmarks 112

A.3.1.1 Partial Factor Productivity (PFP).. 112

A.3.1.2 Total Factor Productivity (TFP) 112

A.3.2 Econometric Cost Functions 114

A.3.2.1 Advantages of Econometric Cost Functions 114

A.3.2.2 Disadvantages of Econometric Cost Functions 115

A.3.3 Stochastic Frontier Analysis 117

A.3.3.1 Advantages of SFA 117

A.3.3.2 Disadvantages of SFA 117

A.3.4 Data Envelope Analysis 118

A.3.4.1 Advantages of DEA 118

A.3.4.2 Disadvantages with DEA 119

A.3.5 Some Evidence from Other Markets.. 126

APPENDIX FOUR: REVIEWS OF PERFORMANCE MEASURES AND STANDARDS IN DIFFERENT JURISDICTIONS....................................... 131

A.4.1 UK.. 131

A.4.2 Netherlands 141

A.4.3 Mississippi "PEP"... 143

A.4.4 Mexico.. 149

A.4.5 Victoria, Australia.. 153

REFERENCES.. 156

1. Introduction and Executive Summary

1.1 Introduction

This is a particularly challenging time for the Ontario Energy Board ("OEB" or "the Board"). The Board continues to be the economic regulator of more than 80 electricity distributors, two large gas distributors (with storage and/or transmission assets as well), five electricity transmission utilities, Ontario Power Generation, and the Independent Electricity System Operator. Traditionally, the Board's regulatory roles involved protecting consumers, promoting efficient and low cost energy services, and maintaining the financial viability of regulated energy utilities.

These obligations remain, but legislative and policy changes have substantially broadened the Board's responsibilities. Most importantly, the Green Energy and Green Economy Act (GEA) has added three new regulatory objectives to the OEB's economic regulation tasks. These objectives are promoting energy conservation and demand management; facilitating the development of a smart grid; and encouraging electricity generation from renewable energy sources. In broad terms, the GEA compels the OEB to consider the *social* costs associated with energy use at the same time that it attempts to ensure that energy is provided efficiently and at the least cost to utility consumers (subject to the constraint that utilities remain financially viable). These two, broad regulatory roles may conflict in practice, since complying with GEA mandates will impose costs on energy networks.

Ratemaking complexities are magnified in Ontario by the structure of regulated energy industries in the Province. Electricity distribution is particularly diverse, with six large distributors serving a majority of Ontario ratepayers and approximately 80 small distributors serving the remainder. Because the number of electricity distributors is likely to remain large, the regulatory burdens associated with "traditional" ratemaking are not likely to abate, nor will the OEB be expand its Staff to be commensurate with its increased regulatory responsibilities. In this more complex and demanding environment, there may be a greater need than ever for regulatory innovation, and enhanced regulatory "technologies," that can enable the Board to do more with its existing human resources.

One means of addressing these regulatory challenges could involve greater use of performance standards. This would be consistent with ratemaking trends in the Province, because in recent years the OEB has placed greater emphasis on the use of performance standards in electricity regulation. For example, the Board has established standards for some customer service-related aspects of distribution service. It is also considering a reliability standards regime for distributors and reporting requirements for transmission reliability. In the "third generation" incentive regulation plan approved for electricity distributors, the Board also used econometric and unit cost benchmarks to identify three efficiency "cohorts" of electricity distributors. Each cohort was in turn assigned a different "stretch factor" component of the X factor in a multi-year, "inflation minus X" rate adjustment mechanism.

While these performance measures and standards may be useful for promoting some regulatory objectives, they may not be sufficient for the challenges the Board is currently facing. In particular, complying with the GEA mandates will require new investments from Ontario's electric distribution and transmission utilities. To mitigate the impact of these additional investments on customer rates, the Board must ensure that utilities' approved network investment plans are implemented as efficiently as possible.

Few of the performance measures that have been used to date, or are under consideration, focus on efficient network investment. For example, the studies used to set differential stretch factors in third generation incentive regulation were based entirely on distributors' operations, maintenance and administrative (OM&A) costs, not their capital costs. This benchmarking work is clearly not sufficient for assessing *overall* cost performance. Current benchmarking methods also do not assess potential tradeoffs among performance metrics, such as the potential to reduce OM&A costs by "gold plating" capital or the linkages between efficient costs and appropriate quality levels. Additional work would be needed if the existing, cost benchmarking studies are to assist the Board in evaluating whether distributors are implementing their investment plans efficiently, or providing "value for the money" to consumers.

A broader conceptual framework may prove valuable for helping OEB Staff and stakeholders understand and evaluate energy network performance in a comprehensive, holistic manner. A comprehensive framework could also assist the Board as it endeavors to ensure that utilities' capital investment plans are implemented as efficiently as possible. To promote this end, the framework should also consider how to link performance evaluations with regulatory consequences.

This last issue is complicated by the fact that the Board administers a number of different regulatory approaches simultaneously. Distributors and transmitters file cost of service applications (*e.g.* for "rebasing" rates when incentive regulation plans expire) and can file company-specific rate proposals or multi-year incentive regulation mechanisms.[1] In principle, a broad conceptual analysis should address performance measures and regulatory design issues for all the ratemaking approaches that the Board administers. At the same time, the Board will naturally prefer robust standards, measures and mechanisms that can be applied across a variety of regulatory approaches. Developing a huge panoply of standards and measures would almost certainly be unwieldy and could complicate, rather than facilitate, the Board's achievement of regulatory objectives, including reducing regulatory burdens.

The Board has asked Pacific Economics Group Research (PEG) to advise Staff on Performance Definition and Measurement Issues as the OEB works toward developing a New Regulatory Framework for Electricity. PEG's first task was to prepare a "Concept Paper" that provides an overall framework and structure for analyzing the relevant issues. Among the topics to be addressed in the Concept Paper are the following:

- Clearly defining terms such as "outputs," "standards," "productivity," "efficiency," and "performance" that will be used throughout the consultation
- Assessing how outputs, inputs, capital costs, productivity, efficiency and performance should be measured under the various regulatory approaches the OEB administers
- Developing some basic principles that should be used to evaluate the potential use of performance standards in regulation
- Analyzing different incentive mechanisms that could be helpful for ensuring that utilities comply with regulatory objectives as efficiently as possible
- Discussing the basic process the OEB should consider when evaluating utilities' performance *ex post*

[1] Although incentive regulation plans are typically developed for the industry as a whole, utilities are not prohibited from proposing alternate, mulit-year mechanisms.

1.2 Executive Summary

The results of PEG's research can be briefly summarized. Regarding the basic definitional issues, "outputs" are the goods or services that firms provide to their customers. "Outcomes" are the end-states experienced by either customers or companies themselves after outputs have been provided. With respect to companies, outcomes can be changes in behavior, knowledge or functioning that result from transactions between companies and their customers. For customers, outcomes would reflect changes in their (subjective) utility. For example, in the case of an electricity distribution network, "delivered electricity" (measured in kWh) would be an example of an output. "Customer satisfaction" would be an example of an outcome for the customer.

In the current environment, it is important to recognize that energy networks are providing both "traditional" and "new" network services to their customers. The line between the traditional and the new, however, is not always clear. Some of the investments necessary to comply with GEA mandates may have implications for how traditional outputs are provided, while some assets that help perform traditional functions more efficiently (*e.g.* smart meters) may prove valuable in helping networks cope with the challenges of delivering power from more diverse and less centralized supply sources to end-users. Regulators may therefore need to take a broader view of how traditonal outputs are being provided, and be sensitive to the potential linkages between investments needed for the "new" marketplace and the network outputs that have traditionally been subject to economic regulation.

A standard is a quantitative benchmark that is used to evaluate how effectively regulatory objectives are being achieved. Standards can be established at different levels. For example, there may be a "minimum standard," which reflects the minimum acceptable level of achievement of a given objective by any regulated network. An "average standard" reflects an expectation of how effectively, on average, a network industry is expected to achieve an objective. There may also be a "superior" standard, which could lead to rewards for utilities that achieve at levels beyond this standard.

Performance refers to how effectively a network is achieving the desired regulatory objective. This notion of performance is obviously closely related to the concept of a standard. A standard is what is used to judge actions or outcomes that are expected to contribute towards regulatory objectives. Performance reflects the measure of how effectively utilities are in fact achieving regulatory standards.

Productivity measures the transformation of inputs into outputs. In the present context, "inputs" refer to the resources an energy network procures in order to provide network outputs. Total factor productivity (TFP) measures the relationship between all the outputs provided by a utility and all the inputs that the utility procured to provide those outputs. Partial factor productivity (PFP) measures the relationship between the utility's comprehensive output and a more narrow measure of inputs.

Economists typically distinguish between two types of efficiency: productive efficiency and allocative efficiency. Productive efficiency refers to the degree to which a firm produces the maximum potential output given available technologies. There are both supply- and demand-side dimensions of allocative efficiency. Demand-side allocative efficiency concerns how efficiently a firm is pricing its goods or services. In the short-run, prices become allocatively efficient to the extent that they reflect the firm's marginal cost of service. Supply-side allocative efficiency refers to using the optimal *mix* of inputs in production, for given levels of input prices. For regulated industries, regulatory cost can also be considered as a dimension of efficiency.

Productivity is a broader concept than efficiency. Certainly, improved productive and allocative efficiency can contribute to TFP growth. However, TFP growth can also result from factors beyond the network's control, including broader technological change, the realization of scale economies, and environmental and policy-related effects.

Incentives can be defined as *ex ante* regulatory rules that: 1) encourage behavior by utilities that promotes desired regulatory outcomes; and 2) if executed successfully, will lead to financial benefits for utilities. This definition appears consistent in spirit with Staff's distinction between "alternative" and "incentive" mechanisms in its Discussion Paper on the treatment of infrastructure investment.

The appropriate measures of network outputs can differ somewhat depending on the application. When setting rates in either cost of service proceedings or multi-year, inflation minus X formulas, the appropriate output choices are clearly the utility's billing determinants *i.e.* the outputs for which customers are explicitly billed. When selecting outputs for cost comparisons, however, there is more room for flexibility. Output measures for cost analyses should be selected on how strongly they correlate with the cost of providing regulated services. Billing determinants like customer counts will still be important output measures in these analyses, but they are not the only potential measures. Any variable that can be defined as an output and impacts network costs is appropriate when undertaking cost comparisons.

There are many facets of the quality of service provided by energy networks. We believe the measured service quality indicators used in network regulation should ideally satisfy four, common sense criteria:

- they should be related to the aspects of service that customers value;
- they should focus on monopoly services;
- utilities should be able to affect the measured reliability; and
- the indicators should be sensitive to "pockets" of system quality problems.

There are two main types of inputs for energy networks: operations and maintenance (O&M) inputs, and capital. Measuring capital inputs is far more complex than measuring O&M inputs. Although capital measurement concepts can be organized in many sensible ways, one straightforward approach is to divide the discussion between the measurement of capital *stocks* and capital service *flows*. Capital stock measures pertain to the quantity of capital that exists at a given point in time. Capital flows measure the services provided by that stock of capital over a timer period. Typically, capital flows are more relevant for the capital input quantities or capital cost measures that are used in TFP or cost benchmarking research.

Determining the flow of capital services is usually not straightforward. One reason is that capital goods are durable and provide a flow of services over a multi-year period. While companies may record the price of the capital good when it is purchased and the date when that good is retired (or replaced), they rarely attempt to measure the flow of services provided over a given period. A related problem is that capital services are usually implicitly provided within a single enterprise rather than via arms length transactions in the marketplace. Firms often consume capital goods over a multi-year period, which implies that the provider and the ultimate user of capital services are located in the same economic enterprise.

Determining the flow of capital services is usually not straightforward. One reason is that capital goods are durable and provide a flow of services over a multi-year period. While companies may record the price of the capital good when it is purchased and the date when that good is retired (or replaced), they rarely attempt to measure the flow of services provided over a given period. A related problem is that capital services are usually implicitly provided within a single enterprise rather than via arms length transactions in the marketplace. Firms often consume capital goods over a multi-year period, which implies that the provider and the ultimate user of capital services are located in the same economic enterprise. Conceptually, the flow of capital services associated with an asset could be observed as the amount that consumers pay to rent that asset for a given period. In practice, these flows are typically measured with capital service price formulas that include terms that correspond to taxes, the opportunity cost of capital (*i.e.* the cost of funds), depreciation, and capital gains.

Estimates of TFP growth for network industries are typically used to calibrate the X factor in index-based, incentive regulation plans for network utilities. TFP in most regulatory applications is measured using indexing methods. When this is the case, TFP growth is defined as the growth in an index of comprehensive output quantity minus the growth in an index of comprehensive input quantity. In ratemaking applications, the output quantity index should be measured using the networks' billing determinants. The input quantity index will be measured using subindexes of O&M and capital inputs. In addition to estimating historical TFP trends using indexing methods, econometric methods can be used to estimate TFP growth. Such an approach is well-suited for projecting TFP growth when there is a lack of historical, time series data. However, this approach is also considerably more complex than index-based TFP estimates.

There are a variety of methods that can be used to quantify, and "benchmark," a network's cost or quality performance. This Paper considers four performance measurement techniques: index-based methods; econometric cost functions; stochastic frontier analysis (SFA); and data envelope analysis (DEA). Index-based methods can be used to measure performance, as well for setting the terms of rate adjustment formulas. Econometric methods use statistical and regression techniques to quantify the drivers of network cost. SFA is similar in many respects to other econometric cost models, although it estimates an inefficiency factor for each firm. DEA does not estimate the parameters of a cost function but instead uses linear programming techniques to "envelope" data on sample firms that relate outputs to inputs. DEA is therefore essentially a technique for identifying what are known in economics as isoquant or isocost curves and in measuring the distance of individual firms from the efficient cost (production) frontier reflected in that isocost (isoquant).

In general terms, a regulatory "standard" gauges how effectively a utility is achieving regulatory objectives. There can be different standards that reflect regulators' expectations for what constitutes acceptable, average, desirable and/or superior performance. Many paths can be taken towards using performance standards to encourage long-run regulatory objectives. When evaluating different alternatives, regulators should consider how options differ in terms of risk and information requirements. The approach that is most appropriate in any given situation will depend on a number of factors, including the institutional environment and the amount and quality of data that are available. In all cases, however, several factors should be kept in mind when making the competitive market paradigm operational.

First, movements towards long-run efficiency levels will take place gradually. One reason is that adjusting company operations to achieve greater efficiencies is usually costly. Any movement towards benchmark-based performance targets should therefore take place gradually.

It is also not reasonable to impose "frontier" performance standards on all firms in the industry since this does not allow returns to be commensurate with performance. Companies must always have "room" to outperform the benchmark that is reflected in the prices they face. This enables the firm to be appropriately rewarded for superior performance. If the industry's best-observed practice is imposed on all firms, any firm that fails to achieve this standard will earn below average returns. This would be true even for superior performers that nevertheless fall short of the industry's best performance. This outcome is clearly contrary to having returns be commensurate with performance and thus is not consistent with effective regulation.

It is also important to recognize that there will be considerable uncertainty about what constitutes a "frontier" performance level. Targets established through benchmarking should be cognizant of this uncertainty. Regulators should not impose performance standards for which there is significant probability that well-managed utilities will fail to achieve these targets. The benchmarks should therefore make appropriate allowance for the uncertainty associated with attaining the target performance levels.

A variety of mechanisms can be used to create incentives for energy networks to achieve regulatory objectives. This Paper considers: indexed rate caps; revenue caps; earnings sharing mechanisms; benchmark regulation plans; plan termination provisions; menu approaches; and targeted incentives. We also consider regulatory approaches that rely more on regulatory discretion and judgment, rather than established incentive regulation plans, as the means for creating appropriate behavioral incentives.

Indexed rate caps are very familiar to stakeholders in Ontario, since this approach has been applied several times in the Province in electricity and gas distribution regulation. Indexed rate caps adjust allowed rates by the growth in an inflation factor, minus an X factor, plus or minus a Z factor. The inflation factor is the growth rate in an external price inflation measure. The X-factor is also sometimes called a "productivity factor" since, in North American proceedings, its value is typically estimated using estimates of industry TFP growth. The Z-factor adjusts the allowed change in rates for reasons other than inflation and productivity trends. The main rationale for Z-factors is to recover the impact that changes in government policy have on the company's unit cost.

Indexed base rate caps create strong performance incentives in two primary ways. One is that indexed rate caps are determined by inflation and productivity data that are "external" to the utility itself. This means that during the term of the rate indexing plan, energy networks have very strong incentives to reduce their unit costs, since any changes in their company's unit cost that are less than the change in the inflation minus X formula will go straight to the "bottom line" and boost company profitability. In addition, indexed rate caps can create balanced incentives to control all costs. This results from the fact that indexed rate caps apply to both capital and O&M costs, and controlling capital and operating costs both contribute to improved productivity.

Comprehensive revenue caps sometimes are also sometimes called revenue decoupling mechanisms. Under these types of plans, an electricity network's allowed revenues do not depend on its energy (*i.e.* kWh) deliveries. Revenue caps can be updated over time using indexing mechanisms. Revenue caps can also be established on a per customer basis, and adjusted over time to reflect growth in the number of customers served by the utility. In principle, revenue caps can also be constant and adjusted only in conjunction with rate rebasings.

Indexed revenue caps are identical in many respects to indexed rate caps, except revenues rather than utility rates are subject to formula-based restrictions. Comprehensive revenue caps create the same incentives for productive and supply-side allocative efficiency as indexed rate caps. The reason is that both use external data in the adjustment formula and both are comprehensive mechanisms that apply to all network costs. In addition, comprehensive revenue caps create stronger incentives to pursue conservation and demand management (CDM) objectives, since the mechanism recovers losses in network revenues that occur when customers reduce their energy consumption.

Earnings-sharing mechanisms (ESMs) adjust a company's allowed rates when its rate of return has been in an established range in a recent historical period. The mechanisms are set in advance of their use and typically function for several years. The most widely used earnings measure is return on equity (ROE).

ESMs have some important advantages in regulatory applications. One is their ability to mitigate risk and keep earnings within what are deemed to be acceptable bounds. This property is, of course, greater when ESMs are symmetric and apply to "under" as well as "over" earnings. ESMs also share benefits *as they are realized* under the plan. This naturally leads to some alignment of shareholder and customer interests.

On the downside, it must be recognized that ESMs are not inherently incentive regulation mechanisms. Indeed, compared with traditional cost of service regulation where utilities retain *all* the benefits of unit cost reductions between rate reviews, sharing earnings with customers *reduces* utilities' incentives to pursue productive and allocative efficiencies. ESMs will only be a form of incentive regulation if they extend the period of "regulatory lag" between rate reviews compared with the cost of service regulation.

ESMs also do not by themselves guarantee that customers benefit from an incentive regulation plan. ESMs also increase the regulatory focus on annual earnings computations, which can exacerbate concerns with inherently controversial issues like utility-affiliate transactions and cost allocations. Earnings reviews associated with ESMs can raise regulatory costs and reduce net benefits to all stakeholders. The calculation and review of earnings in some approved ESMs has in fact been contentious.

Benchmark regulation involves evaluating one or more indicators of company activity against external performance standards (or benchmarks). A benchmark plan's key features are the performance indicators, the performance benchmarks, and the rate adjustment (or award) mechanism. A comprehensive benchmark plan is one in which benchmarking mechanisms cover substantially all the facets of company performance that matter to customers. In contrast, a non-comprehensive benchmark specifically targets a small set of performance areas.

The performance benchmarks used in benchmark plans are also varied. A common benchmark is a company's average performance on the selected indicator in a period just prior to plan commencement. A company is therefore rewarded or penalized for its performance relative to recent history.

Comprehensive benchmark plans can in principle create strong performance incentives. Plans are comprehensive and can therefore promote allocative as well as productive efficiency. On the other hand, comprehensive benchmark plans can be challenging to design and implement. Addressing these challenges could raise the costs of implementing and administering such plans. Benchmark plans can be designed in many different ways. Some of these plan designs will involve comprehensive benchmarking evaluations, which requires decisions to be made over the appropriate choice of benchmarking techniques and appropriate values or targets to be reflected in the comprehensive performance standards.

Plan termination provisions are another, increasingly important aspect of incentive regulation. One important consideration here is simply the term of the incentive regulation plan. Provisions that are established for resetting rates when the plan ends are also important.

Longer plan terms clearly strengthen performance incentives. Longer terms are especially useful in encouraging initiatives that involve up-front costs to achieve long-run efficiency gains. On the downside, longer plan terms can increase both business and regulatory risk. This makes them less suitable for businesses undergoing rapid change or trends that are not otherwise captured in the design of the incentive regulation plan.

Rate reset provisions are rules that allow rates at the end of an incentive regulation plan not to be immediately "trued up" to the company's cost of service. Rate reset provisons are important because they affect the incentives created by the regulatory mechanism. If a full cost-based rate true-up can be avoided, performance incentives will be strengthened. Regulatory costs may also be reduced because, in principle, incentive-based rate reset provisions reduce concerns that firms are simply deferring cost increases and may be "gaming" the costs reported in the test year used to set rebased rates. These mechanisms are especially important for term benefits. Indeed, if regulation is designed to encourage greater focus on long-term outcomes, then the rules governing how rates are set when multi-year plans expire become even more critical.

The basic idea between a regulatory menu is that the regulator would develop of series of alternative regulatory options, present these options to the utility, and the utility would be allowed to select the approach that best suits its circumstances. It is argued that a menu approach allows regulation to accommodate the diverse needs and circumstances of different utilities in an efficient manner. Some stakeholders argued for the adoption of a menu approach in Ontario's third generation incentive regulation plan for electricity distributors.

However, this proposal was rejected during the review process for two reasons. First, the menu proposal assumed that the PBR plan would include an ESM, which the Board did not approve. Second, and more fundamentally, there were significant unanswered questions regarding the design of the proposed menu. One basic concern is that it was never explained how permitting companies to choose from a menu would necessarily benefit *customers*. If companies are presented with a variety of regulatory options, they will clearly select the alternative that is expected to be most profitable. However, this is not sufficient for an appropriately-designed incentive plan, which should lead to "win-win" outcomes for companies and customers. It was never clear that the proposed menu would lead to win-win outcomes. This, and related design issues, will need to be addressed if regulatory menus are to receive serious consideration in Ontario as regulatory mechanisms.

Incentives mechanisms can also be targeted at achieving very specific regulatory outcomes. More targeted incentives may be more appropriate for rewarding utilities for achieving one-time, discrete regulatory objectives. Examples of such discrete regulatory objectives include quickly connecting a renewable electricity generator in a distributor's territory, or adding transmission capacity in a given area to eliminate transmission congestion. These are important regulatory outcomes that can be identified and targeted fairly easily. Incentive-based approaches focused on targeted regulatory outcomes would include incentive-based ROEs and project-specific capital structures.

In addition to incentive mechanisms, the Board can use discretion and judgment as a means of creating incentives for networks to achieve regulatory objectives. One example is the use of prudence reviews in cost of service proceedings. The Board can scrutinize networks' costs and operations more closely if certain performance measures fail to conform with established standards. For example, Staff can apply greater scrutiny to the cost of service applications of networks in the bottom third of the benchmarking evaluations that are currently used to set stretch factors in third generation incentive regulation. Alternatively, networks could be "fast tracked" for regulatory approvals if their measured performance satisfies certain standards. In the UK, Ofgem has established a kind of fast-tracking of the applications of companies if their rate applications are deemed to be "well justified."

Ultimately, Staff will need to recommend a framework that brings together performance measures, performance standards, performance evaluation techniques, and regulatory mechanisms in a manner that is most likely to promote desired regulatory outcomes. Moreover, this framework must be adaptable to the variety of regulatory approaches that the Board will continue to administer. The Paper presents some initial, and highly preliminary, ideas on how such a framework can be developed. These thoughts are primarily intended to facilitate discussion and prompt alternate ideas from other parties. Clearly, these evaluation issues are complex and require substantial input from stakeholders. It is hoped that the concrete performance evaluation/regulatory consequences framework presented in this document will stimulate more ideas and better informed alternative proposals than would result in the absence of such a "strawman" proposal.

A significant amount of work has been to date in Ontario which can serve as a foundation for developing a holistic and comprehensive performance evaluation framework.

Notwithstanding this work, fleshing out this framework would involve some challenges, including:

- The availability of capital stock and expenditures data in Ontario would have to be explored, to ascertain the quality and extent of available data
- Following from the capital data analysis, it would be necessary to explore the feasibility of developing long-run TFP trends, and total cost benchmarking models, for Ontario's electricity networks, and whether the quality of these TFP and total cost measures are sufficient for regulatory applications
- Stakeholders and Staff would need to consider the merits of alternative benchmarking approaches and whether and how the models currently used by Staff should be changed (other than changing the application of the models from O&M costs to total costs)
- Stakeholder and Staff should consider how best to define "appropriate" and "desirable" performance standards associated with different regulatory objectives
- Stakeholders and Staff should carefully consider the merits of rate reset provisions, and whether these mechanisms can play a useful role in promoting the Board's regulatory objectives
- Stakeholders and Staff should consider whether the available reliability data in Ontario are of sufficient quality to establish reliability performance standards
- There are significant challenges associated with developing performance measures, standards and evaluation techniques related to the Board's new duties under the GEA. This appears to be particularly true with respect to promoting efficient investment planning by networks, promoting timely and efficient connection of renewable generators, and encouraging the development of a smart grid. Our preliminary, "strawman" evaluation framework did reveal that a significant amount of additional work is necessary in these areas. This is perhaps not surprising, since these are in fact new challenges for the Board, so there is necessarily less of an informational or analytical record to rely on when considering how best to address these challenges.

2. Basic Definitions

This chapter will define and discuss some basic concepts that will be important during the Board's consultation. We discuss, in turn, outputs and outcomes; performance standards; productivity and efficiency; performance; and regulatory incentives.

2.1 Outputs and Outcomes

In general terms, "outputs" are the goods or services that firms provide to their customers. "Outcomes" are the end-states experienced by either customers or companies themselves after outputs have been provided. With respect to companies, outcomes can be changes in behavior, knowledge or functioning that result from transactions between companies and their customers. For customers, outcomes would reflect changes in their (subjective) utility. For example, in the case of an electricity distribution network, "delivered electricity" (measured in kWh) would be an example of an output. "Customer satisfaction" would be an example of an outcome for the customer.

The Board has recently discussed the importance of outcomes, and how this differs from its "traditonal output-based focus," in its most recent draft Business Plan, which says:

> Outcome-based performance measures can capture the Board's effect on the sector. The Board's traditional focus on (delivering outputs from specific) initiatives looks at how efficiently the Board operates. Outcome-based performance measures are sector indicators, and would be monitored over time to evaluate whether the intended outcome has been achieved. The achievement of these outcomes would then be the measure of the Board's performance. However, with many of the Board's initiatives, the outcomes may not be apparent for several years.
>
> The objective of the outcome-based approach is to establish a systematic framework to monitor and evaluate the effectiveness of Board activities. This assessment framework will take time to develop given the challenges involved with identifying measures to accurately assess the achievement of outcomes over the short, medium and long term. Until this framework is developed, the Board will continue to rely primarily on its traditional output-based approach to measure achievement.[2]

---

[2] Ontario Energy Board, *Draft 2011-2014 Business Plan*, January 4, 2011, p. 13.

The distinction is subtle, but important. Outputs can be successfully provided without evaluating whether the delivered output satisfactorily addresses end users‘ needs or desires. An outcomes-based approach is designed to be more responsive to end-users‘ needs and ensure that the regulated sector is providing “value for the money.“ In competitive markets, firms naturally have strong incentives to satisfy customers‘ preferences and not passively provide outputs to the marketplace, because if the firms did not their customers would ultimately take thier business elsewhere. These incentives are not as strong for regulated monopolies, or for regulatory agencies, since consumers of their services cannot choose among alternate service providers. An outcome-based regulatory approach is therefore compatible with the desire to be more pro-active and responsive to consumers‘ preferences, as in competitive markets.

In the current environment, it is important to recognize that energy networks are providing both “traditional“ and “new“ network services to their customers. The line between the traditional and the new, however, is not always clear. Below we discuss the relationship between energy networks‘ traditional and new functions, as well as the relationship between energy network outputs and networks‘ output quality.

2.1.1 Traditional Network Services

Traditionally, the main function of electricity transmission networks has been to move bulk power from generation stations to distribution or other high-volume delivery points. The traditional purpose of distribution networks has been to receive power in bulk from points on high-voltage transmission grids and distribute it to consumers in assigned territories. Delivery involves reducing the voltage of bulk power supplies to the levels used in end-use electrical equipment. To satisfy consumer demands, distributors construct and maintain power delivery networks that establish physical contact with almost every business and household in their service territory.

Because interruptions in power delivery are costly to customers, transmission and distribution utilities are expected to design and operate distribution networks to assure reliable deliveries. One important design requirement is that the capacity of the delivery system must be able to accommodate customers‘ peak demands. For transmission utilities, those are the demands at peak times at designated delivery points. For distributors, networks must have sufficient capacity to meet peak demands for all customers throughout the distributor's assigned territory.[3] Distributors must also endeavor to connect customers rapidly to the network. End use electrical equipment is also designed to operate within a narrow range of voltage levels. Thus, in addition to providing power supplies that are as continuous and uninterrupted as possible, distributors must attempt to conform to technical standards affecting the quality of power deliveries (*e.g.* regarding voltage, waveform, and harmonics).

In sum, both transmission and distribution utilities are expected to deliver power continuously at all points in time when their customers demand it. Outputs for all energy networks therefore include total delivered electricity (over a defined interval, such as a month) and electricity delivered on peak. In addition, distributors provide the service of connecting end-users to the electricity grid and extending the network to connect new customers in their assigned territories. Although some transmission networks may also deliver to end-users, these are typical few in number and do not vary substantially over time.

2.1.2 "New" Network Services

Under the Green Energy and Green Economy Act, three new objectives have been added to the OEB's traditional economic regulation taks: promoting energy conservation and demand management; facilitating the development of a smart grid; and encouraging electricity generation from renewable energy sources. Energy networks naturally play a critical role in realizing these objectives. Distributors must directly invest in "smart grid" technology, which is integrated into their existing networks. Both transmission and distribution networks must invest in additional infrastructure to connect renewable generators to the grid. This essentially involves the same "connection" and "peak demand" outputs that distributors have traditionally provided, but in the "new" marketplace energy networks will provide these outputs upstream (to renewable generators and related sources of electricity supplies) as well as downstream (to end-users).

It is also worth noting that many of the investments that networks must make to facilitate GEA mandates are also useful for traditonal distribution functions. For example,

[3] In practice, this means designing distribution networks to accommodate a diverse number of "local" peak demands throughout their service territory, which may not coincide at the same points in time.

smart grids or Advanced Metering Infrastructure (AMI) is critical to the energy marketplace of the future. At its most basic level, AMI is designed to automate the process for recording customers' power consumption, but it can also create a much wider array of benefits. AMI systems generally involve three interrelated components. The first is the metering units themselves, which are far more sophisticated than the "accumulation meters" that have essentially been in place since the industry's inception. The second is the information networks that are used to transmit data on customer consumption to the utility. Some AMI networks also allow data to flow in two directions, from the customer to the company and from the company to the customer. The third component is the meter data management system, where data on customer consumption and market conditions are stored and accessed.

AMI provides a number of benefits to energy distribution networks. Automated meter reading saves costs that would otherwise be incurred from manual meter reads. AMI can also provide "real time" information on the operation of the distribution system, which allows companies to locate faults that lead to power interruptions more quickly and accurately. In addition to enhancing the reliability of service provided to customers, better information on fault location can be used to optimize the size and dispatch of work crews, thereby reducing operating costs. AMI can also monitor the loading and condition of distribution system components, which can help companies optimize their inspection and maintenance cycles as well as extend the periods for replacing capital equipment. Automated meter reads also tend to improve billing accuracy and the timeliness with which bills are produced, thereby improving cash flow and the quality of billing service provided to customers.

In addition to providing these benefits for energy networks and their customers, more sophisticated metering systems will be increasingly necessary for distributors to cope with the more diverse and "distributed" (*i.e.* less centralized) nature of new generation technologies. Nearly all distribution systems are "radial" or designed for power to flow in one direction (from the bulk transmission system to the end user). Distributed generation (DG) units that are connected to the distribution network can lead to power flows in more than one direction, potentially decreasing the stability of electrical systems. This can affect the extent to which connected loads and generators interact with each other and, particularly when outages occur, the presence of DG units can lead to broader system instabilities. DG can also complicate the restoration of service whenever faults on distribution lines occur.

AMI is critical for helping distributors cope with these challenges. "Real time" information on the loading of distribution system components can be critical for monitoring the impact of DG units on the stability of the overall distribution system and for efficiently dispatching a portfolio of renewable (including wind) and distributed generators. Distribution AMI investments are therefore an important and increasingly essential complement to the renewable and DG units that are becoming more prominent in the energy marketplace.

Conservation and demand management (CDM) is also an important part of the GEA. Policymakers want consumers to respond naturally to the price signals from the marketplace *e.g.* by reducing consumption during peak hours when energy prices are typically highest. Lower demand pressures at the peak will tend to reduce energy prices and greenhouse gas (GHG) emissions, since energy and line losses are usually greatest during peak hours. Lower peak demands can lead to less energy consumption and reduced GHG emissions and more efficient use of network infrastructure as energy use is shifted from peak to non-peak hours.

As discussed, transmission and distribution (and power generation) infrastructure must all be sized to accommodate peak demands, so reducing peak usage will tend to defer the need for "traditional" energy infrastructure investments. Pushing energy investments into the future saves costs and also increases the probability that R&D devoted to cleaner generation technologies will have come to fruition and can be used when investments are ultimately required. Effective CDM can therefore contribute to a cleaner and lower-cost efficient energy supply and delivery system both now and in the future.

AMI is critical for ensuring optimal CDM. Two-way AMI communication systems can relay price signals in real time from the marketplace back to consumers. Visual displays can let customers know the prices they are paying for power being used in their homes and businesses at that moment, and this information can be used to adjust their consumption accordingly. CDM can be further enhanced if automated direct load control (DLC) devices are installed on customer premises. DLC devices can be programmed to slow consumption (*e.g.* through less frequent cycling of air conditioning units) or eliminate it entirely when power prices hit established thresholds. Automated demand response of this type can be a very effective tool for disciplining the energy marketplace, reducing greenhouse gases and enhancing overall efficiency, but more sophisticated and expensive AMI systems are necessary for achieving these benefits.

The increasing importance of DG and its relationship to AMI has already been discussed, but the relationship between DG and network infrastructure is also complex. DG units can provide voltage control and ancillary services such as spinning reserves that can help networks manage system stability. Energy networks can therefore benefit directly from owning, operating and dispatching DG units, and Ontario's distribution system code has in fact been amended to allow networks to own renewable generation assets.[4]

It should also be recognized that DG can serve as a substitute for energy network investments. Because DG is located closer to customer loads than more centralized generation sources, the need for transportation capacity to move power from supply to demand points is reduced. Networks can therefore use DG to avoid or defer the investments that would otherwise be needed to augment energy transportation capacity. Locating generation closer to end uses also reduces line losses and the energy that must be generated to meet final demands, thereby contributing to lower GHG emissions. Greater reliance on DG also reduces the need for, and defers investment in, larger generation stations, which again increases the probability that cleaner technologies will be utilized when those investments are ultimately made. All of these factors demonstrate that DG can be an important "input" into network operations, with positive benefits in terms of operational flexibility and promoting energy market objectives. Networks should therefore in principle consider DG when evaluating investment choices.

These inter-relationships suggest that there is not a bright line between networks' "traditional" and "new" functions. Some of the investments necessary to comply with GEA mandates may have implications for how traditional outputs are provided, while some assets that help perform traditional functions more efficiently (*e.g.* smart meters) may prove valuable in helping networks cope with the challenges of delivering power from more diverse and less centralized supply sources to end-users. Regulators may therefore need to take a broader view of how traditonal outputs are being provided, and be sensitive to the potential linkages between investments needed for the "new" marketplace and the network outputs that have traditionally been subject to economic regulation.

---

[4] See *Amendments to the Distribution System Code and Affiliate Relationships Code for Electricity Distributors and Transmitters*, Board File No EB-2009-411, March 11, 2010.

2.1.3 Outputs and Output Quality

It may be argued that because maintaining reliability is such an important function for electricity networks, reliability (and perhaps other aspects of service quality) should also be viewed as a network output. However, economists typically view quality as an attribute of any given good or service, rather than a separate product in and of itself. In fact, economic analyses of product quality typically view products as a bundle of attributes or characteristics. Each characteristic is desirable in the sense that it satisfies consumer tastes and preferences. Since all characteristics are valuable to consumers, consumers generally prefer 'more' of each desirable attribute rather than less.

However, higher quality comes at a price. It is typically costly to add quality characteristics to a product or to provide 'more' of any given attribute. In competitive markets, the amount and number of quality attributes that firms choose to bundle with their products is ultimately limited by consumers' willingness to pay. Economists therefore believe that each quality attribute carries an implicit price that, in turn, is reflected in the overall price of the product or service in the marketplace.[5]

This implies that while aspects of network service quality are not outputs in and of themselves, the reliability and quality of network outputs is nevertheless a large source of the value of network outputs to customers. It is therefore important for networks to provide appropriate levels of service quality and reliability. Performance metrics and incentive mechanisms can be valuable tools for ensuring that network quality is maintained or, depending on customers' valuations of incremental quality improvements, improved. These performance metrics, standards and incentive mechanisms will also be separate from those designed to encourage utilities to provide outputs in a cost-effective manner. This follows

[5] The implicit prices for various quality attributes can be quantified through statistical methods and aggregated in so-called hedonic price indexes that summarize overall quality differences between products. Clearly, quality attributes are rarely priced explicitly in the marketplace, but it does not follow that the estimation and use of hedonic prices is simply an academic exercise. One example where these economic concepts are applied is by the Bureau of Labor Statistics of the US Department of Labor, which computes hedonic prices indices and adjusts for changes in the quality of some products when it computes the U.S. Consumer Price Index (CPI). For example, CPI calculations control for quality changes in personal computers. The quality of PCs has been increasing at the same time that their prices have fallen. The real decline in PC prices is, therefore, even greater than reflected in their list prices, since consumers are getting more for their money. Alternatively, if a firm were to offer a new PC that had quality levels equal to those of a PC ten years ago, it would certainly fetch a lower price than the higher-quality new models that are available. Hedonic price indexes adjust PC prices so that they reflect the price declines associated with a PC of constant quality.

from the fact that cost and quality are positively related (*i.e.* higher quality comes at higher cost, and lower cost can lead to lower quality). Any regulatory mechanism designed to promote lower cost can therefore unintentionally encourage quality reductions unless there are countervailing regulations or mechanisms focused on maintaining quality.

2.2 Standards and Performance

A standard is a quantitative benchmark that is used to evaluate how effectively regulatory objectives are being achieved. Standards can be established at different levels. For example, there may be a "minimum standard," which reflects the minimum acceptable level of achievement of a given objective by any regulated network. Any utility that falls below a minimum standard could invite regulatory scrutiny or sanction, possibly including financial penalties. There may also be an "average standard" which reflects an expectation of how effectively, on average, a network industry is expected to achieve an objective. There may also be a "superior" standard, which could lead to rewards for utilities that achieve at levels beyond this standard.

Performance refers to how effectively a network is achieving the desired regulatory objective. This notion of performance is obviously closely related to the concept of a standard. A standard is what is used to judge actions or outcomes that are expected to contribute towards regulatory objectives. Performance reflects the measure of how effectively utilities are in fact achieving regulatory standards.

2.3 Productivity and Efficiency

2.3.1 Productivity

Productivity measures the transformation of inputs into outputs. In the present context, "inputs" refer to the resources an energy network procures in order to provide network outputs. Total factor productivity (TFP) measures the relationship between all the outputs provided by a utility and all the inputs that the utility procured to provide those outputs. Partial factor productivity (PFP) measures the relationship between the utility's comprehensive output and a more narrow measure of inputs. For example, labor productivity would measure the productivity of a utility with respect to its use of labor inputs only.

In most utility applications, TFP and PFP are measured with indexes that aggregate several types of output and inputs into comprehensive output quantity and input quantity metrics. Each dimension of output quantity and input quantity is measured by what is sometimes referred to as a subindex. A TFP *level* index is defined as the ratio of an output quantity index to a comprehensive input quantity index.

$$TFP = \frac{Output\ Quantities}{Input\ Quantities}. \quad [1]$$

TFP therefore represents a comprehensive measure of the extent to which firms convert inputs into outputs. Comparisons can be made between firms at a point in time or for the same firm (or group of firms) at different points in time. The latter metric is a measure of TFP growth, and the trend in a TFP index is the difference between the trends in the component output quantity and input quantity indexes.

$$trend\ TFP = trend\ Output\ Quantities - trend\ Input\ Quantities. \quad [2]$$

The measures for PFP are analogous. A PFP level index is defined as the ratio of an index of comprehensive output quantity to an input quantity subindex, such as an index of O&M input. In this example, the growth in O&M PFP would be equal to the growth in comprehensive output quantity minus the growth in O&M input quantity. Mathematically, it can be shown that the growth in TFP can be decomposed into a weighted average of the growth in PFP for the different inputs used in production.

TFP grows when the output quantity index rises more rapidly (or falls less rapidly) than the input quantity index. TFP can rise or fall in a given year but in most industries in industrial countries typically trends upward over time. Over shorter periods, however, TFP growth is more volatile and can certainly decline over a given, multi-year period.

Several factors contribute to TFP growth. In Appendix One, we present a mathematical derivation which shows that TFP can be decomposed into six different components. Although this analysis is somewhat technical, it is useful for understanding the relationship between total factor productivity and efficiency, as well as other concepts that are often discussed when discussing the cost structure of energy networks. This will become more clear after the concept of efficiency is defined below.

2.3.2 Efficiency

Economists typically distinguish between two types of efficiency: productive efficiency and allocative efficiency. For regulated industries, regulatory cost can also be considered as a dimension of efficiency. Regulatory costs reflect the costs of administering the regulatory regime in which the utilities operate, rather than the efficiency of the regulated firms *per se*.[6] In practice, however, most regulatory costs are typically recovered from regulated ratepayers and therefore certainly relevant when evaluating the impact of alternative regulatory arrangements on stakeholders' welfare.

Productive efficiency refers to the degree to which a firm produces the maximum potential output given available technologies. If a firm is not productively efficient, it will be using an excessive amount of inputs to produce the outputs it provides to its customers. This naturally means that the firm will be producing above the minimum cost of service for its output level.

Minimum production costs depend on choices for both "variable" and "fixed" inputs. Variable inputs are resources whose levels can be adjusted quickly at a given point in time. In contrast, rapid adjustments in fixed inputs are either not feasible or are prohibitively expensive. In the short run, productive efficiency depends primarily on meeting demand with a minimum-cost mix of variable inputs. This is sometimes referred to as "static" productive efficiency. In the longer run, all inputs are variable and the cost-effective use of capital equipment is also a central efficiency concern. This is sometimes termed "dynamic" productive efficiency.

There are both supply- and demand-side dimensions of allocative efficiency. Demand-side allocative efficiency concerns how efficiently a firm is pricing its goods or services. In the short-run, prices become allocatively efficient to the extent that they reflect the firm's marginal cost of service.[7] [8] In the longer run, updating the firm's mix of services to meet evolving consumer preferences becomes an important concern.

---

6 Regulatory cost can be considered a specific type of "transaction cost," which are increasingly recognized as critical in how institutions and economies are structured and evolve over time.

7 This implicitly includes the marginal cost of the quality attributes that are bundled within the product.

8 In regulated industries, it is well-known that marginal cost pricing will not typically allow a utility to recover its entire cost of service. When this is true, certain tariff structures (Ramsey pricing, or with multiple billing determinants a Coase Tariff) can allow firms to recover their costs with a mimimum departure from the consumption levels that would occur with marginal cost pricing.

Supply-side allocative efficiency refers to using the optimal *mix* of inputs in production, for given levels of input prices. For example, if the marginal productivity of labor relative to the price of labor is greater than the marginal productivity of capital relative to the price of capital, firms could produce the same amount of output at lower cost by using relatively more labor and less capital. It is sometimes argued that under cost of service regulation, utilities in fact have incentives to use overly capital-intensive production techniques.[9] [10]

2.3.3 The Relationship Between Productivity and Efficiency

With these efficiency concepts defined, it is instructive to return to the decomposition of TFP growth that was discussed at the end of Section 2.3.1. Appendix One shows that the growth rate in TFP has been decomposed into six terms. The first is the **scale economy effect**. Economies of scale are realized if, when all other variables are held constant, changes in output quantities lead to reductions in the unit cost of production.[11]

The second term is the **nonmarginal cost pricing effect**. This is equal to the difference between the growth rates of two output quantity indexes. The first is the index used to compute TFP growth, which in ratemaking applications should be constructed by weighting each output by its share of regulated revenue. Hence, the first term is the growth in an output quantity index constructed using revenue weights. The other output quantity index, denoted by $\dot{Y}^{\varepsilon}$, is constructed using cost elasticity weights. It can be shown that revenue weights will differ from cost elasticity weights if prices are not proportional to marginal costs. Accordingly, this term is interpreted as the effect on TFP growth resulting from departures from marginal cost pricing.[12]

---

[9] This is sometimes referred to as the Averch-Johnson effect.

[10] Regarding the final efficiency dimension, costs are incurred in utility regulation. These include, most obviously, the resources (*e.g.* accountants, lawyers, data collection/collation and consultation) that are dedicated to the regulatory process. Senior company officials are also drawn into the regulatory arena on a part-time basis. This can divert management attention from market developments, and performance may suffer as a result. Reduced regulatory costs therefore increase the efficiency of the regulatory process and can encourage behavior that further promotes productive and allocative efficiency.

[11] Technically, this will be the case if the sum of the cost elasticities with respect to the output variables in the cost function is less than one.

[12] See Denny, Fuss and Waverman, p. 197.

The third term is the **cost share effect**. This measures the impact on TFP growth of differences in the growth of input price indexes based on optimal and actual cost shares. This term will have a non-zero value if the firm does not utilize the optimal input mix.

The fourth term is the **Z variable effect.** It reflects the impact on TFP growth of changes in the values of variables in the Company's service territory or broader institutional environment (*e.g.* regulatory or government policy) that are beyond management control.

The fifth term is **technological change**. It measures the effect on TFP growth of a proportional shift in the cost function. A downward shift in the cost function is equivalent to an "increase" or improvement in technology.

The sixth term is the **productive inefficiency effect**. This measures the effect on TFP growth of a change in the firm's productive inefficiency factor. Firms decrease their productive inefficiency as they approach the cost frontier, which represents the lowest cost attainable for given values of output quantities, input prices, and other business conditions.

Three of these six components of TFP growth reflect changes in efficiency, and three do not. The three components that measure changes in efficiency are: 1) the non-marginal cost pricing effect, which reflects changes in demand-side allocative efficiency; 2) the cost share effect, which reflects changes in supply-side allocative efficiency; and 2) the productive inefficiency effect, which reflects changes in productive efficiency.

The technological change effect reflects changes in the technology available to the firm but not how efficiently the firm is producing given that technology. This is consistent with the view that technological change is typically exogenous to the energy network itself.[13] Accordingly, this component of TFP growth cannot reflect changes in the firm's efficiency *per se*, since efficiency must reflect endogenous behavior on the part of the firm.

The Z factor component by definition reflects impacts on the firm's TFP that result from factors outside its control. This includes factors within its defined service territory or the broader regulatory/institutional environment. An example of a broad institutional change

---

[13] This view is not appropriate for all companies, particularly firms like Apple or Boeing that engage in extensive research and development (R&D) and are accordingly responsible for significant technological innovations within their respective industries. Nevertheless, the assumption that technological change is exogenous appears valid for energy networks. Even though energy utilities sometimes provide funding for institutes that promote technological change within their industries, utilties themselve rarely undertake R&D designed to enhance their own energy delivery capability.

impacting measured TFP could be the additional investment costs networks must incur to comply with a specific government policy directive. An example of a factor within the service territory that can impact measured TFP growth is a movement towards less dense development patterns. If development with the territory becomes more spatially dispersed, networks would need to construct more infrastructure assets to connect new customers compared with the assets that were previously installed to connect existing customers. New connections therefore require more inputs relative to outputs than in the past, but utilities have no control over the resulting negative impact on their TFP since they have an obligation to provide service within their territory.

Similarly, if an increase in the number of customers served leads utilities to realize scale economies, the increase in measured TFP reflects factors beyond company control. Such scale economy effects simply reflect the underlying characteristics of the technology, and the fact that the incremental cost of serving an additional customer can be less than the utility's average cost of serving the existing customer base. When this occurs, the average cost of production falls whenever output (which utilities have an obligation to serve) within a given territory expands. This is an exogenous effect rather than an increase in efficiency resulting from endogenous, deliberate choices on the part of the network to reduce unit costs.

This discussion shows that productivity is a broader concept than efficiency. Certainly, improved productive and allocative efficiency can contribute to TFP growth. However, TFP growth can also result from factors beyond the network's control, including broader technological change, the realization of scale economies, and Z-factor effects.

Interestingly, TFP and efficiency can also move in opposite directions. Consider a utility experiencing rapid growth and thereby realizing considerable scale economies, which simultaneously devotes less effort to managing costs and thereby becomes less productively efficient. Even though this firm's efficiency has declined, it's measured productivity would increase if the magnitude of the scale economy effect was greater than the magnitude of productive inefficiency effect. Alternatively, consider a firm that successfully improves it's productive efficiency at the same time that it must comply with a costly new government mandate. If the improvement in the firm's productive and allocative efficiency is less than the magnitude of the Z-factor effect, it's measured TFP will decline in spite of its improved efficiency performance.

2.4 Regulatory Incentives

It is also important to define the concept of "incentives" carefully, particularly where infrastructure investment is concerned. This is evident from Staff's June 2009 Discussion Paper on the Regulatory Treatment of Infrastructure Investment for energy networks, particularly in the "Meaning of 'Incentive'" section on pp. 15-16. This part of the Paper discusses a number of regulatory "incentives" that the US Federal Energy Regulatory Commission (in Order 679) made available to utilities in order to promote infrastructure investment. Staff writes:

> FERC Order 679 identifies seven "incentives" that are available to all jurisdictional public utilities: the ROE adder, allowing CWIP in rate base prior to the asset coming into service, the hypothetical capital structure, accelerated depreciation, recovery of costs of abandoned facilities, deferred cost recovery, and single-issue ratemaking. While for the purposes of this paper, staff has distinguished classes of mechanisms differently than FERC, staff shares FERC's concern as explained in its Order 679 regarding the meaning of "incentives" in the context of infrastructure development. The alternative mechanisms described in this Discussion Paper are intended to provide "incentives" to construct appropriate infrastructure, but they do not constitute an "incentive" in the sense of a "bonus" for good behavior. Rather, if adopted by the Board, each should be applied in a manner that is rationally tailored to the risks and challenges faced in constructing the infrastructure.[14]

It is clear from this passage that Staff is referencing seven specific FERC "incentives," although FERC uses this term with reservations. Staff emphasizes that it shares FERC's concern over the meaning of regulatory "incentives" for infrastructure development. Staff then distinguishes "alternative mechanisms" from "incentive mechanisms" in the balance of the Paper.

As indicated above, "alternative mechanisms" are intended to encourage the construction of appropriate infrastructure but not provide a "bonus" for good behavior. Examples of alternative mechanisms are accelerated cost recovery provisions such as including construction work in progress (CWIP) in rate base before the asset comes into service and "contract term" depreciation, which allows depreciation of assets over shorter asset lives and therefore enhanced cash flow for utilities. These mechanisms are distinct from

---

[14] *Staff Discussion Paper on the Regulatory Treatment of Infrastructure Investment for Ontario's Electricity Transmitters and Distributors*, EB-2009-0152, June 5, 2009, p. 15-16.

two FERC options that Staff singles out in the Discussion Paper: 1) an incentive-based ROE and 2) project-specific capital structures, for new investments that benefit customers.[15] Staff writes that "(f)or the purposes of this Discussion Paper, staff has classified these two mechanisms in particular as "incentives" because they provide "cost plus" compensation to the regulated entity for its investment."[16]

In the present context, we will define "incentives" as *ex ante* regulatory rules that: 1) encourage behavior by utilities that promotes desired regulatory outcomes; and 2) if executed successfully, will lead to financial benefits for utilities. This definition appears consistent in spirit with Staff's distinction between "alternative" and "incentive" mechanisms, since the former do not involve an explicit linkage between financial rewards and "good behavior" *i.e.* behavior expressly designed to promote desired regulatory outcomes. It should also be noted that regulatory incentives provide only an opportunity for financial rewards, not a guarantee. The realization and magnitude of such rewards depends on the utility's performance, or how successfully its behavior promotes the desired regulatory outcome(s).

---

[15] In Order 679, FERC explicitly allows an incentive-based ROE for "new investments in transmission facilities that benefit by ensuring reliability or reducing the cost of delivered power by reducing transmission congestion." FERC focuses on these benefits from transmission-related investments since FERC regulates transmission and not distribution.

[16] Staff Discussion Paper, *op cit*, p. 25.

3. Measurement

Practical utility regulation obviously requires measured variables for output, costs, productivity, and efficiency. This chapter will consider appropriate measurement of many of the key concepts defined in Chapter Two. We consider, in turn, measurement of outputs; costs and inputs; productivity; and efficiency.

3.1 Outputs

The appropriate measures of network outputs can differ somewhat depending on the application. When setting rates in either cost of service proceedings or multi-year, inflation minus X formulas, the appropriate output choices are clearly the utility's billing determinants *i.e.* the outputs for which customers are explicitly billed. When selecting outputs for cost comparisons, however, there is more room for flexibility. We begin by discussing the selection of appropriate output measures for setting rates and then turn to output measures in cost analyses.

3.1.1 Appropriate Output Measures in Setting Rates

Nearly all TFP measures used in ratemaking use utilities' billing determinants to measure outputs. Recently, some jurisdictions have debated whether "unbilled" activities provided by network utilities should also be used to measure outputs in TFP studies. An example of such an unbilled output may be energy security, or redundancy in the delivery network. It has been argued that these is an important network function that imposes costs on utilities, so excluding it from TFP analyses will lead to biased TFP estimates.[17]

While this argument may have some surface appeal, its weakness becomes apparent on closer inspection. This is perhaps easiest to see under cost of service regulation. "Unbilled" activities like security are no less important when utilities are regulated by cost of service methods than when rates are set by indexing formulas that utilize information on industry TFP growth. It also remains essential for these costs to be recovered in regulated

---

[17] See Economic Insights (2009), *Total Factor Index Specification Issues*, Report to the Australian Energy Market Commission.

prices that are established. How is this done in cost of service regulation? The costs of providing security are included in the utility's overall cost of service. Prices for the different billing determinants of the regulated utility are then set so that, in aggregate, they recover the costs of these and other activities. Energy security is therefore an unbilled activity in cost of service regulation, but the costs of providing energy security are still recovered through billed outputs *i.e.* billing determinants.

Indeed, there is no other way to recover the costs of 'unbilled' activities than through the prices charged for billed outputs. Costs are recovered through the revenues earned from customers, and revenues can only be earned from billed outputs. The issue of "unbilled activities" is therefore irrelevant in cost of service regulation. There are many utility activities for which customers are not billed explicitly, yet the costs of these activities will be recovered from the prices charged on the outputs that *are* billed. It follows that the effective, measured outputs to be used when setting cost of service-based rates are the utility's billing determinants, not any activities for which customers are not billed.

The same is also true when measured TFP trends are used to set the terms of multi-year adjustment formulas. The initial prices at the outset of an incentive regulation plan are determined through cost of service regulation, where initial prices reflect the costs of energy security and other unbilled activities. The indexing formula adjusts the change in these prices. Any change in the costs of energy security and other "unbilled" outputs are necessarily recovered through the prices charged for the billed outputs. The **only** change in output quantities that can recover the costs of these activities are billed outputs. These are the outputs that, accordingly, should be reflected in the measure of industry TFP which are used to set price changes.

This intuitive result is confirmed in the mathematical logic that underpins the development of index-based regulation. The indexing logic relies on what is sometimes referred to as the competitive market paradigm *i.e.* that utility tariff adjustments should be set at a rate that is consistent with how prices evolve in competitive markets. The indexing logic therefore examines long-run changes in revenues and costs for an industry. In the long run, the trend in revenue (R) for an industry equals the trend in its cost (C).

$$Trend\ R = Trend\ C \tag{3}$$

The trend in the revenue of any industry will be equal to the sum of trends in revenue-weighted output price indexes (P) and revenue-weighted output quantity indexes (Y).

$$Trend\ R = Trend\ P + Trend\ Y \tag{4}$$

The growth rate in the cost incurred by an industry is the sum of the trends in a cost share-weighted input price index (W) and a cost-share weighted input quantity index (X).

$$Trend\ C = Trend\ W + Trend\ X \tag{5}$$

Substituting (4) and (5) into equation (3) and rearranging, we find

$$\begin{aligned} Trend\ P &= (Trend\ W + Trend\ X) - Trend\ Y \\ &= Trend\ W - (Trend\ Y - Trend\ X) \\ &= Trend\ W - Trend\ TFP \end{aligned} \tag{6}$$

This is the basic result of the indexing logic. It shows that the change in an industry output price index can be decomposed into changes in the industry's input price index minus changes in its TFP index. When this result is applied to utility regulation, it implies that allowed changes in utility prices (the left-hand side variable in (4)) can be linked to industry input price inflation minus changes in industry TFP. If the chosen inflation factor (such as the CPI) is a good proxy for long-run trends in industry input prices, then it is appropriate to calibrate the X factor using an estimate of the trend in the regulated industry's TFP.

In this indexing logic, the measured output term first appears in equation (4). This equation shows that the change in revenue can be decomposed into a change in output prices and output quantities. The change in output quantities in this equation is the same output quantity trend that appears in the TFP trend measure in equation (6). Equation (4) therefore draws a direct link between the outputs that are used to measure TFP and revenues. In other words, the outputs that are used in the TFP measure *must* have a direct link to the revenues of the regulated industry. If this was not the case, then the index decomposition in (4) – which gives rise to the output quantity index used in the TFP measure – would not be satisfied.[18]

More specifically, equation (4) has two direct implications for the output quantity specification. One is that the specific output quantities that are used to compute the output

[18] More technically, equation (4) says that the revenue, price and output quantity indexes that are used in TFP-based regulation must satisfy what is known as the product test.

quantity index must be the billing determinants that are used in the tariffs for the regulated sector. No other output quantity measures can be compatible with equation (4) in the logic above. The second implication of equation (4) is that each billing determinant should be weighted by its revenue share when computing the output quantity index. Again, this is necessary for the changes in revenues to be decomposable into changes in output prices and output quantities. If output quantities were weighted by anything other than each output's share of revenues, equation (4) would not be satisfied (except by chance).

This discussion shows that, when setting regulated rates under any ratemaking approach, network outputs should be measured by the network's billing determinants. For most residential and small commercial customers, these billing determinants are the customer counts (to which the monthly customer charge is applied) and kWh deliveries, perhaps differentiated by the amount of kWh delivered or time of electricity use. Many larger commercial and industrial customers also have charges for peak demand, so their measured kW peak demands should also ideally be used. If indexing techniques are used to measure TFP, these outputs should be aggregated into a comprehensive index of output quantity using each output's share of regulated revenue.

In practice, it is often difficult to satisfy these criteria. For example, the reported kW data of vertically integrated utilities may reflect power flows and sales for resale on the transmission network rather than deliveries to end-users on the distribution network. Data on revenue shares are often also difficult to obtain. When this is the case, changes in kWh are often taken as a proxy for changes in kW, and cost elasticity shares that are estimated econometrically are a second-best estimate of revenue shares.

3.1.2 Output Measures for Cost Analysis

When undertaking analysis of network cost drivers, or making cost comparisons among different energy networks, output measures should be selected on how strongly they correlate with the cost of providing regulated services. Billing determinants like customer counts will still be important output measures in these analyses, but they are not the only potential measures. Any variable that can be defined as an output and impacts network costs is appropriate when undertaking cost comparisons.

One important output measure that can be considered in cost applications is the spatial distribution of end-users in a distributor's territory. Power distribution networks transport electricity directly into the premises of end-users, and the location of those end users directly impacts the infrastructure needed for delivery and, in turn, distribution costs. Distributors serving largely rural territories typically have lower customer densities than distributors serving urban areas; all else equal, the delivery assets per customer are therefore greater for more ruralized territories. In applied cost analyses, the spatial distribution of customers is sometimes proxied by the total circuit km of distribution line, or the total square km of territory served. Provided customer numbers is also used as a cost measure, either of these additional variables will reflect the impact of different levels of customer density within a territory on electricity distribution costs.

Peak demand is another variable that can be included as a output measure in cost analysis. Of course, peak demand is a billing determinant for some customers, but peak demand will also be an important cost driver for smaller customers whose peak demands are not metered (and for which peak kW deliveries is therefore not a billing determinant). The reason is that delivery systems must be sized to accommodate peak demands, so there is a direct relationship between customers' peak demands and the costs of the necessary power delivery infrastructure.

It can also be appropriate to consider measures that reflect outputs that are provided by some networks but not others in the samples used for cost analyses. For example, distributors can differ in the extent to which they provide subtransmission service. Some distributors have no subtransmission assets, while others have a significant amount of subtransmission capital. The latter distributors are effectively providing more delivery services, since they are are undertaking functions that the transmission utility is providing on behalf of the former set of companies. Ideally a cost study would control for this difference to ensure that "apples to apples" comparisons are made between networks.

Cost analyses can also include variables that reflect the quality of the service provided. Examples include reliability measures of SAIFI or SAIDI, or measures that may be available of energy security. As discussed in Chapter Two, these are not output measures *per se*, but they are clearly quality attributes of the network services that are being provided.

Providing “more” of these attributes is costly, so it would be appropriate to include accurate measures of reliability or security in cost analyses.

3.1.3 Measuring Quality

There are many facets of the quality of service provided by energy networks. We believe the measured service quality indicators used in network regulation should ideally satisfy four, common sense criteria:

- they should be related to the aspects of service that customers value;
- they should focus on monopoly services;
- utilities should be able to affect the measured reliability; and
- the indicators should be sensitive to “pockets” of system quality problems.

First, indicators should be linked to aspects of utility service that customers actually value. This may seem obvious, but a strict application of this criterion excludes indicators that have been included in some plans. For instance, the reliability of service delivered to customers is an appropriate service quality indicator while tree trimming expenses generally is not.

Second, indicators should focus on the quality of the activities for which there are few if any alternative suppliers. This is consistent with the principle that regulation, including regulation of service reliability, is less necessary in competitive markets. Market forces are likely to create acceptable quality levels when products are available from multiple providers.

Third, utilities should be able to influence measured quality through their own behavior. It is nonsensical to evaluate a company’s reliability performance using indicators that are largely or entirely unrelated to management actions. If random or unforeseen incidents can affect important quality dimensions, the impact of these events should ideally be eliminated from the indicators.

Fourth, it is often sensible to have indicators that are measured on less than a system-wide basis. This is because system-wide measures may mask persistent service quality problems for “pockets” of customers. An example may be circuit reliability performance standards.

Overall, the choices for reliability indicators should balance the needs of comprehensiveness and simplicity. The selected indicators should not focus on some areas while ignoring other service quality attributes that are important to customers, because performance may deteriorate in the non–targeted areas. Comprehensiveness can be achieved simply by adding indicators to a plan. However, regulatory costs also rise as the regulatory plan includes more indicators since more utility and regulator resources must be devoted to measuring, reporting, and monitoring the selected indicators.

For energy networks, the most important aspect of service quality is undoubtedly reliability, or the continuity of the basic power delivery service. Electric utilities are expected to provide a continuous power supply at all times, so interruptions in power supply constitute a diminution in service quality. Reliability is often measured by the frequency and duration of power interruptions. Reliability is most often measured at the level of the entire system, although it can also be measured for subsets of the network such as for operating areas or specific circuits. The most typical measures used in utility regulation are:

- the System Average Interruption Frequency Index (SAIFI), or the number of sustained interruptions that is experienced annually by an average customer on the system
- the System Average Interruption Duration Index (SAIDI), or the number of minutes of sustained power interruptions that are experienced annually by an average customer on the system
- the Customer Average Interruption Duration Index (CAIDI), or the average duration of a sustained interruption experienced annually by a customer on the system[19]
- the Momentary Average Interruption Frequency Index (MAIFI), or the number of momentary interruptions that is experienced annually by an average customer on the system

[19] SAIDI is equal to the product of SAIFI and CAIDI, so if any two of these indicators are measured the third can be computed.

There are also analogues of each of the reliability measures above for subsets of the network. An example might be a "circuit SAIFI," which measures the number of annual outages experienced by an average customer on a specific circuit. Reliability indicators can also focus on thresholds for restoring power to customers.

Although these industry metrics are well-known, there can be significant differences in how these indicators are measured across networks. For example, the definition of "sustained" and "momentary" outages differs among utilities, but in most cases a sustained outage is either one that lasts at least one minute or five minutes. A momentary outage is any loss of power experienced by a customer that is not "sustained."

These service reliability metrics must generally be collected directly within the utility itself. There is considerable variation in how reliability measures such as SAIFI and SAIDI are defined and calculated across utilities. Sources of difference include:

- *Which interruption events are excluded from the metrics* Utilities can differ in which outages are included or excluded from SAIFI and SAIDI statistics. For example, four Australian jurisdictions (the Australian Capital Territory, New South Wales, South Australia, Victoria) exclude planned outages while it is rare for planned outages to be excluded in Canada, the US, and Europe. In Ontario, planned outages are not excluded from the metrics. Some electric utilities are still vertically-integrated, and their reliability measures will include generation, transmission and distribution outages, while others are stand-alone distributors and their outages reflect only outages at the distribution level. While vertically-integrated reliability measures can be separated into those resulting from the generation, transmission and distribution systems (and most are usually distribution-related), a failure to do so will lead to inherently misleading comparisons among some utilities' reliability measures.

  The largest source of discrepancies in outage exclusions across utilities concerns major event days. In Ontario, there are no standardized rules for excluding major events from reported reliability metrics, and some LDCs do not exclude any events. In most jurisdictions, however, nearly all utilities exclude these events from recorded reliability statistics because major events and storms are atypical and idiosyncratic, so including them can lead to a distorted perception of the utility's underlying reliability performance. However, utilities have adopted different definitions of what qualifies as "major" or "catastrophic" events. One traditional approach that has been adopted in a number of jurisdictions is to define any event as exceptional if at leads to interruptions for at least 10% of customers on the system. Any such widespread outage would accordingly be "normalized" out of reported, system-wide reliability indicators. This standard currently applies to the measured reliability of Maritime Electric, San Diego Gas & Electric, Kansas utilities, and Pennsylvania utilities, among others.

In 2002-2003, there was an effort by the Institute for Electrical and Electronic Engineers ("IEEE") to standardize the definition of major event days across utilities. This culminated in IEEE Standard 1366, which is sometimes referred to as the "Beta Method."[20] This standard has been promulgated worldwide, and an increasing number of utilities are adopting it as a basis for their officially reported reliability statistics. This standard does lead to greater comparability of reliability

---

[20] The main steps for identifying an major event day under Standard 1366 are the following:

- A major event day is a day in which daily SAIDI exceeds a threshold value $T_{MED.}$
- In calculating daily SAIDI, interruption durations that extend into subsequent days are assigned to the day on which the interruption begins. This technique ties the customer-minutes of interruption to the instigating events.
- The major event day identification threshold value T_{MED} is calculated at the end of each reporting period for use during the next reporting period. For utilities that have six years of reliability data, the first five are used to determine T_{MED} and that threshold is applied during the sixth year.
- The methodology for calculating T_{MED} is as follows:
  - Values of daily SAIDI for a number of sequential years, ending on the last day of the last complete reporting period, are collected.
  - If any day in the data set has a value of zero for SAIDI, those SAIDI data are excluded from the analysis.
  - The natural logarithm of each daily SAIDI value in the data set is calculated.
  - The average of the logarithms, α , of the data set is calculated.
  - The standard deviation of the logarithms, β , of the data set is calculated.
  - The major event day threshold, T_{MED}, is calculated by using the equation (this value should in theory give an average of 2.3 major event days per year)

    $$T_{MED} = e^{\alpha + 2.5\beta}$$

  - Any day with daily SAIDI greater than the threshold value T_{MED} is designated a major event day, and data for this day is removed from SAIFI and SAIDI performance to provide a "normalized" measure of performance.

statistics among utilities, but there are still a number of factors that can lead to differences in reliability measures.

- *Step restoration* When utilities restore power after widespread outages, restoration typically proceeds in "steps," where some phases of a circuit are restored before others. Companies vary in the extent to which they track customer minutes of interruption in response to partial restoration of circuits. This can affect both the "start" and "stop" times of a given interruption and the total minutes of the recorded outage.

- *Degree of automation* Companies differ in the extent to which they rely on manual or automated systems (such as outage management systems, or OMSs) to record reliability data. It is quite common for companies' measured frequency and duration of outages to rise substantially after they move to more automated recording systems. This implies that manual systems for measuring interruption data tend to miss or undercount the frequency and duration of outages.

For these and related reasons, there is often significant variation in how companies measure and record reliability indicators. In principle, reliability measurement can be standardized among electric utilities in a jurisdiction, but doing so is likely to take considerable effort. It would also lead to inconsistency between the past and standardized reliability measures for many utilities.

There are other issues that should be recognized when evaluating measured reliability data. Most importantly, the measured reliability of network service can also vary because of external business conditions that are beyond managerial control. Networks have an obligation to provide service to customers in assigned territories. Power delivery also requires direct connection and delivery into the homes and businesses of end users. The conditions of a utility's service territory and customer base can therefore affect the cost and measured quality of service for the delivery networks that utilities construct and maintain. These business condition variables can also vary considerably among companies. The list of relevant business conditions that can impact different aspects of service quality includes:

- weather (*e.g.* winds, storms, lightning, extreme heat and cold)
- vegetation (contact with power lines)
- the amount of undergrounding mandated by local authorities (reducing the contact of power lines with foreign objects but typically increasing the duration of interruptions that do occur)
- the degree of ruralization in the territory (typically increasing the exposure of feeders to the elements and lengthening response times when faults occur)
- the difficulty of the terrain served
- the mix of residential, commercial, and industrial customers (*e.g.* industrial and large commercial customers value power reliability more than smaller customers and are often willing to pay more for it; a greater share of such customers may therefore be correlated with better reliability indices)
- in the short run, it should also be noted that the age of the utility's network can also affect its reliability performance, although in the longer term this variable is subject to managerial control

In addition to varying across distributors, some of these business conditions are quite volatile and unpredictable over time. This is particularly true for weather. This implies that business conditions can lead not only to systematic differences in measured quality across companies, but year-to-year fluctuations in some quality indicators.

In short, appropriate measurement of reliability is particularly challenging. Measured reliability statistics can vary substantially for a given utility over time, or across utilities because of differences in their territories and customer bases. Care must therefore be taken when reliability measures are used in regulation. This is especially true when attempting to make comparisons of reliability performance across networks in an effort to estimate differences in underlying reliability performance.

3.2 Inputs

3.2.1 O&M Cost and Inputs

The indexing logic presented in Section 3.1.1 shows that, just as there is a direct link between regulated revenues and the outputs used to set regulated rates, there is a link between regulated costs and the input quantity measures that should be used in ratemaking. This is apparent in equation (5), which decomposes the change in regulated cost into a change in an input price index and an input quantity index. It follows that the change in input quantities can be measured as the change in the associated cost of the input and the inflation in that input's price.

This relationship applies to operations and maintenance (O&M) costs and, in fact, this is how PEG measures O&M input quantities. In Ontario, extensive data are available on the operations of electricity distributors. Cost data are gathered chiefly from the Trial Balance reports. These reports are filed annually by distributors, as provided for under Section 2.1.7 of the Board's Electricity Reporting and Record Keeping Requirements ("RRRs"). The reported costs are expected to conform with Ontario's Uniform System of Accounts ("USoA").

The available cost data include detailed itemizations of OM&A expenses. The itemizations include the cost of "labour with payroll burden" (presumably salaries and wages) for the following six distribution activities:

- transformer station equipment operation;
- distribution station equipment operation;
- overhead distribution lines and feeders operation;
- underground distribution lines and feeders operation;
- customer premises operation; and
- sentinel lights maintenance.

However, no comparable labor cost itemization exists for other distribution functions, or for any customer care or administrative and general (A&G) functions.

An important supplemental source of Ontario cost data is the Performance Based Regulation ("PBR") reports. These are prepared annually by distributors as provided for under Section 2.1.5 of the Board's RRRs. One potentially important item in these reports is labor's share of OM&A expenses for operation and maintenance (Distribution OM&A), billing and collection, and administration.

The available OEB O&M cost data have a number of strengths that support their use in TFP research. Like the data collected on the FERC Form 1 in the US, the trial balance cost data are highly detailed. The use of a USoA also facilitates standardized reporting.

One important problem with the O&M data is inconsistencies in the allocation of labor expenses between distributor activities. Staff observed in its November 2006 notice that distributors report most customer care labor expenses as administrative expenses. We have found that this problem also extends to distribution labor expenses for many companies.

A related problem is the poor quality of the publicly available data on the salary and wage component of OM&A expenses. On the US FERC Form 1, the salaries and wages assigned directly to OM&A expenses are reported on an itemized basis for all major power distributor activity groups (distribution, customer accounts, customer service and information, and administration and general). Uncertainty regarding the share of labor in OM&A expenses reduces the accuracy of productivity indexes that can be developed for Ontario distributors, since these indexes require information on cost shares.

3.2.2 Capital Cost and Inputs

Measuring capital inputs is far more complex than measuring O&M inputs. Although capital measurement concepts can be organized in many sensible ways, one straightforward approach is to divide the discussion between the measurement of capital *stocks* and capital service *flows*. Capital stock measures pertain to the quantity of capital that exists at a given point in time. Capital flows measure the services provided by that stock of capital over a timer period. Typically, capital flows are more relevant for the capital input quantities or capital cost measures that are used in TFP or cost benchmarking research.

3.2.2.1 Measuring Capital Stocks

It is often easier to observe the quantity of new capital added to the capital stock in any given year than the quantity of the capital stock itself.[21] Capital stock measures typically

[21] This section follows the discussion in C. Hulten, "The Measurement of Capital," pp. 119-158 in *Fifty Years of Economic Measurement*, E.R. Berndt and J.D. Triplett (eds.), Chicago: the University of Chicago Press.

infer the latter from the former, by adding up past capital additions into an overall measure of the capital quantity. This summation of past capital additions must also recognize that older capital additions may be less productive than more recent additions, since capital is "used up" in the production process.

The perpetual inventory equation is the primary method of developing capital stock measures. The perpetual inventory equation adds up investments of different vintages to arrive at a total capital quantity measure. Each vintage is weighted by a number $\emptyset$ that takes a value between zero and one to reflect the possibility that capital will become less productive over time. If $\emptyset = 0$, then capital is essentially fully used up and no longer productive and should not be counted in the capital stock. This can be expressed by the following equation

$$K_t = \emptyset_o I_t + \emptyset_1 I_{t-1} + \ldots + \emptyset_T I_{t-T} \qquad [7]$$

where $\emptyset = 1$ and *t-T* is the date of the oldest surviving capital addition.

The $\emptyset$ terms here are sometimes called efficiency weights, to reflect the fact capital decays over time and become less efficient in providing capital services. It is necessary to determine an efficiency sequence to use the perpetual inventory equation. Three such efficiency patterns have received the greatest attention in the economic literature and applied capital measurement.

The first is the *one hoss shay* pattern. This approach assumes that, provided they have not been retired from use, all capital goods have the same efficiency regardless of the year they were put in place. An example could be a light bulb, which provides the same lighting services from the time it is installed until the time it burns out, at which point its efficiency drops immediately to zero.[22]

The second efficiency pattern is *straight line decay*. Here, efficiency deteriorates in equal increments in each year of the useful life of the asset.[23] The third pattern of efficiency is known as *geometric decay*. Here efficiency decays at a constant *rate* in each year.[24]

---

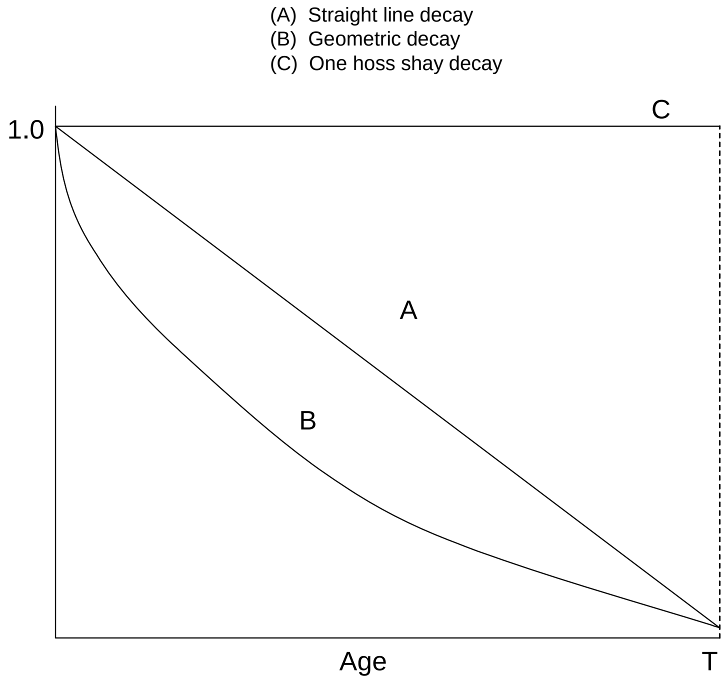
[22] Thus a one hoss shay pattern of efficiency would be

$\emptyset_0 = \emptyset_1 = \ldots = \emptyset_{T-1} = 1,\ \emptyset_{T+t} = 0 \; for \;\; t = 0,\ 1,\ 2\ldots$

[23] In other words, with straight line decay, $\emptyset_{t-1} - \emptyset_t = \frac{1}{T}$. This is equivalent to the following efficiency pattern

These three efficiency profiles over time can be summarized in the figure below.

Figure One: Efficiency Profiles



It should be noted that these patterns apply to the physical decay of the capital assets. This is distinct from the notion of economic depreciation. Depreciation refers to a change in the *value* of an asset. This change will be reflected in the price that a new user would be willing to pay for a capital asset. The difference between that price and the original price of that asset (expressed in constant, inflation adjusted dollars) would equal the total depreciation in the asset value.

Economic depreciation is clearly related to economic decay, but the concepts are distinct. This is most easily seen by considering the one hoss shay case. A rational consumer would not pay the same price for a light bulb that was installed two years ago as one that was installed yesterday. Even if both provide the same services when the light is turned on, the

---

$$\emptyset_0 = 1,\ \emptyset_1 = 1 - \frac{1}{T},\ \emptyset_2 = 1 - \frac{2}{T}, \ldots$$

[24] Let this rate of constant decay be given by the parameter δ. This implies that with geometric decay $\emptyset_0 = 1,\ \emptyset_1 = (1-\delta),\ \emptyset_2 = (1-\delta)^2, \ldots$

older light bulb is closer to the end of its useful life and therefore less valuable. Thus the value of the light bulb has depreciated, even if its physical efficiency in terms of the services it provides has not. In general, economic depreciation and physical capital decay are interdependent but different concepts, and the pattern of economic depreciation over time will differ from that of economic decay.

The one exception to this rule is for geometric decay. It turns out mathematically that, when capital decays at a constant rate, the value of the capital asset also depreciates at a constant rate over time. Thus the pattern of economic depreciation will coincide with the pattern of physical capital decay in the case of geometric decay, and only in this case. In each year, then, the value of a capital good that is subject to geometric decay will decline at a rate of δ each year.

In practical terms, measuring the stock of capital typically begins with a *benchmark* capital stock, or (price deflated) value of capital in some base year.[25] If there was a full series of capital stock additions, then it would not be necessary to start with a benchmark capital stock, for the perpetual inventory equation could be applied to all capital additions since the beginning of the enterprise. In practice, however, it is almost never possible to obtain the full historical series of capital stock changes, so capital stock measurement must begin with a benchmark value in a base year.

The base year for the capital stock should be as distant from the present as is practical. As the base year becomes more distant, the value for the latest values of the capital stock depend more on observed values for capital additions, which have been added to this benchmark value. The value for the benchmark capital stock therefore becomes relatively less important in terms of the values of capital quantities that are computed.

The following perpetual inventory equation is used to compute subsequent values of the capital quantity index (*i.e.* the capital stock):

$$XK_t = (1 - d) \cdot XK_{t-1} + \frac{VI_t}{WKA_t}. \qquad [8]$$

[25] Benchmark capital stocks are typically deflated by a "triangularized weighted value" of capital asset prices over the 20 or 30 year period preceeding the year of the benchmark capital value.

Here, the parameter, d, is the economic depreciation rate, VI_t is the value of gross additions to the utility plant and WKA_t is an index of utility plant asset prices.

3.2.2.2 Measuring Capital Flows

Applied TFP and benchmarking work are typically concerned with the *flow* of services provided by capital goods rather than the stock of capital that exists at any point in time. For example, a TFP measure is theoretically designed to relate the flow of outputs provided to customers to the flow of inputs used to produce those outputs over a given time period. Since capital is the major input for electricity networks, this implies that network TFP measures must develop measures of the flow of capital services that utilities provide.

Determining the flow of capital services is usually not straightforward. One reason is that capital goods are durable and provide a flow of services over a multi-year period. While companies may record the price of the capital good when it is purchased and the date when that good is retired (or replaced), they rarely attempt to measure the flow of services provided over a given period. A related problem is that capital services are usually implicitly provided within a single enterprise rather than via arms length transactions in the marketplace. Firms often consume capital goods over a multi-year period, which implies that the provider and the ultimate user of capital services are located in the same economic enterprise.

Conceptually, the flow of capital services associated with an asset could be observed as the amount that consumers pay to rent that asset for a given period. This data is available for some property assets (*e.g.* apartments and commercial rental space), but there is a paucity of such data for most capital goods. An early description of this problem is found in Griliches (1963), who wrote

> Ideally, the available flow of services would be measured by machine-hours or machine-years. In a world of many different machines, we would weight the different machine-hours by their respective rents. Such a measure would approximate most closely the flow of productive services from a given stock of capital and would be on a par with man-hours as a measure of labor input.[26]

---

[26] Griliches, Z. (1963), "Capital Stock in Investment Functions: Some Problems of Concept and Measurement", in *Measurement in Economics,* C. Christ et. al (editors), Stanford: Stanford University Press; reprinted in Z. Griliches (1988), *Technology, Education, and Productivity*, New York: Basil Blackwell.

This idea implies that the flow of capital services can be measured as the product of a *rental price* and a capital quantity stock. This product also gives the (implicit) outlay on capital services over the period, or the capital cost. Given a capital quantity stock measure, like that developed through the perpetual inventory equation, a measure of capital service flows therefore only requires imputing the rental prices of capital goods.[27]

In the general methodology, capital cost in a given year *t*, CK_t, is therefore the product of a capital rental price (also referred to as a capital service price index) WKS_t and a capital quantity index, XK_{t-1}.

$$CK_t = WKS_t \cdot XK_{t-1}. \qquad [9]$$

The capital rental or service price index may be thought of as the annual cost (including the opportunity cost) of owning a unit of plant.[28]

Each capital quantity index is constructed using inflation-adjusted data on the value of utility plant. Each service price index measures the trend in the hypothetical price of capital services from the assets in a competitive rental market. The price and quantity indexes require a consistent mathematical characterization of the process of plant deterioration. A common formula for the capital service price index, WKS_t, is:

$$WKS_t = \left(CK_t^{taxes} / XK_{t-1}\right) + r_t \cdot WKA_{t-1} + d \cdot WKA_t - \left(WKA_t - WKA_{t-1}\right) \qquad [10]$$

The four terms in this formula correspond to taxes, the opportunity cost of capital, depreciation, and capital gains.

3.2.2.3 Capital Data for Ontario Distributors

There are significant challenges associated with measuring capital in Ontario. Accurate and standardized capital cost measures require years of consistent, detailed plant additions data. However, PBR data on plant additions in Ontario are only available since 2002, which reduces the reliability of the capital cost and quantity measures that can be computed for Ontario distributors. In particular, measured capital costs will be highly

---

[27] These rental prices are imputed via the theoretical equality between the value of an asset and the discounted value of services it provides. That is, in a hypothetical rental market, the price that an asset would fetch at a point in time would be based on the value of services it would provide to the user over its remaining useful life.

[28] The capital service price in general has four components: 1) the opportunity cost of capital, also known as the cost of funds; 2) depreciation; 3) capital gains; and 4) taxes levied on capital.

sensitive to our estimate of the quantity of capital on hand in 2002. This "benchmark year" calculation requires a suitably weighted index of construction costs over the past forty years.

It is possible that additional capital additions data may be available from other sources. For example, first generation incentive regulation for electricity distributors developed capital measures for a sample of distributors using capital additions for multiple years, obtained from a survey of municipal electric utilities. Some historical data could also possibly be obtained from MUDBANK, although both the availability and quality of this data source is presently unknown. If the Board intends to estimate the growth in TFP for Ontario electricity distributors and/or develop a total cost benchmarking model, assessing the quality of availability of alternative capital data should be an important part of this consultation.

Given the challenges with the capital stock data in Ontario, it may be argued that "physical" capital metrics, such as the km of distribution line, are an acceptable proxy for the monetary capital measures that are typically used in TFP and cost analyses. In third generation incentive regulation, this was advocated by the Coalition of Large Distributors (CLD's), and they proposed an alternate TFP estimate that relied in part on physical capital metrics. The Board rejected CLD's TFP estimate and said its "greatest concern" with their methodology was their measurement of capital. This was an appropriate decision because there are almost no circumstances in which physical capital measures should ever be used in TFP or cost research. The relative merits of physical and monetary measures of capital are discussed extensively in Appendix Two of this Report.

3.3 Productivity

3.3.1 Index-Based Estimates of TFP Trends

Estimates of TFP growth for network industries are typically used to calibrate the X factor in index-based, incentive regulation plans for network utilities. TFP in most regulatory applications is measured using indexing methods. When this is the case, TFP growth is defined as the growth in an index of comprehensive output quantity minus the growth in an index of comprehensive input quantity. As discussed in Section 3.1.1, in ratemaking applications, the output quantity index should be measured using the networks' billing determinants. The input quantity index will be measured using subindexes of O&M and capital inputs. The measurement of these specific inputs was described in Section 3.2.

In most regulatory proceedings where TFP trends have been estimated using indexing methods, long-run TFP trends have been estimated using about 10 years worth of historical data. The Board used a somewhat longer, 18 year period to measure industry TFP growth in third generation incentive regulation in Ontario. Whatever period is selected should be sufficient for smoothing out short-term fluctuations in TFP that can arise because of changes in output (*e.g.* kWh deliveries that are sensitive to changes in weather and economic activity) and the timing of different types of expenditures. This long-run historical TFP trend is then assumed (either implicitly or explicitly) to be a reasonable proxy for the TFP growth that is expected over the term of the indexing plan.[29]

This is not always an appropriate assumption. For example, it is often not warranted to assume that TFP growth measured for short historical periods will be a good proxy for future trends. Shorter sample periods are more likely to be distorted by factors such as the timing of expenditures or unusual output growth. There is accordingly less confidence that past TFP trends are a good proxy for the future trend if the available data only allows TFP to be calculated for a relatively short period. As discussed, a general rule of thumb in regulatory proceedings is that a minimum of 10 years of data are needed to calculate a generally reliable estimate of the industry's long-run TFP trend.

Another instance where the industry's past TFP trend may not be appropriate going forward is when utilities face capital investment needs that differ substantially from the recent past. This issue was discussed extensively in third generation incentive regulation, and to accommodate diversity in capital expenditures the Board approved an optional, incremental capital module where distributors could request additional rate relief to fund incremental capital investments. The need to allow flexibility in networks' capital spending may be greater since the passage of the GEA.

3.3.2 Econometric Estimation of TFP Trends

In addition to estimating historical TFP trends using indexing methods, econometric methods can be used to estimate TFP growth. Such an approach is well-suited for projecting TFP growth when there is a lack of historical, time series data. The econometric approach

[29] Although, a consumer dividend is also sometimes added to this historical TFP trend to reflect the expected acceleration in TFP relative to the industry's historical norms when a firm becomes subject to PBR.

essentially uses statistical methods to estimate the underlying "drivers" of TFP growth, such as technological change and the realization of scale economies. Statistical techniques can estimate the impact of each of these sources of TFP growth by using data from electricity distributors operating under a wide variety of business conditions. Once those underlying TFP "drivers" are estimated, they can be combined with data on the changes in the business condition variables that apply for either individual electricity distributors or for groups of distributors. This information can then be brought together using a methodological framework that is detailed in Appendix One of this report.

The econometric approach to estimating TFP growth has a number of potential advantages. One is that it is rigorous and has a strong foundation in statistical methods and the economics literature. This approach can also be tailored to reflect the specific business conditions, and "TFP drivers," of the Ontario power networks. A TFP decomposition model can be operationalized using data from the electricity networks themselves on their identified TFP drivers. We can, for example, calculate productivity trends for individual Ontario utilities, or groups of utilities, that are specific to their operating scale and their expectations concerning output growth, undergrounding, and other business conditions. This allows TFP trends to be customized to the special operating conditions of individual utilities while at the same time ensuring that the price cap index remains "external," since the TFP driver parameters are estimated using large datasets and are thus insensitive to a company's performance while subject to the PBR plan.

There are also regulatory precedents for using econometric methods to estimate TFP growth. Econometric decompositions of TFP growth have been presented in California regulation. For example, CPUC staff have estimated the expected productivity growth of individual utilities that are specific to their operating scale. The most recent gas distribution IRM in Ontario included consideration of econometric projections of TFP growth for both Enbridge and Union Gas. Econometric methods have also been used to project partial factor productivity (PFP) growth for gas distribution operating expenditures in Victoria, Australia.

The main disadvantage of the econometric approach is its complexity. Econometrics often involves technically complex statistical methods. The TFP estimates that result from econometric modeling therefore tend to be less transparent and not as easy to understand as those resulting from indexing methods. While unnecessary complexity should be avoided in regulatory proceedings, it is not always practical or desirable to rely on simpler, index-based TFP estimates when calibrating the terms of indexing formulas. This would be the case, for example, if the available time series data was either too short, or distorted by transitory factors, and therefore did not yield reliable estimates of long-term TFP trends.

3.4 Performance

There are a variety of methods that can be used to quantify, and "benchmark," a network's cost or quality performance. This chapter will briefly describe four performance measurement techniques: index-based methods; econometric cost functions; stochastic frontier analysis (SFA); and data envelope analysis (DEA). These can be viewed as the primary methods, but there are variants on some of these basic models.[30] Appendix Three presents additional details on these methods, as well as an analysis of the pros and cons of each as a performance measurement technique.

3.4.1 Index-Based Methods

Index-based methods can be used to measure performance, as well for setting the terms of rate adjustment formulas. A total factor productivity (TFP) index is a comprehensive performance measure that includes all of the inputs and outputs of an economic unit. In contrast, a partial factor productivity (PFP) index is a partial performance measure. TFP and PFP indexes can compare productivity performances between firms at a point in time or for the same firm (or group of firms) at different points in time.

3.4.2 Econometric Methods

A cost function is a mathematical relationship designed to capture the relationship between the cost of service and business conditions. Business conditions are aspects of a company's operating environment that may influence its activities but cannot be controlled. Economic theory can guide the selection of business condition variables in cost function

[30] For example, stochastic frontier analysis is similar to "thick frontier analysis" and "distribution free analysis;" a discussion of the differences between these frontier estimation techniques is found in Bauer, P., A. Berger, G. Ferrier, and D. Humphrey (1998), "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods," *Journal of Economics and Business*, 50: 85-114. However, the latter two estimation techniques are rarely used in utility benchmarking. In the interests of brevity, we therefore deal only with the four benchmarking methods listed above, but the discussion is generally applicable to benchmarking techniques that may be closely related to these methods.

models. According to theory, the total cost of an enterprise depends on the amount of work it performs - the scale of its output - and the prices it pays for capital goods, labor services, and other inputs to its production process.[31] Theory also provides some guidance regarding the nature of the relationship between outputs, input prices, and cost. For example, cost is likely to rise if there is inflation in input prices or more work is performed.

In addition to output quantities and input prices, networks confront other operating conditions due to their special circumstances. Unlike firms in competitive industries, energy networks are obligated to provide service to designated customers within a given service territory. Many utility services are also delivered directly into the homes, offices and businesses of end-users. Utility cost is therefore sensitive to the circumstances of the territories in which they provide delivery service.

One important factor affecting cost is customer location. This follows from the fact that utility services are delivered over networks that are linked directly to customers. The location of customers throughout the territory therefore directly affects the assets that utilities must put in place to provide service. Different spatial distributions for customers can have different implications for network cost.

Cost is also sensitive to the mix of customers served. The assets needed to provide delivery service will differ somewhat for residential, commercial, and industrial customers. Even more importantly, different types of customers have different levels and temporal patterns of demand and different load factors.

In addition to customer characteristics, cost can be sensitive to the physical environment of the service territory. The cost of constructing, operating and maintaining a given network will depend on the terrain over which that network extends. These costs will also be influenced by weather and related factors. For example, costs will likely be higher in areas with high winds, a propensity for ice storms or other severe weather that can damage equipment and disrupt service. Operating costs will also influenced by the type and density of vegetation in the territory, which will be at least partly correlated with precipitation and

---

[31] Labor prices are usually determined in local markets, while prices for capital goods and materials are often determined in national or even international markets.

other weather variables. To a great extent, these conditions accompany the particular territory that the power distributor is required to serve and are therefore beyond management control.

Econometric cost functions require that a functional form be specified that relates cost to outputs, input prices, and other business conditions. Parameters are associated with the variables specified in this cost function. Econometric methods are then used to estimate the parameters of cost function models. Econometric estimates of cost function parameters are obtained using historical data on the costs incurred by utilities and measurable business condition variables that are included in the cost model.

3.4.3 Stochastic Frontier Analysis

Stochastic frontier analysis (SFA) is similar in many respects to other econometric cost models. SFA also specifies a functional form that relates cost to outputs, input prices, and other business conditions. The same business condition variables would be used in SFA as in econometric cost functions. Parameters of SFA models are estimated using historic data on the variables used in the cost function.

However, SFA differs in that it also estimates an inefficiency factor for each firm. SFA is specifically focused on estimating the minimum cost of production.[32] The actual total cost (C_i) incurred by company, i, in providing service is assumed to be the sum of the minimum achievable cost (C_i^*) and an inefficiency factor.

$$C_i = C_i^* + inefficiency_i$$

SFA uses econometric methods to isolate and measure this inefficiency factor. While not estimating firm inefficiency directly, it should be noted that econometric cost functions can also be specified that distinguish between inefficiency and other random factors that are not reflected in the business condition variables. We discuss this issue in Appendix Three.

3.4.4 Data Envelope Analysis

Data envelope analysis (DEA) represents a much different approach towards measuring performance. It does not estimate the parameters of a cost function and is therefore often described as "non-parametric." Instead, DEA uses linear programming

---

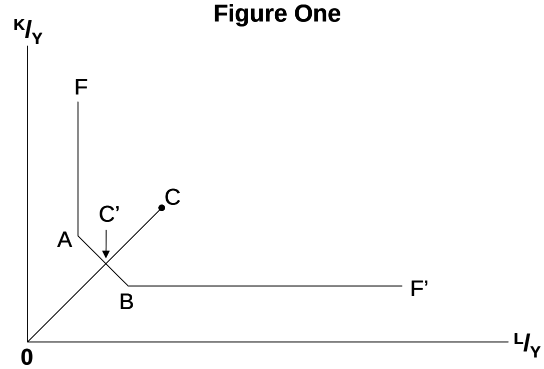
32 Alternatively, SFA can be focused on estimating maximum production frontiers.

techniques to "envelope" data on sample firms that relate outputs to inputs. DEA is therefore essentially a technique for identifying what are known in economics as isoquant or isocost curves and in measuring the distance of individual firms from the efficient cost (production) frontier reflected in that isocost (isoquant).

In a basic input-oriented DEA model, the relative efficiency of a firm is determined by assigning weights to firm inputs and outputs such that the ratio of aggregated outputs to aggregated inputs is maximized. This linear programming problem is subject to the constraint that the efficiency score cannot exceed a value of one for a firm using the same set of weights. The result of this process will be an efficiency measure for each firm that takes a value between zero and one. These efficiency scores are relative to "peers" identified through the analysis and which set the efficiency "frontier." The DEA efficiency score has the intuitive interpretation that, relative to the peers, it measures the amount by which a firm can radially contract all of its inputs while still producing the same level of output.

This can perhaps be clarified through a visual example. In Figure One, there are two inputs, capital (K) and labor (L). The X axis in this figure is labor per unit of output (L/Y) while the Y axis is capital per unit of output (K/Y).

Figure One



In this example, the points A, B and C refer to specific firms that are identified as peers. It can be seen that firms A and B are using fewer capital and labor inputs per unit of output than firm C. The DEA technique would construct a piece-wise linear frontier through points A and B, which is identified by the line FABF'. This line is the production frontier. The efficiency of firm C is measured relative to this frontier, and the efficiency measure is equal to OC'/OC. Suppose this value turns out to be 0.6. This implies that firm C is 40% below the production frontier, and it can reach the frontier by reducing both its capital and labor inputs by 40%. Under input-oriented DEA, the firm's measured inefficiency is therefore equal to the entire difference between its position and the constructed efficiency frontier.

The basic input-oriented DEA model can be expanded in various ways. Technically, this occurs by modifying the linear programming problem to relax various assumptions. These more DEA sophisticated models will break down the sources of efficiency into "technical" or productive efficiency, and (supply side) allocative efficiency.[33]

DEA can also be modified to include second-stage regressions that regress DEA efficiency scores on other business condition variables. The results of these regressions can then be used to adjust the efficiency scores resulting from the DEA analysis. The primary reason for undertaking such regressions rather than including all relevant business condition variables in the linear programming problem is that increasing the number of inputs in DEA analysis tends to reduce the number of peers that are identified for any firm. Having fewer peer firms can artificially inflate the efficiency measure. Indeed, in the limit, if enough inputs are introduced in the analysis, then no firm may be identified as a peer for any other firm. The DEA measure therefore becomes one for all firms by default, which is clearly an unrealistic result.

[33] For example, the model above assumes constant returns to scale in the relationship between inputs and outputs. This assumption can be relaxed. Doing so would allow the technical efficiency measure above to be decomposed into scale efficiency and "pure" technical efficiency. Other assumptions can be relaxed that allow further decomposition into congestion efficiency and allocative efficiency.

4. Setting Standards

In general terms, a regulatory "standard" gauges how effectively a utility is achieving regulatory objectives. There can be different standards that reflect regulators' expectations for what constitutes acceptable, average, desirable and/or superior performance. This chapter discusses some principles that should be kept in mind when regulators consider adopting standards in utility regulation.

4.1 Standards and Performance Benchmarking

"External" performance measures are often termed benchmarks or performance standards. In a regulatory context, a standard will be external to a utility if the utility's own actions cannot influence the value of the standard. Benchmarking is one specific means of establishing external standards. In utility regulation, benchmarking involves comparing one or more utility performance measures to external performance standards. Compared with external benchmarks more generally, benchmarking differs in that it relies on *direct* and *explicit* comparisons between a company's performance and the external performance standard.

External standards can in principle lead to significant benefits in utility regulation. It is widely acknowledged that, compared with using a utility's own costs to set rates, external benchmarks create stronger performance incentives. Stronger performance incentives will, in turn, promote utility behavior to control costs and comply with regulatory objectives (such as maintaining appropriate service quality) more efficiently, which can benefit both customers and shareholders.

The role of benchmarking *per se* in promoting effective regulation is less certain. In principle, benchmarking can play a potentially valuable role in promoting effective regulation. Benchmarking can be a tool for ensuring that regulation replicates the operation and outcomes of competitive markets. Creating incentives for utility operations that are comparable to competitive markets would ultimately create benefits for both consumers and shareholders. Thoughtful and rigorous benchmarking studies can be helpful for establishing objective performance standards that strengthen incentives and increase the potential benefits from utility services.

While this potential exists, it must also be recognized that benchmarking is simply a tool, and like any regulatory tool it can be abused. Inappropriate benchmarking can be destructive and contrary to the goal of effective regulation. For example, “bad” benchmarking studies can set unrealistically demanding performance standards. Such standards can lead to prices that do not recover the costs of even an efficiently run company. While this may be corrected over time (*e.g.* in an updated benchmarking study), there may still be lasting damage. Utilities are highly capital-intensive enterprises and continually raise debt and equity capital. The use of inappropriate benchmarking studies by regulators can raise a utility’s cost of capital as investors demand risk premiums to compensate for heightened regulatory risks. These higher costs would ultimately be reflected in higher prices. Bad benchmarking studies can therefore reduce long-run benefits to both customers and shareholders.

Benchmarking that is biased in favor of companies can lead to similarly undesirable results. Here, customers would either pay unreasonable prices for utility services or shareholders would enjoy superior returns even though the utilities do not exhibit superior performance. Ultimately, benchmarking must be designed so that returns are commensurate with performance. Any benchmarking approach that is not compatible with this goal does not promote sound public policy.

In light of these concerns, it should always be remembered that benchmarking can either promote or frustrate effective regulation depending on how it is applied. By increasing risk, “bad” benchmarking applications may unintentionally raise costs and therefore frustrate rather than promote desired regulatory objectives. Benchmarking is therefore a double-edged sword, and its impact depends on the understanding and care with which it is wielded in practice.

4.2 Principles for Establishing Appropriate Cost Standards

It is widely believed that effective utility regulation should replicate the operation and outcomes of competitive markets. One reason is that competitive market forces create maximum incentives to operate efficiently. Firms in competitive markets that do not produce efficiently have lower profits as sales are lost to more efficient rivals. Reduced profits, in turn, create pressures to reduce costs. Similarly, firms that choose non-optimal prices or do not produce the products that consumers demand lose sales to competitors. Profits thereby decline, leading to changes in marketing behavior that satisfy consumer demands. Economic theory has also established that competitive markets often create the maximum amount of benefits for society.[34] For these and related reasons, a "competitive market paradigm" is useful for establishing effective performance standards. Below we consider how competitive markets operate and the implications for setting appropriate standards in utility regulation.

One important aspect of competitive markets is that prices are external to the costs or returns of any individual firm. By definition, firms in competitive markets are not able to affect the market price through their own actions. Rather, in the long run, the prices facing any competitive market firm will change at the same rate as the growth in the industry's unit cost.

Competitive market prices also depend on the *average* performance in the industry. Competitive markets are continually in a state of flux, with some firms earning more and others less than the "normal" rate of return on invested capital. Over time, the average performance exhibited in the industry is reflected in the market price.[35]

Taken together, these features have the important implication that in competitive markets, returns are commensurate with performance. A firm can improve its returns relative to its rivals by becoming more efficient than those firms. Companies are not disincented from improving efficiency by the prospect that such actions will be translated into lower prices because the prices facing any individual firm are external to its performance. Firms that attain average performance levels, as reflected in industry prices, would earn a normal return on their invested capital. Firms that are superior performers earn above average

[34] This is sometimes known as the "First Fundamental Welfare Theorem" of economics.

[35] This point has also been made in the seminal article, *Incentive Regulation for Electric Utilities* by P. Joskow and R. Schmalensee. They write "at any instant, some firms (in competitive markets) will earn more a competitive return, and others will earn less. An efficient competitive firm will expect on average to earn a normal return on its investments when they are made, and in the long run the average firm will earn a competitive rate of return"; p. 11.

returns, while firms with inferior performance earn below average returns. Regulation that is designed to mimic the operation and outcomes of competitive markets should allow for this important result.

There are many ways to set long-run cost standards. Below we discuss two potential approaches. These options differ in terms of the information they require to be implemented, the risks that they create for energy networks and, potentially, the strength of incentives that are created to improve network performance.

The riskier approach, which involves a very strong application of the competitive market paradigm, is to set a long-run regulatory standard whereby regulated rates reflect performance levels that would be expected for an average firm in a competitive industry. Economic research may be helpful in determining this target. For example, competitive markets can be examined to establish how close firms are, on average, to superior performers in the industry. This can provide evidence of the impact that competition ultimately has on the performance of a typical firm relative to the industry's superior performing firms.

Benchmarking can also be useful in achieving this objective. Benchmarking can assess utility performance levels relative to the norm and superior performance levels in the industry. Benchmarking can therefore set objective performance targets that are superior to the industry norm and that move utilities in the direction of better performance levels that would be expected under competition.

While this approach has some conceptual appeal, it also entails considerable risks. Most importantly, it places great weight on knowledge that is difficult to attain and inherently uncertain, such as the relationship between average and superior performance levels in competitive industries. It also relies heavily on the accuracy of benchmarking methods. These methods are in their infancy in utility regulation and will be particularly uncertain about what constitutes the industry's performance "frontier." This approach will therefore be especially risky if regulators believe that regulation should move all companies to the frontier. Overall, this method places a premium on sharing speculative performance gains with customers and therefore puts utilities at risk if these gains do not materialize.

A simpler and less risky approach is to include a stretch factor component to the X factor in a multi-year, "inflation minus X" rate adjustment formula. Benchmarking can be used to inform the value of the stretch factor as well as determine when it is appropriate to remove the stretch factor. For example, the stretch factor can be eliminated when benchmarking studies demonstrate that the company's cost performance is significantly lower than expected, by a certain threshold amount (say 10% below the predicted cost of service). This result implies that the utility's customers are already benefiting from superior performance levels. Like the first option, this benefit-sharing approach does not depend directly on the company's actual performance gains, but it places less emphasis on speculative and uncertain information.

Ontario's third generation incentive regulation plan for electricity distributors does, in fact, use benchmarking evidence to inform the relative valuations of stretch factors. Distributors are assigned into one of the three efficiency "cohorts" depending on how they perform on two separate benchmarking studies. Relatively more efficient distributors were assigned lower stretch factors. Implicitly, this approach is consistent with designing rate regulation to encourage energy networks to improve their cost performance and to reward them when they do.

Many paths can be taken towards using performance standards to encourage long-run cost performance objectives. When evaluating different alternatives, regulators should consider how options differ in terms of risk and information requirements. The approach that is most appropriate in any given situation will depend on a number of factors, including the institutional environment and the amount and quality of data that are available. In all cases, however, several factors should be kept in mind when making the competitive market paradigm operational.

First, in competitive markets, movements towards long-run efficiency levels will take place gradually. One reason is that adjusting company operations to achieve greater efficiencies is usually costly. Companies must in general devote resources towards improving their performance, and payoffs from those actions in improved efficiency typically take time to materialize. This process can be expected to be especially long for industries such as power distribution where assets are dedicated to serving particular customers (*e.g.*

directly delivering to a customer's premises) and therefore have less value in alternative uses. It is particularly costly to adjust operations in this case since many assets have secondary market values far below their current values. Discarding existing capital can therefore lead to large capital losses which, in turn, tends to increase the rigidity of capital stocks. For this and related reasons, any movement towards benchmark-based performance targets should take place gradually.

It should also be remembered that price levels in competitive markets reflect the industry's average efficiency performance. This means that firms with superior performance earn above average returns. This is true even in the long run.[36] This implies that it is not reasonable to impose "frontier" performance standards on all firms in the industry since this does not allow returns to be commensurate with performance. Companies must always have "room" to outperform the benchmark that is reflected in the prices they face. This enables the firm to be appropriately rewarded for superior performance. If the industry's best-observed practice is imposed on all firms, any firm that fails to achieve this standard will earn below average returns. This would be true even for superior performers that nevertheless fall short of the industry's best performance. This outcome is clearly contrary to having returns be commensurate with performance and thus is not consistent with effective regulation.

It is also important to recognize that there will be considerable uncertainty about what constitutes a "frontier" performance level. Targets established through benchmarking should be cognizant of this uncertainty. Regulators should not impose performance standards for which there is significant probability that well-managed utilities will fail to achieve these targets. The benchmarks should therefore make appropriate allowance for the uncertainty associated with attaining the target performance levels.

---

[36] There are both short-run and long-run equilibria in competitive markets. In the short run, equilibrium occurs whenever quantity supplied equals quantity demanded. But the industry will not be in long-run equilibrium if average returns in the industry are not equal to the competitive rate of return, defined to be the opportunity cost of capital. For example, if average industry returns exceed the competitive rate of return, long-run equilibrium is established as new firms enter the industry and existing firms expand their production, thereby increasing supply and driving down prices and average returns. This process continues until the industry's average return equals the competitive rate of return. For evidence that superior performers continue to earn above-average returns even in the long run, see L. Schwalbach, U. Grabhoff, and T. Mahmood, "The Dynamics of Corporate Profits," *European Economic Review*, October 1989, 1625-1639.

4.3 Service Quality Standards

Setting standards can also be important for promoting service quality goals. In practical terms, two main sources of information can be used to set standards and deadbands in regulatory plans. The first option is peer performance. In principle, peer–based benchmarks may be attractive since they are commensurate with the operation and outcomes of competitive markets, where firms are penalized or rewarded for their price and reliability performance relative to their competitors. In practice, however, peer–based benchmarks are challenging. One reason is that uniform data are not generally available for some utility metrics, such as reliability. Differences in measure definitions would make peer data difficult to compare and inappropriate as benchmarks. Even if measures are defined comparably across utilities, peer benchmarks should control for differences in utility business conditions that affect performance. Controlling for the impact of business conditions on expected system reliability performance is complex. While peer-based benchmarks are rare, this approach has been used in Norway, Sweden, and the Netherlands.

The alternative is the utility's own performance on an indicator. For example, benchmarks could be based on average performance on a given indicator over a recent period. Reliability assessments would then depend on how an individual utility's measured reliability levels differ either positively or negatively from its recent historical experience. Historical benchmarks are used in a number of North American jurisdictions including Massachusetts and California.

The use of past utility performance to set standards is appealing in many respects. Historical benchmarks reflect a company's own operating circumstances. Historical data will reflect the typical external factors faced by the company if the period used to set benchmarks is long enough to reflect the expected temporal variations in these factors. Longer periods are more likely to achieve this goal than shorter periods and are therefore preferred. If only short time series are available at the outset of a regulatory plan, benchmarks can be updated at the outset of future plans as more data become available. The rules for updating benchmarks should be spelled out clearly in advance to create the appropriate performance incentives and minimize administrative burdens.

A potential concern with using a company's past performance to set benchmarks will arise if the utility has historically registered substandard reliability performance. If this is the case, the benchmark would reflect a level of inefficiency in system reliability delivery. A more objective standard of system reliability performance may then be appropriate and would benefit of customers. However, evaluating whether a company's historical system reliability performance is substandard requires controlling for factors beyond companies' control that can impact their system reliability performance. This can be done by applying the performance benchmarking techniques discussed in Chapter Three to service quality metrics, although these studies can be complex and will certainly entail greater administrative costs than simpler system reliability regulation approaches.

One important criterion for setting standards is that benchmarks should be calculated on the same basis as the reliability indicators. If the data used to measure the indicator are not comparable to those used to set the benchmark, the regulatory plan will not lead to an objective comparison of the company's measured performance relative to the benchmark. This is almost literally a case of 'comparing apples to oranges'. Discrepancies between measured and historical benchmark performance can arise if utilities change the measurement systems used to record, such as installing a new outage management system (OMS) to record reliability.

Standards and deadbands should also reflect external business conditions in a utility's service territory. A failure to control for these business conditions in a regulatory benchmark can expose utilities to arbitrary and unfair performance evaluations. For example, consider a plan where a utility is rewarded or penalized depending on how its measured reliability compares to that of another utility. Assume that both companies measure every indicator in the same way. This plan would still lead to unreasonable penalties or rewards if one utility had a more ruralized territory with very low population densities. Not controlling for the effect of business conditions in that service territory would tend to handicap the more rural utility and, over time, lead to penalties that did not reflect its real performance.[37]

[37] For example, suppose the company in the more demanding territory really had worse service quality performance than the other firm in a given year; this plan would lead to penalties both for worse performance and because one firm had more demanding conditions that made it more difficult to provide the same level of

Third, all else equal, standards should be as stable as possible during the regulatory plan. Stable benchmarks give utility managers more certainty over the resources they must devote to providing adequate service, as reflected in those benchmarks. It is harder for managers to hit a 'moving target', particularly if operational changes can only be implemented over longer periods. Stable benchmarks therefore promote more effective, longer–term service quality programs.

In some cases, however, a lack of data available at the outset of regulatory plan may make it more difficult to set benchmarks that are viewed as reliable over the term of a multi–year plan. This would be true if the information systems used to record reliability data had changed recently or if there was little confidence that a short data series reflected typical external business conditions for the utility. If this is the case, benchmarks can be updated using data that becomes available during the term of the plan, but this should be done according to well–defined rules that are established at the outset of the plan. An example would be a benchmark equal to a moving average of a company's historical performance on an indicator, until 10 years of historical data are available. This type of approach has been implemented in Massachusetts. Setting benchmarks according to such objective rules creates as much stability as is feasible given data constraints.

Although historical averages of company performance will reflect typical external factors faced by a company, they will not control for shorter–term fluctuations in external factors around their norms. As noted, some business conditions that can affect measured quality are quite volatile from year to year. Weather is the salient example.

One way to accommodate year–to–year fluctuations in external factors is by measuring indicators on a multi-year basis. For example, a regulatory plan could target a three-year moving average of SAIFI and SAIDI rather than the SAIFI and SAIDI values registered each year. Measuring indicators over multiple years will tend to smooth out the impact of random factors on indicator values and lead to a more reasonable measure of the

---

service as the other firm. In principle, a firm in a more demanding territory could also have better system reliability performance and yet still register worse measured reliability performance because of the impact of its more demanding business conditions. Here, the company is penalized even though it is a superior performer. In both cases the company's penalties do not reflect its real reliability performance unless adjustments are made to the plan to reflect differences in the companies' service territories.

company's underlying service quality performance. New Jersey LDCs and Alberta-based Enmax both use a 5 year rolling average to calculate their system reliability benchmarks.

Another way to accommodate year-to-year fluctuations in external factors is through deadbands. Suppose, by way of example, that the value of a reliability indicator is known to fluctuate in a certain range due to external factors. The mean value of this indicator over a suitable historical period would reflect the typical long run external business conditions faced by the utility. Variation in the company's performance around this historical mean will accordingly reflect short run fluctuations in those business conditions. Deadbands should therefore reflect the observed variability in measured system reliability performance. One straightforward measure of this year–to–year variability is the standard deviation of the reliability indicator around its mean. San Diego Gas & Electric, based in California has a deadband around their SAIDI and SAIFI performance before penalties or rewards begin to accrue. New Zealand has gone the furthest to minimize volatility in their SAIDI and SAIFI performance by incorporating a regime which uses both a deadband and even allowing an LDC to be non-compliant in one out of three years before penalties are considered.

5. Creating Incentives

A variety of mechanisms can be used to create incentives for energy networks to achieve regulatory objectives. This chapter will briefly describe and assess the main incentive regulation approaches. We analyze, in turn: 1) indexed rate caps; 2) revenue caps; 3) earnings sharing mechanisms; 4) benchmark regulation plans; 5) plan termination provisions; 6) menu approaches; and 7) targeted incentives. We then turn to regulatory approaches that rely more on regulatory discretion and judgment, rather than established incentive regulation plans, as the means for creating appropriate behavioral incentives.

5.1 Indexed Rate Caps

Indexed rate caps are the most common form of incentive regulation in the world today. This approach is also very familiar to stakeholders in Ontario, since it has been applied several times in the Province in electricity and gas distribution regulation. Under an indexed rate-cap plan, restrictions are placed on changes in prices for regulated services. The limits are called caps since utilities are often free to charge rates that are less than the maximum allowed. The mechanisms for restricting allowed rate growth vary, but the most general approach is for changes in utility prices to be measured using actual price indexes (APIs). An API can in principle be calculated for each specified "basket" of utility services. Growth in each API is capped by a price cap index (PCI), as expressed by the following formula:

$$\Delta API \leq \Delta PCI. \qquad [15]$$

While PCI formulas vary from plan to plan, it is generally true that the PCI growth rate (ΔPCI) is the difference between an inflation factor (*P*) and an X-factor (*X*), plus or minus a Z-factor (Z). The standard formula is

$$\Delta PCI = P - X \pm Z. \qquad [16]$$

The inflation factor, *P*, is the growth rate in an external price inflation measure. Three types of inflation measures have been used in approved rate-cap plans. These may be termed macroeconomic, industry-specific, and peer price measures.[38] The X-factor is also sometimes

[38] Macroeconomic inflation measures are summary measures of price growth for a wide range of goods and services and are typically computed by government agencies. Examples include the chain-weighted price index for gross domestic product (GDPPI), consumer price indexes (CPIs), and producer price indexes (PPIs). Industry-specific inflation measures are expressly designed to track inflation in the prices of the relevant utility

called a "productivity factor" since, in North American proceedings, its value is typically estimated using estimates of industry TFP growth.

The Z-factor adjusts the allowed change in rates for reasons other than inflation and productivity trends. The main rationale for Z-factors is to recover the impact that changes in government policy have on the company's unit cost. Absent such adjustments, government officials can change policies in ways that raise the company's unit cost, confident that there will be little or no effect on rates.

In the third generation incentive regulation plan approved for Ontario's electricity distributors, the rate indexing plan also included an incremental capital module (ICM). The ICM was an optional mechanism that distributors could access if the "core" inflation minus X formula did not provide enough revenue to finance the network's required capital investment. Distributors could only use the ICM to recover capital expenditures in excess of a threshold determined by a formula. The threshold values varied by company and were designed to ensure that the use of two separate rate adjustment mechanisms – a core inflation minus X formula, plus the ICM – did not lead to a double-counting and over-recovery of the distributor's capital spending.[39]

Indexed rate cap plans are multi-year rate adjustment mechanisms. In almost all cases, the plans are in effect for a defined period of time. When a given plan expires, rates at the end of the plan are typically reviewed using cost of service methods. This often leads to a rate "rebasing" in which rates are updated and become the base to which an updated, rate indexing formula is then applied.

Indexed base rate caps create strong performance incentives in two primary ways. One is that indexed rate caps are determined by inflation and productivity data that are "external" to the utility itself. This means that during the term of the rate indexing plan,

---

inputs. Such measures are constructed using sub-indexes that reflect trends in the prices of major input categories. The weights assigned to the inflation sub-indexes reflect the percentage each input category represents in utility cost. Peer price indexes are based on the prices charged by other service providers. For example, a peer price index for the bundled power service of a Midwestern United States utility might be constructed from the retail price trends of other Midwestern utilities.

[39] In addition to the threshold, company requests to use the ICM had to satisfy two other eligibiity criteria to be approved: 1) the amounts should be directly related to the claimed cost driver, clearly non-discretionary, and clearly outside the base upon which rates were derived; and 2) the amounts to be incurred must be prudent. See the *Report of the Board on 3rd Generation Incentive Regualtion for Ontario's Electricity Distributors, July 14, 2008.*

energy networks have very strong incentives to reduce their unit costs, since any changes in their company's unit cost that are less than the change in the PCI will go straight to the "bottom line" and boost company profitability. Indexed rate caps create incentives for utilities to improve increase productivity efficiency, as defined in Chapter Two.

In addition, indexed rate caps can create balanced incentives to control all costs. This results from the fact that indexed rate caps apply to both capital and O&M costs, and controlling capital and operating costs both contribute to improved productivity. This contrasts with a common critique of cost of service regulation, which is that it creates weak incentives to control capital vis-a-vis operational spending since, under rate of return regulation, utilities' earnings are linked directly to their regulated rate base. Hence, indexed rate caps can create incentives for networks to increase supply-side allocative effiency.

It should be noted, however, that these incentives are weakened by the rate rebasings that typically occur when an index-based rate plan expires.[40] Rates at these reviews are determined using cost of service methods. Accordingly, the well-known incentive problems with cost of service regulation re-emerge at rate rebasings. This means that while networks may have strong incentives to reduce unit costs while indexed rate caps are in effect, they may also have strong incentives to report cost increases for the year on which rebased rates will be established. This can make it difficult in practice to distinguish genuine cost reductions from cost deferrments (particularly deferred capital investment) over time. Section 5.5 will discuss incentive mechanisms that are designed to redress this concern.

5.2 Revenue Caps

Comprehensive revenue caps sometimes are also sometimes called revenue decoupling mechanisms. Under these types of plans, an electricity network's allowed revenues do not depend on its energy (*i.e.* kWh) deliveries. Revenue caps can be updated over time using indexing mechanisms. Revenue caps can also be established on a per customer basis, and adjusted over time to reflect growth in the number of customers served by the utility. In principle, revenue caps can also be constant and adjusted only in conjunction with rate rebasings.

---

[40] There may also be concerns associated with Z factor mechanisms that pass through costs into rates, although in practice this depends on how effectively the regulator evaluates the prudence of Z factor requests.

Indexed revenue caps are identical in many respects to indexed rate caps, except revenues rather than utility rates are subject to formula-based restrictions. Under a comprehensive revenue cap, indexing formulas apply to a company's regulated revenue, not its rates. Comprehensive revenue caps typically include some type of a balancing account mechanism to ensure that actual revenues are similar or equal to the index-based revenue requirement. The balancing account contains any mismatch between actual and allowed revenues until rates can be adjusted to eliminate the mismatch.

As with indexed price caps, indexed revenue caps typically apply indexing formulas that include an inflation measure, an X-factor, and a Z-factor. Compared with the formula presented earlier, a revenue cap index also often includes some adjustment to reflect the effect of output growth on cost and, hence, required revenues. An explicit term for such an adjustment may be called an output factor, which is denoted by *Y*. An index-based restriction on revenue requirement growth may then be written as:

$$\Delta\, \textit{Revenue Requirement} = \mathrm{P} - \mathrm{X} + \mathrm{Y} \pm \mathrm{Z}\,. \qquad [17]$$

The X and Y terms, as here described, are sometimes captured in a consolidated X.

Revenue caps can also be non-comprehensive. Non-comprehensive revenue caps adjust only a portion of the company's rates or revenue requirement. An example is a cap on the revenue requirement (allowed cost) for O&M expenses. As with comprehensive revenue caps, partial caps are usually developed using indexes, which include an inflation measure, X-factor, and Z-factor and should include an adjustment for output quantity growth. Unlike comprehensive indexing mechanisms, however, the X factor in non-comprehensive revenue adjustment mechanism would be established using estimates of partial factor productivity (PFP) growth rather than TFP growth. For example, the X factor in a formaul adjusting allowed O&M cost would be calibrated using an estimate of O&M PFP growth, rather than TFP growth, for the relevant network industry.

Comprehensive revenue caps create the same incentives for productive and supply-side allocative efficiency as indexed rate caps. The reason is that both use external data in the adjustment formula and both are comprehensive mechanisms that apply to all network costs. In addition, comprehensive revenue caps create stronger incentives to pursue conservation and demand management (CDM) objectives, since the mechanism recovers losses in network revenues that occur when customers reduce their energy consumption.

Compared with comprehensive mechanisms, non-comprehensive revenue adjustmetnt formulas can create weaker performance incentives. Networks are encouraged to reduce the costs associated only with the set of costs that are the subject of the plan, rather than comprehensive costs. Non-comprehensive mechanisms therefore do not create balanced performmance incentives or promote allocative efficiency, and regulators would need to assess the prudence of expenditures that are outside the mechanism while a non-comprehensive revenue indexing plan is in effect.

5.3 Earnings Sharing Mechanisms

Earnings-sharing mechanisms (ESMs), sometimes called "sliding-scale" mechanisms, adjust a company's allowed rates when its rate of return has been in an established range in a recent historical period. The mechanisms are set in advance of their use and typically function for several years. The most widely used earnings measure is return on equity (ROE). In Ontario, the OEB has approved the use of ESMs in some incentive regulation plans but not in others.

Approved ESMs can vary in several ways. One important difference is customers' share of surplus and/or deficit earnings relative to allowed (or target) ROE. Customers' share of earnings may differ for different levels of realized earnings. For example, the ESM may specify that customers receive 50% of the difference between actual and allowed ROE for all earnings that are within allowed ROE and 200 basis points above allowed ROE, but customers' share increases to 75% of earnings between 200 basis points and 400 basis points above allowed ROE. It is common for customers' share of incremental earnings to increase as actual earnings rise, although this is not always the case.

Many ESMs also feature what is often called a deadband. A deadband is simply a range around allowed ROE in which earnings fluctuations do not prompt rate adjustments. Accordingly, shareholders retain all earnings that are within the deadband.

Some ESMs are "symmetric" and allow both "excess" and "deficit" earnings to be shared with customers. In other cases, ESMs only share earnings that exceed the target ROE. ESMs can also feature earnings "caps" or floors. An earnings cap would be a maximum ROE. All earnings beyond this point would be returned to customers. An earnings "floor" is a minimum ROE beyond which earnings are not allowed to fall. If the ROE floor is breached, there may be automatic rate adjustments that restore earnings to the target level. Alternatively, breaching the ROE floor can lead to a hearing that establishes new rates.

Some ESMs also distribute benefits to shareholders and, particularly, customers in innovative ways. One approach that has been implemented by MidAmerican Energy has been for all or a portion of customers' share of earnings to be used to defray the costs of future investments. For example, utilities can book customers' share of earnings as a regulatory liability that is used to fund plant investments that are expected in the future.

ESMs have some important advantages in regulatory applications. One is their ability to mitigate risk and keep earnings within what are deemed to be acceptable bounds. This property is, of course, greater when ESMs are symmetric and apply to "under" as well as "over" earnings.

ESMs also share benefits *as they are realized* under the plan. This naturally leads to some alignment of shareholder and customer interests. It can also lead to less pressure on regulators to choose aggressive productivity stretch factors in order to share the benefits of incentive regulation, which can never be known with certainty *ex ante*.

On the downside, it must be recognized that ESMs are not inherently incentive regulation mechanisms. Indeed, compared with traditional cost of service regulation where utilities retain *all* the benefits of unit cost reductions between rate reviews, sharing earnings with customers *reduces* utilities' incentives to pursue productive and allocative efficiencies. ESMs will only be a form of incentive regulation if they extend the period of "regulatory lag" between rate reviews compared with the cost of service regulation.

ESMs also do not by themselves guarantee that customers benefit from an incentive regulation plan. Customers may complain if network earnings exceed the target ROE but are within the "deadband" and therefore do not reach the sharing range. In addition, ESMs increase the regulatory focus on annual earnings computations, which can exacerbate concerns with inherently controversial issues like utility-affiliate transactions and cost allocations. Earnings reviews associated with ESMs can raise regulatory costs and reduce net benefits to all stakeholders. The calculation and review of earnings in some approved ESMs has in fact been contentious.

5.4 Benchmark Regulation

5.4.1 Comprehensive Benchmark Plans

Benchmark regulation involves evaluating one or more indicators of company activity against external performance standards (or benchmarks). The standards are external to the extent that they are insensitive to the actions taken by the subject utility's managers. Evaluations and rate adjustments are accomplished by formal mechanisms that are established in advance and typically function for several years.

A benchmark plan's key features are the performance indicators, the performance benchmarks, and the rate adjustment (or award) mechanism. The performance indicators used in approved benchmark plans vary greatly in scope. As with revenue caps, there are both comprehensive and non-comprehensive benchmark plans. A comprehensive benchmark plan is one in which benchmarking mechanisms cover substantially all the facets of company performance that matter to customers. Comprehensiveness can be achieved by having many indicators that cover separate performance dimensions, or by having fewer but more broadly focused indicators.[41] In contrast, a non-comprehensive benchmark specifically targets a small set of performance areas.

The performance benchmarks used in benchmark plans are also varied. A common benchmark is a company's average performance on the selected indicator in a period just prior to plan commencement. A company is therefore rewarded or penalized for its performance relative to recent history.

An alternative approach, which is an example of "yardstick regulation" or statistical benchmarking, is to use the corresponding performance for a group of "peer" utilities. The utility is therefore rewarded or penalized based on its performance relative to a group of comparable firms. The peer can be defined to be all utilities in the same region as the company subject to the plan. The peer group in this case would be a proxy for the regional industry. In principle, peers can also reflect the entire national industry.

The rate adjustment mechanisms in approved benchmark plans vary. One approach is to adjust rates directly based on the relationship between measured and benchmark

---

[41] Examples of broad-based performance indicators include retail price indexes, unit cost indexes, and TFP indexes.

performance. An alternative is to adjust allowed returns rather than rates. For example, allowed returns can increase when the utility's measured performance exceeds the benchmark levels, and allowed returns can be reduced if measured performance falls below benchmark levels. This approach can lead to rate changes indirectly, especially if the benchmark plan is integrated with an ESM. In this application, the allowed ROE "floats" with the utility's performance, and the ESM then operates according to pre-established formulae that depend on the relationship between the utility's actual and allowed returns.

In all benchmark plans, a major design issue is the customer sharing percentage. Another is whether the mechanism features deadbands in which fluctuations around the benchmark do not induce rate adjustments.

Benchmarking plans sometimes provide supplemental adjustments to rates rather than serving as the sole basis for rate changes. Several rate adjustment mechanisms can, in principle, coincide with a benchmarking plan. At one extreme, rates may be adjusted for the actual trend in a company's unit cost. At the other, rates may be predetermined for several years.

Comprehensive benchmark plans can in principle create strong performance incentives. Plans are comprehensive and can therefore promote allocative as well as productive efficiency. Plans can also incorporate service quality and cost control objectives in a single mechanism.[42]

On the other hand, comprehensive benchmark plans can be challenging to design and implement. Addressing these challenges could raise the costs of implementing and administering such plans. Benchmark plans can be designed in many different ways. Some of these plan designs will involve comprehensive benchmarking evaluations, which requires decisions to be made over the appropriate choice of benchmarking techniques and appropriate values or targets to be reflected in the comprehensive performance standards.

---

[42] For example, this has been true of several incarnations of the Performance Evaluation Plans (PEP) for Mississippi Power, which have been in effect since 1986.

5.4.2 Non-Comprehensive Benchmark Plans

Non-comprehensive benchmark plans are similar in many respects to comprehensive benchmark plans. They also involve performance indicators, performance benchmarks, and award mechanisms. The main difference is that a non-comprehensive plan does not cover all dimensions of company performance.

Traditionally, many benchmark plans for energy utilities have been non-comprehensive and include a few narrowly focused performance indicators. Indicators measuring fuel procurement, generator management, and demand-side management (DSM) have historically been common. For example, in a 1986 survey on incentive regulation, Joskow and Schmalensee identified forty-three plans targeting the peformance of power generation plants in nineteen states.[43] In the gas distribution industry, there are numerous approved benchmarking plans for gas procurement cost.

Probably the most important set of non-comprehensive benchmark mechanisms are service quality incentives (SQIs). SQIs are a form of non-comprehensive benchmark regulation that rewards or penalizes a utility depending on the relationship between its measured service quality and quality benchmarks.

There have also been some intriguing applications of non-comprehensive benchmark plans that have been applied overseas. For example, Norway has implemented an innovative approach to benchmarking power distribution reliability. Beginning in 2001, prices for each distributor were adjusted to include an allowance for “energy not supplied” (ENS). The ENS measure is analogous to SAIDI, and an *expected* value for ENS was determined for each distributor in the country. The expected ENS was generated using an econometric model in which ENS is a function of a variety of business condition variables, including weather and the length of the network. Model parameters were estimated using historical data from the Norwegian power distribution industry. Each company’s expected ENS was then determined by multiplying the parameter estimates by the average values of the business condition variables expected for a given company.

Each year, the distributor’s annual ENS is compared to the benchmark, expected value. This difference is then multiplied by the value of reliability. This valuation is also

[43] Paul R. Joskow & Richard Schmalensee (1986), *Incentive Regulation for Electric Utilities*, 4 YALE Journal of Regulation.

tailored to each distributor to reflect its customer mix. If the difference is positive (i.e. reliability has been better than expected), it is added to the company's capped revenue for the following year. If the difference is negative (i.e. reliability has been worse than expected), it is subtracted from the company's capped revenue for the following year.

Another interesting benchmark application has been applied to network investments. Several jurisdictions have used engineering-based methods to simulate the costs of an efficient "reference network" for specific utilities. The models apply engineering based principles and costs to information on the locations of customers and loads in a specific territory, and generate estimates of the cost of constructing an efficient network to provide the necessary outputs. This engineering-based benchmark is then used to evaluate investment requirements for individual network utilities. This model has long been used in Chile and has since been adopted in a number of other Latin American countries. Engineering based models have also been developed and applied in Spain and, most recently, Sweden.

5.5 Plan Termination Provisions

Plan termination provisions are another, increasingly important aspect of incentive regulation. One important consideration here is simply the term of the incentive regulation plan. Provisions that are established for resetting rates when the plan ends are also important.

There has been a trend in incentive regulation towards longer plans. Three-year plans were typical during the 1990's. More recently, five-year terms have become standard, and some plans have been approved with considerably longer terms, including the ten-year plans for the power distribution services of National Grid in Massachusetts and New York. As part of its "RPI-X@20" review, in October 2010 the UK also announced that it plans to extend the term of plans for electricity distributors from five to eight years.

Longer plan terms clearly strengthen performance incentives. Longer terms are especially useful in encouraging initiatives that involve up-front costs to achieve long-run efficiency gains. That is one reason why longer plan terms are often approved for utility mergers. Both National Grid plans mentioned above involved mergers.

On the downside, longer plan terms can increase both business and regulatory risk. This makes them less suitable for businesses undergoing rapid change or trends that are not otherwise captured in the design of the incentive regulation plan. However, these risks may be mitigated by allowing utilities to access optional “modules“ such as the ICM that allow them to address company-specific circumstances.

As discussed, most incentive regulation plans (including those approved to date in Ontario) include a full cost of service review when the plan expires. This is not always the case, however. There have been examples where the rates established at the end of an expired incentive regulation plan were computed by using information on the company’s cost of service *and* an external benchmark. This was true with seveal plans approved for National Grid in the Northeastern US. In each, rates were rebased using a weighted average of the network’s cost of service and a total cost benchmark. Since rates were not entirely “trued-up“ to the company’s cost of service, utilities were allowed to keep some of the benefits associated with superior performance even beyond the term of the incentive regulation plan.

A different, innovative approach towards rate updates has been applied in some UK and Australian plans. In some cases, regulators have created “efficiency carryover mechanisms“ that allowed estimated efficiency gains achieved in an incentive regulation plan to be distributed to customers in increments over the term of the successor plan. This is done by computing company-specific cost benchmarks for operating and capital costs for each year in a five year indexing plan, before that plan takes effect.[44] These operating and capital cost benchmarks are determined as part of the same regulatory review that establishes allowed rate changes for the successor incentive regulation plan. While the plan is in effect, the regulator then monitors the network’s actual operating and captial expenditures and computes the difference between actual and benchmark costs. For each cost category, the difference between the actual and benchmark cost is the measured “efficiency,“ which is then distributed to customers as rate reductions in five equal increments over the next five years. For example, efficiencies in year one of a plan are phased out in years two through five of the current plan, and year one of the next plan. Efficiencies in year two of a plan are phased out in years three through five of the current plan, and years one and two of the next plan. This mechanism enables the efficiencies associated with cost savings to be retained by the network for exactly five years, regardless of the year those efficiencies were realized.

[44] For capital, the benchmarks are actually the carrying costs (cost of funds plus depreciation) associated with capital expenditures during each year of the plan. This allows actual capital expenditures (capex) to be compared with the projected, benchmark capex *ex post*.

Rate reset provisions are important because they affect the incentives created by the regulatory mechanism. If a full cost-based rate true-up can be avoided, performance incentives will be strengthened. Regulatory costs may also be reduced because, in principle, incentive-based rate reset provisions reduce concerns that firms are simply deferring cost increases and may be "gaming" the costs reported in the test year used to set rebased rates. PEG has developed an "incentive power" model that simulates the impact of different regulatory designs on utilities' behavior, and our results show that allowing partial cost of service true ups when plans expire has a very powerful impact on the strength of incentives. These mechanisms are especially important for encouraging efficient investment behavior and initiatives that involve up-front costs and long term benefits.

Indeed, if regulation is designed to encourage greater focus on long-term outcomes, then the rules governing how rates are set when multi-year plans expire become even more critical. Utilities will not invest in initiatives with payback periods (*i.e.* the time it takes for the present value of future financial benefits to exceed the upfront cost of the initiative) that exceed the term of the incentive regulation plan unless there are regulatory assurances that they will be allowed to retain future benefits. In practice, this means that utility time horizons on some investments are naturally limited by the term of the multi-year plan. Efficiency carry-over mechanisms or partial true-ups can extend companies' effective time horizons and may thereby be effective in promoting longer-term thinking and long-term "value for the money."

One mechanism for rewarding utilities for creating long-term outcomes could be to capitalize the value of savings measured by an efficiency carry-over mechanism or partial true-up plan update. The utilities' regulatory asset value would then reflect both the net value of dollars invested (*i.e.* a measure of capital inputs) as well as a measure of long-term value the utility has provided to customers (*i.e.* a measured regulatory outcome). This latter value could be a measure of long-term efficiency gains or, in principle, a measure of the social benefits the utility has generated by complying with GEA mandates in an expedited and effective manner. Although we are not aware of any precedents for this type of mechanism, linking utilities' capital stocks (and therefore long-term company earnings) to measures of the long-term value created by their actions should be a strong motivator and encourage longer-term planning and initiatives.

On the other hand, the details and empirical foundation of any such mechanism would clearly require careful scrutiny. Customer groups and Staff members may understandably be reluctant to allow less than complete rate true-ups unless (or until) they fully understand the potential benefits for customers as well as shareholders. Incentive-based plan termination provisions deserve serious consideration going forward, but if there is any interest on the part of stakeholders these mechanisms will clearly need to be investigated and vetted carefully.

5.6 Menu Approaches

Many economists have been interested in using "menu" approaches in utility regulation for some time. The basic idea behind a regulatory menu is that the regulator would develop of series of alternative regulatory options, present these options to the utility, and the utility would be allowed to select the approach that best suits its circumstances. It is argued that a menu approach allows regulation to accommodate the diverse needs and circumstances of different utilities in an efficient manner.

In spite of the theoretical interest, there have been few examples of approved "menu" approaches. Perhaps the best example is the Information Quality Incentive (IQI) which has been approved several times for energy networks in the UK. The IQI is a type of menu in that it allows for a variety of different ways for translating utilities' capital plans into allowed rates, according to rules that are specified in a mathematical matrix. However, the IQI is more complex than a simple menu, and it requires detailed, company-specific benchmarks to be implemented. The basic structure of the IQI and its application to UK electricity distributors is described in Appendix Four.

There has also been some interest in the menu approach in Ontario. For example, in the first generation incentive regulation plan for electricity distributors, the Staff Draft Rate Handbook recommended an inflation minus X plan where distributors were allowed to select their X factor using a menu approach. Under this approach, a menu of six alternative X factor and allowed return on equity (ROE) combinations was developed, with higher values for X associated with higher allowed ROE levels and *vice versa*. Companies would then be allowed to select the X factor- ROE combination that most appealed to their risk-incentive preferences.

In January 2000, however, the OEB rejected this approach as too complex for a first generation PBR plan. It also did not believe that there was a well-developed analytical foundation supporting the specific menu of X factor and ROE combinations. Instead of this menu approach, the OEB opted for a more conventional PBR plan where the inflation minus X formula used a single X factor.

Some stakeholders also recommended an alternate menu approach be established in the third generation incentive regulation plan. However, this proposal was rejected by stakeholders during the review process for two reasons. First, the menu proposal assumed that the PBR plan would include an earnings sharing mechanism (ESM), which the Board did not approve because it of its expressed confidence in the values for the productivity and stretch factors and because of the regulatory burdens associated with earnings calculations. That decision essentially eliminated the viability of the proposed menu.

Second, and more fundamentally, during the working group consultations with stakeholders, there were significant unanswered questions regarding the design of the proposed menu. One basic concern is that it was never explained how permitting companies to choose from a menu would necessarily benefit *customers*. If companies are presented with a variety of regulatory options, they will clearly select the alternative that is expected to be most profitable. However, this is not sufficient for an appropriately-designed incentive plan, which should lead to "win-win" outcomes for companies and customers. It was never clear that the proposed menu would lead to win-win outcomes.

Relatedly, it was not clear whether and how the X factor-ROE tradeoffs presented in the menu were reasonable. This is a critical issue. In any menu approach, the menu options must be calibrated so that the party selecting from the menu will be induced to select an alternative that benefits customers and shareholders alike. Whether these incentives are created depends on the linkages between the variables that are paired for each option, as well as the linkages among the different items on the menu. Evaluating these relationships and their implications for customer and shareholder welfare is likely to be complex, and these issues were not addressed in the proposal. Design considerations like these probably illustrate why menu approaches towards regulation are appealing in theory but much less common in practice.

Menu approaches have now been proposed but not approved at least twice in Ontario. The menu approach is intriguing in theory but poses significant design challenges. Those complex issues would need to be addressed if a menu approach is to be considered again.

5.7 Targeted Incentives to Achieve Regulatory Outcomes

Essentially all the incentive mechanisms discussed above (indexed rate caps, revenue caps, benchmark mechanisms and the plan termination provisions) are designed to encourage and reward excellence. The underlying rationale for each is to motivate utility managers to be innovative and discover new ways to provide existing utility outputs more efficiently. This is naturally an important regulatory objective and necessary to comply with the Board's economic regulation obligations.

However, incentives mechanisms can also be targeted at achieving very specific regulatory outcomes. More targeted incentives may be more appropriate for rewarding utilities for achieving one-time, discrete regulatory objectives. Examples of such discrete regulatory objectives include quickly connecting a renewable electricity generator in a distributor's territory, or adding transmission capacity in a given area to eliminate transmission congestion. These are important regulatory outcomes that can be identified and targeted fairly easily. While it is in the public's interest for these objectives to be realized as expeditiously and effectively as possible, the incentive mechanisms appropriate to achieving these targeted aims could be different from those promoting "discovery" and performance excellence more generally.

Incentive-based approaches focused on targeted regulatory outcomes would include what the Staff called "incentive mechanisms" in its Discussion Paper on the Regulatory Treatment of Infrastructure Investment. These incentive mechanisms are an incentive-based ROE and project-specific capital structures, for new investments that benefit customers. Staff "has classified these two mechanisms in particular as "incentives" because they provide "cost plus" compensation to the regulated entity for its investment."[45] These incentive mechanisms would appear to be especially appropriate when applied to relatively narrow investments that are designed to achieve well-defined, targeted regulatory objectives.

---

[45] Staff Discussion Paper, *op cit*, p. 25.

5.8 Regulatory Judgment/Discretion

In addition to incentive mechanisms, the Board can use discretion and judgment as a means of creating incentives for networks to achieve regulatory objectives. One example is the use of prudence reviews in cost of service proceedings. The OEB always applies a certain amount of judgment when assessing the prudence and reasonableness of networks' reported costs. The Board can scrutinize networks' costs and operations more closely if certain performance measures fail to conform with established standards. For example, Staff can apply greater scrutiny to the cost of service applications of networks in the bottom third of the benchmarking evaluations that are currently used to set stretch factors in third generation incentive regulation.

Alternatively, networks could be "fast tracked" for regulatory approvals if their measured performance satisfies certain standards. The OEB can condition approvals of cost of service applications on the quality and thoroughness of information provided by the applicant, as well as the network's performance on certain cost and quality measures. In the UK, Ofgem has established a kind of fast-tracking of the applications of companies if their rate applications satisfy similar criteria. These changes are an important part of the new Revenue using Incentives to deliver Innovation and Outputs (RIIO) framework that Ofgem is implementing to refine and supplement the RPI-X approach that it has administered for the last 20 years.

It should be noted that the current benchmarking studies do not pertain to capital spending. However, similar benchmarking approaches may be able to be developed for capital spending as well. Ideally, these models would be developed and estimated using Ontario data, but the feasibility of such an approach depends largely on data availability. Nevertheless, even if the information is simply used to inform regulatory judgment, additional "external" data on capital data could aid Staff and the OEB as they assess networks' capital spending proposals.

6. Evaluating Performance

Over the course of this project, Staff will need to recommend a framework for applying the concepts discussed in this Paper to Ontario's regulatory challenges. More precisely, Staff needs to develop a framework that brings together performance measures, performance standards, performance evaluation techniques, and regulatory mechanisms in a manner that is most likely to promote desired regulatory outcomes. Moreover, this framework must be adaptable to the variety of regulatory approaches that the Board will continue to administer. These approaches include multi-year incentive regulation approaches and cost of service rate reviews, and perhaps assessing company-specific proposals as well.

This chapter presents some initial, and highly preliminary, ideas on how such a framework can be developed. These thoughts are primarily intended to facilitate discussion and prompt alternate ideas from other parties. Clearly, the issues to be addressed in this consultation are complex and require substantial input from stakeholders. It is hoped that the concrete performance evaluation/regulatory consequences framework presented here will stimulate more ideas and better informed alternative proposals than would result in the absence of such a "strawman" proposal.

We begin by briefly discussing what is known as the RIIO Model under development in the UK. The acronym "RIIO" stands for (setting) Revenues using Incentives to deliver Innovation and Outputs. We then turn to an illustrative, strawman framework that brings together performance measures, standards, empirical techniques and regulatory mechanisms to evaluate utility performance and use these performance evaluations to promote regulatory outcomes.

6.1 RIIO

The RIIO model grew out of the Office of Gas and Electricity Markets' (Ofgem's) review of the UK's longstanding "RPI – X" regulatory approach. This review was undertaken during the 20th year anniversary of the year in which RPI-X controls were applied to electricity networks. In October 2010, this "RPI-X@20" review led to the announcement of the new RIIO regulatory framework, which Ofgem describes as follows:

> The RIIO model has taken the elements of the old RPI-X framework that work well, adapted other elements to ensure that they are focused on delivery of a sustainable energy sector and long-term value for the money, and added elements to encourage the radical measures needed in innovation and timely delivery. The model is designed to promote smarter gas and electricity networks for a low carbon future.[46]

The RIIO model is fairly complex, and a complete description and analysis of this approach goes beyond the scope of this Report. The Staff's Discussion Paper on Defining and Measuring Performance of Electricity Transmitters and Distributors provides some additional discussion on important features of RIIO. It should be noted, however, that implementing RIIO is still a work in progress and, apparently, many details have still not been finalized.

Even though RIIO is still evolving, it is natural for Staff to consider whether it has any implications for this consultation, since RIIO has been motivated by many of the same concerns that led to the Board's creation of a Renewed Regulatory Framework for Electricity. In both cases, regulatory reform was prompted by broader public policy changes that put greater emphasis on promoting "smart" networks and a green energy economy. Regulators in both the UK and Ontario are attempting to meld this new, green energy emphasis and associated mandates with their traditional economic regulation functions. RIIO represents the outcome of Ofgem's consideration of these issues and the framework it is establishing to cope with its expanded regulatory challenges.

There are some intriguing ideas in the RIIO model that merit consideration in this proceeding. One is simply the notion of linking revenues to outcomes.[47] RIIO will attempt to link networks' revenues to measures of the value they provide to customers. This includes the value of traditional, or what RIIO terms "bread and butter," outcomes like reliability, safety and new connections. Revenues can also be linked to "sustainable development" outcomes such as customer satisfaction and environmental impact. RIIO will bring measures of these outcomes together in a "balanced scorecard" which records how effectively networks

---

[46] Office of Gas and Electricity Markets, *Handbook for Implementing the RIIO Model*, October 2010, p. 1.

[47] In RIIO, Ofgem's use of the term "outputs" is analogous to our use of the term "outcomes." RIIO did not distinguish between these concepts, but our discussion puts more emphasis on measurement and empirical estimation techniques than is evident in Ofgem's discusison of RIIO, and in this context it useful to make this distinction.

deliver these desired regulatory outcomes. This balanced scorecard approach can facilitate comparisons among networks and thereby inform Ofgem's ratemaking.

One concrete way networks can benefit from delivering perceived value for the money is for their regulatory applications to be "fast tracked." A fast-tracked regulatory application would be subject to less regulatory scrutiny than typical regulatory filings, which would in turn allow for an early decision for the network. A fast-tracked regulatory application should reduce regulatory costs for networks and Staff. Fast-tracked applications may also reduce regulatory risk, since the very act of deciding that a network is worthy of "fast tracking" may be a signal of the general reasonableness of the costs and information presented in a company's regulatory filing.

Under RIIO, much of Ofgem's evaluation of company applications will revolve around each network's long-term business plan. These plans will present information on each network's targets for delivering outputs/outcomes over a multi-year horizon for a given level of projected cost. Ofgem will carefully review the plans to see if they are "well justified" using a variety of empirical evaluation tools. Ofgem's assessment "tool kit" is reproduced in Appendix B of the Staff report, where different assessment techniques are arrayed in order of increasing regulatory scrutiny. However, the assessment tools to be applied will be determined on a case-by-case basis that is not known in advance by the networks. Ofgem believes it is appropriate for companies not to know what regulatory tools may be applied to their applications, since this would encourage networks to submit plans that are well-justified as a whole. If companies were aware of the tools to be used, they could be incented to display good performance and value for the money only on a few selected performance measures rather than submitting an overall, well-justified long-term business plan.

While there is some merit in Ofgem's discussion of the need for discretion in regulatory evaluations, there are also some concerns associated with their approach (at least as it has been publicly explained so far; as discussed, RIIO is a work in progress and is still evolving). One is that the absence of a well-developed and transparent evaluation framework runs the risk of becoming highly subjective. Networks would in principle benefit from having a clear understanding of the regulatory outcomes they are expected to achieve as well as the regulatory consequences that result from either succeeding or failing to comply with these expectations. A well-defined, transparent and objective framework should enhance stability and promote more effective long-run planning and behavior assuming, of course, that the framework has a sound conceptual and empirical foundation and will thereby encourage efficient, long-run behavior.

Second, the RIIO approach appears extremely information-intensive. Companies must prepare detailed, long-run business plans and convince the regulator that their plans are especially "well justified" in order to be rewarded. Linking rewards to companies' submitted business plans could lead to a perverse kind of tournament, where networks essentially compete against each other with respect to planning and regulatory persuasion. This could lead to an inordinate amount of effort and resources directed towards the planning process and crafting extremely detailed and captivating business plans, rather than serving the needs of network customers. Not to be glib, but there is a danger that RIIO's emphasis on the well-justified, long-term business plan could unintentionally lead to a regulatory regime that "pays for planning" rather than "pays for performance."

RIIO also imposes information burdens on the regulator. Ofgem will have an extensive assessment toolkit at its disposal that includes detailed reviews of company plans, a battery of sophisticated industry-wide and international benchmarking models, and a variety of engineering, financial and "market testing" tools. Any of the techniques from the toolkit can potentially be deployed in any given regulatory review. By emphasizing regulatory discretion and an extensive array of assessment tools, RIIO may increase the complexity and costs of both the regulator's work and the regulatory process.[48]

If a RIIO-type framework was established in Ontario, these information burdens would be greater in the Province than in the UK. There are only a dozen or so electricity distributors in the UK, while there are more than 80 in Ontario. This naturally means that OEB Staff would need to evaluate more than six times as many long-term business plans as Ofgem in order to implement a RIIO-type regime. It should also be noted that the Staff's principles for implementing a Renewed Regulatory Framework include predictability and transparency. Complying with these objectives would likely involve some modifications of the RIIO approach that has been articulated to date in the UK.

---

[48] In practice, regulators and Staff could control these costs simply by choosing simpler and less costly assessment techniques, but this could also potentially reduce the effectiveness of the regulatory review.

6.2 Illustrative Framework

This section will outline a framework (which incorporates some ideas from RIIO) that could be used to assess electricity networks' performance and link regulatory consequences to firms' measured performances Because this example is only for illustrative purposes, we do not provide extensive commentary on every element of the framework. It is expected that there will be considerable refinement of the framework during the course of the project, and this example is only designed to initiate these discussions.

The framework begins with the Board specifying the desired regulatory outcomes to be achieved. This will include outcomes related to the Board's traditional functions, as well as outcomes related to its new responsibilities. For illustrative purposes, we specify four outcomes associated with both traditional and new network outputs, for a total of eight desired outcomes. They are:

Traditional Network Outputs

1. Efficient utility operations; using the concepts defined in Chapter Two, this outcome is consistent with promoting static productive efficiency.
2. Efficient capital expenditures; using the concepts defined in Chapter Two, this outcome is consistent with promoting dynamic productive efficiency.
3. Balanced cost containment incentives; for example, utilities should have incentives not to "gold plate" their networks in order to reduce operating costs. Utilities should also have balanced cost containment incentives throughout the term of a multi-year incentive plan. This means, for example, that networks should not be "harvesting" their existing assets by conserving on capital expenditures while an incentive regulation plan is in effect, and then concentrating those deferred investments in a base year that is used to set new, cost-of-service based rates. Using the concepts defined in Chapter Two, this outcome is consistent with promoting supply-side allocative efficiency.
4. Maintaining system reliability and appropriate customer service; using the concepts defined in Chapter Two, this outcome is consistent with promoting demand-side allocative efficiency.

New Network Outputs

1. Efficient capital planning, including plans to connect renewable generators in a network's territory; using the concepts defined in Chapter Two, this outcome is consistent with promoting dynamic productive efficiency.
2. Connect renewable generators quickly and efficiently; using the concepts defined in Chapter Two, this outcome is consistent with promoting static productive efficiency.
3. Develop an operational smart grid; using the concepts defined in Chapter Two, this outcome is consistent with promoting static and dynamic productive efficiency and demand-side allocative efficiency.
4. Promote cost effective conservation and demand management (CDM); using the concepts defined in Chapter Two, this outcome is consistent with promoting demand-side allocative efficiency.

Next, the Board would establish specific performance measures associated with a given outcome. These measures would be monitored at cost of service reviews and/or during incentive regulation plans. Many performance measures discussed in this Paper are appropriate in both types of regulatory applications. For example, growth in industry TFP indexes are clearly appropriate for setting the terms of multi-year, index-based incentive regulation plans. TFP level indexes can also be used to assess the performance of a network utility, relative to similar TFP levels indexes for industry aggregates or selected peers, in a cost of service review. These comparisons could provide evidence on the productive efficiency of the utility in question and thereby inform Staff's assessment of the prudence of its costs in the rebasing year. All else equal, relatively good TFP performance for a firm would be evidence that the firm is relatively efficient and therefore that its base year costs are reasonable. Similarly, the O&M econometric and unit cost peer benchmarking studies that PEG developed are currently being used to set stretch factors in distributors' third generation incentive regulation plans. These models could also be used to assess distributors' O&M costs in the base year that is used to set updated, cost-based rates. Relatively good performance on these models would be evidence that base year O&M costs were relatively efficient and therefore reasonable.

We have proposed a number of different performance measures for each of the eight regulatory outcomes. This information is summarized in the second column of Table One. For discussion purposes, we propose the following measures for the associated objectives:

Traditional Network Outputs

1. Efficient utility operations
   - Industry TFP trends
   - Econometric benchmarks of opex cost
   - Unit cost and PFP benchmarks for opex
2. Efficient capital expenditures
   - Industry TFP trends
   - Econometric capex benchmarks
   - Possibly, each company's proposed capital investment plan
3. Balanced cost containment incentives

   Either
   - Actual opex and
   - Actual capex (with separate benchmarks for each)

Or

   - Change in total cost

   The former set of measures would be used with a UK or Victoria-style efficiency carry-over mechanism; the latter would be used in a National Grid-style plan termination provision.

4. Maintaining system reliability and appropriate customer service
   - SAIFI
   - SAIDI
   - Call service and customer service measures currently monitored by the Board

In each instance above, the measures can be applied in either an incentive regulation or cost of service regulatory application.

New Network Outputs

5. Efficient capital planning
   - Proposed investment plan?
6. Connect renewable generators quickly and efficiently
   - Total connection requests
   - Average time to connect
7. Develop an operational smart grid
   - Total spending/number of meters installed?
8. Promote cost effective conservation and demand management (CDM)
   - Energy savings from DSM plans
   - Changes in load shape/demand-shifting?

The next step is to specify performance standards and techniques used to evaluate performance for each of the measures. Our proposed standards and performance techniques are listed directly opposite the associated measures in columns three and four of the Table. In some cases, we have established separate "average" and "desirable" standards. In other instances, there is considerable uncertainty about what the best choices are for standards and evaluation techniques, so we have saved these decisions for future stakeholder deliberations.

The fifth column lists regulatory mechanisms that can be used to promote the given objective. These mechanisms make use of the performance measures, standards, and techniques listed in the three columns to the left. In some cases, these mechanisms are implemented using information derived from multiple performance assessments. For example, the X factor in a multi-year plan has two components: a productivity factor based on a measured TFP trend for the network industry; and a company-specific stretch factor based on benchmarking assessments of each network's cost performance.

This strawman framework could be populated with additional information or amended entirely. It is presented here only to provide a concrete example of how the concepts discussed in this Report can be brought together in a coherent manner. It also illuminates the linkages between outcomes, measured performance variables, defined performance standards, performance assessment techniques, and final regulatory mechanisms.

Table One

Performance Evaluation Framework

	Desired Regulatory Outcomes	Performance Measures	Performance Standards	Performance Assessment Techniques	Regulatory Mechanisms
"Traditional" Utility Outputs	Efficient utility operations (static productive efficiency)	Industry TFP Econometric opex benchmark Opex PFP and Unit Cost levels	Industry average TFP trend Acceptable: "Average" benchmarking performance Desired: "Superior" benchmarking performance Peer group performance	Industry TFP trend indexes Econometric benchmarking Unit cost and PFP benchmarking	Inflation minus *x* price controls *x* = Industry TFP trends and company specific stretch factor
	Efficient capital expenditures (dynamic productive efficiency)	Industry TFP Company capex projections Proposed investment plan?	Industry average TFP trend Acceptable: "Average" benchmarking performance Desired: "Superior" benchmarking performance ?	Industry TFP trend indexes Econometric capex projections IQI?	Inflation minus *x* price controls *x* = Industry TFP trends and company specific stretch factor Incremental Capital Module (ICM)
	Balanced cost control incentives (supply-side allocative efficiency)	Opex Capex Change in total cost	Ex ante projections OR Plan Termination provisions comparing actual cost to revenues under continued application of incentive mechanism	Compare measured performance to performance standard(s) >> both IR and COS	Inflation minus *x* price controls *x* = Industry TFP trends and company specific stretch factor Efficiency carry-over mechanism.
	Maintain reliability and customer service (demand-side allocative efficiency)	SAIFI SAIDI Call Center performance	Company-specific average on selected measure	Compare measured performances to historical average benchmark Econometric benchmarking of reliability? >> both IR and COS	Monitoring with potential "fast track" for superior performances. Longer-term, penalty/reward mechanism
"New" Utility Outputs	Efficient capital planning (dynamic productive efficiency)	Proposed investment plan	?	?	Monitoring with potential for "fast track" for superior performance
	Connect renewable generators (static productive efficiency)	Requests to connect generators Average time to connect	Acceptable: Industry average Desirable: Upper quartile	Compare measured performance to industry average and quartiles	Monitoring with potential for "fast track" for superior performance
	Operational Smart Grid (dynamic productive efficiency and static allocative efficiency)	Total spending/time to conncet smart meters	?	?	Monitoring with potential for "fast track" for superior performance
	Promote CDM (dynamic allocative efficiency)	Energy savings Changes in load shape	Acceptable: Industry average Desirable: Upper quartile	TRC computations	DSM incentives Monitoring with potential for "fast track" for superior performance

Pacific Economics Group Research, LLC

7. Going Forward

This Chapter briefly discusses some important issues that will need to be addressed in this consultation going forward. We begin by reviewing the empirical and regulatory foundations that currently exist in Ontario and which can be built upon and integrated in a holistic, comprehensive framework. We then discuss several key implementation issues and/or decisions.

7.1 Foundations in Ontario

A number of projects conducted with Staff over the last five years provide a strong conceptual and empirical foundation for the current consultation. The Staff Discussion Paper provides an overview of many of these initiatives. Some of the data and empirical tools developed in this work could provide a foundation for further applied work going forward, in particular:

- The gas incentive regulation mechanisms examined in 2006-07 developed a number of innovative productivity measurement techniques, including the use of both econometric and index-based TFP trends. Obviously this work was specific to gas distribution and not electricity networks, but some of these performance measurement methods could potentially be applied to electricity transmission and distribution.[49]
- The operations and maintenance benchmarking work in 2007-08 developed both peer-based O&M unit cost and O&M partial factor productivity (PFP) benchmarking models and econometric O&M cost benchmarking models.[50]
- The Third Generation Incentive Regulation Mechanisms (3rdGenIRM) developed in 2007-08 estimated a long-run TFP trend for electricity distributors, using TFP growth for US electricity distributors between 1988 and 2006 as a proxy. The 3rdGenIRM also used the O&M benchmarking studies discussed above to identify three efficiency "cohorts" within the Ontario electricity distribution industry and recommended different

[49] Lowry, M *et al*, *Rate Adjustment Indexes for Ontario's Natural Gas Utilities*, November 20, 2007.
[50] Lowry, M *et al*, *Benchmarking the Costs of Ontario Power Distributors*, March 20, 2008.

productivity stretch factors for different cohorts, with lower stretch factors for relatively more efficient distributors.[51]

This empirical work can obviously be useful for a performance evaluation framework. TFP trends, econometric benchmarking, and index-based benchmarking relative to peers can be useful tools for promoting a number of desired regulatory outcomes. Each of these tools can also be applied in both incentive-based regimes (*e.g.* 3rdGenIRM) and cost of service reviews.

However, the tools used in 3rdGenIRM can be enhanced in several ways. Most importantly, it would be preferable to rely on TFP trends that were computed entirely from Ontario data rather than based on US proxies. It would also be desirable to develop total cost benchmarking studies instead of simply O&M benchmarking studies. It was not possible to develop long-term TFP trends, and total cost benchmarking models, for Ontario distributors because of the lack of time series data on Ontario's capital stocks and capital expenditures. These data would be necessary to develop estimates of capital input and capital costs. In its 3rdGenIRM Decisions, the Board signaled the importance of developing Ontario-specific TFP trends in future incentive regulation proceedings, which would in turn require greater investigation into available, historical series on capital stocks and expenditures for Ontario electricity networks.

- Service Reliability Standards were investigated in 2009-2010. The Board has recently accepted Staff's recommendation that service reliability measurement issues be investigated fully in the Province, as a prelude to potentially developing standards in the near future.[52]

---

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

[52] Kaufmann, L., *et al, System Reliability Regulation: A Jurisdictional Survey,* May 2010; Staff Report to the Board, *Electricity Distribution System Reliability Standards,* EB-2010-0249, March 31, 2011.

7.2 Further Issues to Address

Notwithstanding the work that has been done to date, developing a holistic and comprehensive performance evaluation framework would involve some challenges. Among the most important issues that would need to be addressed are the following:

- As discussed, the availability of capital stock and expenditures data in Ontario would have to be explored, to ascertain the quality and extent of available data
- Following from the capital data analysis, it would be necessary to explore the feasibility of developing long-run TFP trends, and total cost benchmarking models, for Ontario's electricity networks, and whether the quality of these TFP and total cost measures are sufficient for regulatory applications
- Stakeholders and Staff would need to consider the merits of alternative benchmarking approaches in the context of the Renewed Regulatory Framework, and whether and how the models currently used by Staff should be changed (other than changing the application of the models from O&M costs to total costs)
- Stakeholder and Staff should consider how best to define "appropriate" and "desirable" performance standards associated with different regulatory objectives
- Stakeholders and Staff should carefully consider the merits of efficiency carry-over and plan termination provisions, and whether these mechanisms can play a useful role in promoting the Board's regulatory objectives
- Stakeholders and Staff should consider whether the available reliability data in Ontario are of sufficient quality to establish reliability performance standards
- There are significant challenges associated with developing performance measures, standards and evaluation techniques related to the Board's new duties under the GEA. This appears to be particularly true with respect to promoting efficient investment planning by networks, promoting timely and efficient connection of renewable generators, and encouraging the development of a smart grid. Our preliminary, "strawman" evaluation framework did reveal that a significant amount of additional work is necessary

in these areas. This is perhaps not surprising, since these are in fact new challenges for the Board, so there is necessarily less of an informational or analytical record to rely on when considering how best to address these challenges.

Appendix One: Decomposition of TFP Growth

There are rigorous ways to set X factors so that they are tailored to utility circumstances that differ materially from industry norms (either historically or at a given point in time). This can be done by developing information on the sources of TFP growth and adjusting the X factor to reflect the impact on TFP resulting from differences between a utility's particular circumstances and what is reflected in historical TFP trends. To provide a conceptual foundation for such adjustments, below we consider how the broad TFP aggregate discussed above can be decomposed into various sources of productivity change.

Our analysis begins by assuming a firm's cost level is the product of the minimum attainable cost level C^* and a term η that may be called the inefficiency factor.

$$C = C^* \cdot \eta . \qquad \text{[A1.1]}$$

The inefficiency factor takes a value greater than or equal to 1 and indicates how high the firm's actual costs are above the minimum attainable level.[53]

Minimum attainable cost is a function of the firm's output levels, the prices paid for production inputs, and business conditions beyond the control of management. Let the vectors of input prices facing a utility, output quantities and business conditions be given by $\boldsymbol{W}$ $(= W_1, W_2 \ldots W_J)$, $\boldsymbol{Y}$ $(= Y_1, Y_2 \ldots Y_I)$, and $\boldsymbol{Z}$ $(= Z_1, Z_2 \ldots Z_N)$, respectively. We also include a trend variable (T) that allows the cost function to shift over time due to technological change. The cost function can then be represented mathematically as

$$C^* = g(\boldsymbol{W}, \boldsymbol{Y}, \boldsymbol{Z}, T). \qquad \text{[A1.2]}$$

Taking logarithms and totally differentiating Equation [A1.2] with respect to time yields

$$\dot{C} = \left(\sum_i \varepsilon_{Y_i} \cdot \dot{Y} + \sum_j \varepsilon_{W_j} \cdot \dot{W} + \sum_n \varepsilon_{Z_n} \cdot \dot{Z} \right) + \dot{g}. \qquad \text{[A1.3]}$$

[53] A firm that has attained the minimum possible cost has no inefficiency and an inefficiency factor equal to 1. The natural logarithm of 1 is zero, so if a firm is operating at minimum cost, the inefficiency factor drops out of the analysis that follows.

Equations [A1.1] and [A1.3] imply that the growth rate of *actual* (not minimum) cost is given by

$$\dot{C} = \left(\sum_i \varepsilon_{Y_i} \cdot \dot{Y} + \sum_j \varepsilon_{W_j} \cdot \dot{W} + \sum_n \varepsilon_{Z_n} \cdot \dot{Z} \right) + \dot{g} + \dot{\eta}. \qquad [A1.4]$$

The term ε_{Y_i} in equation [A1.4] is the elasticity of cost with respect to output i. It measures the percentage change in cost due to a small percentage change in the output. The other ε terms have analogous definitions. The growth rate of each output quantity i is denoted by $\dot{Y}$. The growth rates of input prices and the other business condition variables are denoted analogously.

Shephard's lemma holds that the derivative of minimum cost with respect to the price of an input is the optimal input quantity. The elasticity of minimum cost with respect to the price of each input *j* can then be shown to equal the optimal share of that input in minimum cost (SC_j^*). Equation [A1.4] may therefore be rewritten as

$$\begin{aligned} \dot{C} &= \sum_i \varepsilon_{Y_i} \cdot \dot{Y} + \sum_j SC_j^* \cdot \dot{W} + \sum_n \varepsilon_{Z_n} \cdot \dot{Z} + \dot{g} + \dot{\eta}. \\ &= \sum_i \varepsilon_{Y_i} \cdot \dot{Y} + \dot{W}^* + \sum_n \varepsilon_{Z_n} \cdot \dot{Z} + \dot{g} + \dot{\eta}. \end{aligned} \qquad [A1.5]$$

The W^* term above is the growth rate of an input price index, computed as a weighted average of the growth rates in the price subindexes for each input category. The *optimal* (cost-minimizing) cost shares serve as weights. We will call W^* the optimal input price index.

Recall from the indexing logic presented earlier in this document that

$$T\dot{F}P = \dot{Y} - \dot{X} \qquad [A1.6]$$

And

$$\dot{X} = \dot{C} - \dot{W} \qquad [A1.7]$$

The input price index above is weighted using actual rather than optimal cost shares. Substituting equations [A1.6] and [A1.7] into [A1.5], it follows that

$$
\begin{aligned}
T\dot{F}P &= \dot{Y} - \left(\dot{C} - \dot{W}\right) \\
&= \dot{Y} - \left[\left(\sum_{i} \varepsilon_{Y_i} \cdot \dot{Y}_i + \sum_{n} \varepsilon_{Z_n} \cdot \dot{Z}_n + W^* + \dot{g} + \dot{\eta}\right) - \dot{W}\right] \\
&= \dot{Y} - \left[\left\{\left[\left(1 - \frac{1}{\sum \varepsilon_{Y_i}}\right) \cdot \sum \varepsilon_{Y_i} \cdot \dot{Y}_i + \sum_{i} \frac{\varepsilon_{Y_i}}{\sum_{i} \varepsilon_{Y_i}} \cdot \dot{Y}_i\right] + \sum_{n} \varepsilon_{\dot{Z}_n} \cdot \dot{Z}_n + W^* + \dot{g} + \dot{\eta}\right\} - \dot{W}\right] \\
&= \dot{Y} - \left\{\left[\left(\frac{1}{\sum \varepsilon_{Y_i}} - 1\right) \cdot \sum \varepsilon_{Y_i} \cdot \dot{Y}_i + \dot{Y}^{\varepsilon} + \sum_{n} \varepsilon_{\dot{Z}_n} \cdot \dot{Z}_n + W^* + \dot{g} + \dot{\eta}\right] - \dot{W}\right\} \\
&= \cdot \left(1 - \sum \varepsilon_{Y_i}\right) \cdot \dot{Y}_i + \left(\dot{Y} - \dot{Y}^{\varepsilon}\right) - \left(W^* - \dot{W}\right) - \sum_{n} \varepsilon_{\dot{Z}_n} \cdot \dot{Z}_n - \dot{g} - \dot{\eta}
\end{aligned}
$$

The expression above shows that growth rate in TFP has been decomposed into six terms. The first is the **scale economy effect**. Economies of scale are realized if, when all other variables are held constant, changes in output quantities lead to reductions in the unit cost of production. This will be the case if the sum of the cost elasticities with respect to the output variables is less than one.

The second term is the **nonmarginal cost pricing effect**. This is equal to the difference between the growth rates of two output quantity indexes. One is the index used to compute TFP growth. The other output quantity index, denoted by $\dot{Y}^{\varepsilon}$, is constructed using cost elasticity weights. The Tornqvist index that PEG and others typically use to measure TFP should theoretically be constructed by weighting outputs by their shares of revenues. It can be shown that using cost elasticities to weight outputs is appropriate if the firm's output prices are proportional to its marginal costs, but revenue-based weights will differ from cost elasticity shares if prices are not proportional to marginal costs. Accordingly, this term is interpreted as the effect on TFP growth resulting from departures from marginal cost pricing.[54]

The third term is the **cost share effect**. This measures the impact on TFP growth of differences in the growth of input price indexes based on optimal and actual cost shares. This term will have a non-zero value if the firm utilizes inputs in non-optimal proportions.

---

[54] See Denny, Fuss and Waverman, p. 197.

The fourth term is the **Z variable effect.** It reflects the impact on TFP growth of changes in the values of the Z variables that are beyond management control.

The fifth term is **technological change**. It measures the effect on productivity growth of a proportional shift in the cost function. A downward shift in the cost function due to technological change will increase TFP growth.

The sixth term is the **inefficiency effect**. This measures the effect on productivity growth of a change in the firm's inefficiency factor. A decrease in a firm's inefficiency will reduce cost and accelerate TFP growth. Firms decrease their inefficiency as they approach the cost frontier, which represents the lowest cost attainable for given values of output quantities, input prices, and other business conditions.

Appendix Two: Physical and Monetary Capital Measures

Although capital is almost universally measured using monetary rather than physical metrics in utility productivity studies, some consultants continue to advocate the use of physical capital metrics. It should also be noted that the issue of appropriate capital measures was the subject of considerable debate in 2007-2008 when the third generation incentive regulation plan was established for electricity distributors in Ontario. PEG advised the Staff of the Ontario Energy Board (OEB) in this proceeding, and we estimated an industry TFP trend using monetary capital values. London Economics (represented by Julia Frayer) developed an alternative TFP measure which used physical capital measures in part. In its September 2008 final decision, the OEB accepted PEG's approach and wrote that "(o)f greatest concern with Ms. Frayer's approach is the (physical) measurement of capital, which is inconsistent with the prior Ontario TFP studies and does not appear to have been adopted in any jurisdiction other than New Zealand."[55] While this was a fairly decisive rejection of the use of physical capital metrics on the part of the Board, it is worth reviewing the evidence on both sides of this debate in some detail, since the issue of how to measure capital is naturally important in this proceeding.

<u>The Relationship Between Asset Values and Indexing Logic</u>

One important factor supporting the use of monetary, rather than physical, capital values is the indexing logic which demonstrates the role that industry TFP trends can play in adjusting utility rates. This logic shows that only monetary capital values are internally consistent with the TFP trend measures that should be used in rate adjustment mechanisms. Recall that the indexing logic examines long-run changes in revenues and costs for an industry. In the long run, the trend in revenue (R) for an industry equals the trend in its cost (C).

$$Trend\ R = Trend\ C \qquad \textbf{(A2.1)}$$

[55] Ontario Energy Board, *Supplemental Report of the Board on 3rd Generation Incentive Regulation for Ontario's Electricity Distributors*, September 17, 2008, p. 12.

The trend in the revenue of any industry will be equal to the sum of trends in revenue-weighted output price indexes (P) and revenue-weighted output quantity indexes (Y).

$$Trend\ R = Trend\ P + Trend\ Y \tag{A2.2}$$

The growth rate in the cost incurred by an industry is the sum of the trends in a cost share-weighted input price index (W) and a cost-share weighted input quantity index (X).

$$Trend\ C = Trend\ W + Trend\ X \tag{A2.3}$$

Substituting (A2.2) and (A2.3) into equation (A2.1) and rearranging, we find

$$\begin{aligned} Trend\ P &= (Trend\ W + Trend\ X) - Trend\ Y \\ &= Trend\ W - (Trend\ Y - Trend\ X) \\ &= Trend\ W - Trend\ TFP \end{aligned} \tag{A2.4}$$

It can be seen that the trend in (revenue-weighted) prices depends on the difference between the trends in two indexes. The first is a *cost-share weighted* input price index. The second is a total factor productivity (TFP) index. The trend in output quantities used in the TFP index is calculated using revenue-share weights; the trend in input quantities used in the TFP index is calculated using cost-share weights.

In terms of the choices for capital inputs, the critical relationship in this logic is equation (A.3). This equation shows that there is a direct link between the input quantity measure used in TFP calculations and the costs of the industry. In other words, the trend change in the industry's input quantity (which is used, in turn, to compute industry TFP trends) should be associated with trend changes in industry cost. This relationship naturally applies to capital inputs, which account for the largest share of energy network inputs.

Clearly, the total cost of the industry is measured in monetary terms, and internal consistency requires this value to be decomposed into two component indices (for input prices and input quantities) that are measured on the same, monetary basis. This is almost invariably the case for opex inputs, which are measured using the monetary values for operating expenditures. These monetary values are "deflated" using an opex input price index, which functionally divides the monetary value of opex changes into a price change component (reflected in the change in the overall input price index, *W*) and a quantity change component (reflected in the change in the overall input quantity index, *X*). Capital input quantities will be logically consistent with the total cost and opex input quantity measures only if these indices are also calculated using monetary capital values.[56]

The link between monetary capital values and TFP trends is also consistent with how utility prices are set in practice. When prices under a multi-year incentive regulation plan are updated using measures of industry input price and TFP trends, prices at the outset of the plan are typically set to recover the company's cost of service in a base year. These initial year costs include the costs associated with capital assets. When a utility sets its rates to recover the depreciation and carrying costs of these capital goods, it does so with reference to the aggregated monetary values of these disparate assets, net of their depreciation. It follows that if monetary costs – including the monetary costs of physical capital assets - are used to set rates at the outset of a plan but a "physical method" for measuring capital is used to set the X factor, the X factor to *adjust* distribution rates will not be consistent with how those rates were originally set. This internal inconsistency between setting initial rates and adjusting rates over time can only reduce the transparency of the rate adjustment mechanism and perhaps exacerbate rate volatility when prices are updated, thereby undermining the predictability and effectiveness of the incentive regulation regime.

<u>The Relationship Between Physical Capital Values and the Assumed Pattern of Depreciation</u>

It should also be noted that the use of physical capital measures in TFP studies embody certain assumptions about depreciation. A necessary, but not sufficient, condition for using physical capital to measure the capital stock is for capital to obey what is known as "one-hoss shay" depreciation. The defining characteristic of one-hoss shay depreciation is that the asset undergoes *no* physical decay from the time it is installed until the day it is replaced. The classic example of a one-hoss shay "asset" is a light bulb.

The link between one hoss shay depreciation and physical capital can perhaps be clarified by recalling (as discussed in Chapter Two) that TFP growth is designed to measure the flow of services provided by aggregate inputs. The services provided by a given capital

---

[56] Indexes that obey this property are sometimes said to satisfy the "product test"; for example, see Waters, W.G. and J. Street (1998), "Monitoring the Performance of Government Trading Enterprises," *The Australian Economic Review*, Vol. 31, no. 4, p. 368.

good depend on how efficiently that asset is operating compared with its potential. Economists sometimes term this relationship between actual and potential services as the "efficiency units" associated with a given capital good. Whenever there is any physical asset decay, then the efficiency units of older capital must be less than the efficiency units of the newer capital. If this is the case, then old and new capital goods cannot simply be added together and used to measure capital input because there is effectively less input quantity being provided by the older capital goods. Different physical values for capital goods (such as km of distribution line installed in different years) can therefore be added together and used as an overall capital measure only when there is **no** physical decay in assets *i.e.* when there is one-hoss shay depreciation. When this is not the case, then the capital inputs installed in different years must also be adjusted to take account of capital decay that has taken place since the assets were put in place.

The following factual statements are worth emphasizing, when evaluating the claim that physical capital metrics should be used to measure capital in regulatory applications:

- A defining characteristic of one-hoss shay depreciation is that the asset undergoes *no* physical decay from the time it is installed until the time it is replaced.
- The productive services provided by a given capital good depend on how efficiently that asset is operating compared with its potential.
- Economists sometimes characterize the relationship between actual and potential services in terms of the "efficiency units" associated with a given capital good.
- Whenever there is *any* physical asset decay, then the efficiency units of older capital must be less than the efficiency units of the newer capital.
- When this is the case, old and new capital goods cannot be added together to measure capital input because some adjustment of physical capital measures is necessary in this instance to reflect the loss in "efficiency units" as capital goods age.

It follows that physical capital counts can be used to measure capital quantity *only* when the capital stock satisfies one hoss shay depreciation; if there is **any** physical decay in capital over time, you cannot simply add physical counts of assets installed in different years together and obtain an accurate measure of capital input quantity.

PEG does not believe that a one-hoss shay depreciation pattern (*i.e.* zero physical decay in every year an asset is in place) is consistent with day-to-day experience in energy network industries. For example, scores of utilities have implemented "reliability centered maintenance" programs which are designed to optimize system performance and extend asset life. Distribution maintenance involves many concrete decisions about inspection cycles, washing insulators, whether and when to treat or "wrap" wood poles, vegetation management, etc. Even though distribution assets tend to be long-lived, the fact that they involve extensive maintenance programs is a sure sign that there is some physical decay over time. It would be imprudent and unprofitable for utilities to devote resources to asset maintenance unless doing so increased the services effectively provided by these capital inputs. Such maintenance programs would also not be consistent with a one-hoss shay depreciation pattern, where the assets must be providing a constant stream of services *before* maintenance programs are undertaken.[57]

PEG's view is also consistent with how Staff has described depreciation in its Discussion Paper on the Regulatory Treatment of Infrastructure Investment for Ontario's Electricity Transmitters and Distributors. In that Discussion Paper, Staff writes:

> Stations, lines and meter assets, and their components, are subject to deterioration that will eventually impede their ability to function as originally designed. Asset deterioration depends on factors such as geographic environment/location, utilization, age, weather and maintenance practices. As assets deteriorate, equipment performance reliability usually suffers, resulting in increased environmental risks, an increase in potential safety hazards to the public and employees, and decreased system reliability. Ultimately, assets deteriorate to the point that they are no longer able to perform their function in a cost-effective manner, at which point replacement, rather than repair and maintenance, becomes necessary.[58]

---

[57] It has been argued that the presence of maintenance expenditures can be consistent with one-hoss shay depreciation, since agricultural land sometimes includes expenditures to maintain the productivity of given lands and yet land typically is assumed not to depreciate in TFP studies. However, there is an important distinction to be made between "no depreciation" and one-hoss shay depreciation. The difference is that, with very rare exceptions, land is not physically replaced at all, so it is appropriate to assume that there is no depreciation since the concept is inherently designed to measure the extent to which assets are "used up" over time as they are utilized in production. Other than land, all assets will inevitably be completely used up at some point and hence must be replaced (assuming ongoing operation of the enterprise and that the asset has not become technologically obsolete). This disparity between land and other assets implies that the zero depreciation for land assets is not equivalent to one-hoss shay depreciation.

[58] *Staff Discussion Paper on the Regulatory Treatment of Infrastructure Investment for Ontario's Electricity Transmitters and Distributors*, EB-2009-0152, June 5, 2009, p. 6.

Distributors in other jurisdictions have also provided written evidence, in regulatory filings, that their assets exhibit some physical decay. For example, there have recently been extensive problems with bushfires in the Australian State of Victoria. The Victorian Bushfires Royal Commission received and analyzed submissions on the extent to which electricity assets, and distributors' asset management practices, may have contributed to the risk and incidence of bushfires in the State. The Royal Commission Report is replete with evidence that electricity assets deteriorate over time and, therefore, do not display one hoss shay depreciation. For example:

> "SP AusNet provided to the (Royal) Commission the results of a study of its conductor fleet, which noted, among other things, 'The primary issue facing SP AusNet is the increasing age profile and deteriorating performance (2 % per annum) of steel and copper conductor through failure…'"[59]
> "Powercor submitted that the data show that its maintenance regime is working because 'deteriorated assets are…detected before they fail"[60]
>
> "…in severe weather conditions (in particular, high winds) deteriorated tie wires carry real potential to cause fire and are an example of a 'hidden defect.' As Professor Hastings stated '…[a broken tie is] a hidden defect because it is not an in service failure but it is this degraded state. I think a lot of the issue with distribution networks and their situation in relation to high fire danger days is related to keeping the number of these defects which have not yet progressed to failure under control or to a desirably low level"[61]

Other statements could be cited, but these findings are sufficient to show that electricity distribution assets are not characterized by one hoss shay depreciation. The distribution businesses themselves say their assets deteriorate over time, and deteriorated assets can be detected and replaced or repaired before they fail. Moreover, the Royal Commission findings show that it is frankly dangerous to assume, as some consultants have, that electricity distribution assets do not deteriorate since they appear to be providing power delivery services. "Hidden defects" exist before infrastructure assets progress to complete failure, and these hidden defects necessarily entail a loss of productive services (and increased risks) and physical deterioration over time.

---

[59] *2009 Victorian Bushfires Royal Commission*, *Volume II: Fire Preparation, Response and Recovery*, p. 151.

[60] *2009 Victorian Bushfires Royal Commission, op cit*, p. 152.

[61] *2009 Victorian Bushfires Royal Commission, op cit*, p. 152-153.

A corollary of the "no physical decay" condition is that one hoss shay assets also provide unmistakable replacement signals. One-hoss shay capital goods work perfectly until the day they break down, at which point they never work again and must be replaced. This also does not reflect the reality of most energy network assets, as is evident in the Royal Commission Report. Managers have a degree of discretion about when to replace assets and, to a lesser extent, about replacing current labor-based operations with capital equipment (*e.g.* in service restoration). Replacement decisions are, in fact, intertwined with operational and maintenance decisions. The complexity and inter-relatedness of these judgments is not consistent with the transparent simplicity of deciding when to replace a light bulb.

The economics literature also generally supports the notion that energy network assets are not characterized by one-hoss shay depreciation. Indeed, this literature has found exceedingly few assets with one hoss shay depreciation profiles in any industry. One statement of this view comes from an OECD Manual titled *Measuring Capital: Measurement of Capital Stocks, Consumption of Fixed Capital, and Capital Services:*

> "There are probably rather few assets that maintain constant efficiency throughout their working lives. Light bulbs are sometimes cited as potential one-hoss shays, but light-bulbs are too short-lived to be classified as capital goods. More serious contenders might be bridges or dams. With a constant level of maintenance these structures may continue to provide constant rentals for very long periods. In general, however, few examples of the one-hoss shay have been identified in the real world."[62]

The literature also finds that when observers ignore the role of maintenance expenditures, they often incorrectly conclude that assets exhibit one hoss shay depreciation. This has been noted in the *Dictionary of Usage for Capital Measurement Issues*, released in conjunction with the Second Meeting of the Canberra Group on Capital Stock Statistics:

> "The concept of decay is a crucial one in capital measurement. Some additional remarks about input and output decay may clarify the concepts. The division between output decay and input decay is economically, not technologically, determined, because owners can often offset output decay by increased maintenance. However, increased maintenance as a capital good ages implies input decay. Accordingly, when increased maintenance does compensate for output decay, this does not create a one hoss shay asset, because a one hoss shay asset is by definition one with zero decay. There seems to be some confusion on this point in the literature: A good deal of the

[62] OECD Manual. (2001), *Measuring Capital – Measurement of Capital Stocks, Consumption of Fixed Capital and Capital Services.*

anecdotal evidence that has been cited in favor of the plausibility of the one hoss shay model has ignored input decay."[63]

Arguments in favor of one hoss shay depreciation based on "casual experience" or "intuitive appeal" also run contrary to rigorous empirical depreciation studies. For example, when discussing alternative depreciation patterns, Charles Hulten (a depreciation expert) writes that observers often believe "...the one hoss shay pattern commands the most intuitive appeal. Casual experience with commonly used assets suggests that most assets have pretty much the same level of efficiency regardless of their age – a one year old chair does the same job as a 20 year old chair, and so on."[64] However, this author's own academic work shows that this "casual experience" conflicts with more scientific investigations of depreciation. Hulten and Wykoff examined the prices that were actually paid in secondary markets for used capital goods.[65] They found that these prices were most consistent with geometric and not one-hoss shay depreciation patterns. This work has been very influential and is used directly by a number of researchers (including the US Bureau of Economic analysis) to value capital stocks. Surveying the intuitive and empirical arguments, Hulten writes:

> "Taken together, these intuitive arguments (in favor of one hoss shay) above suggest that this is a case in which the econometric evidence leads to the wrong result. However, it may also be true that the intuition, not the econometrics, is faulty. Intuition tends to be based on personal experience of individual cases."[66]

Furthermore, Hulten notes that proponents of one-hoss shay depreciation ignore what is known as the "portfolio effect," *i.e.* the depreciation profile associated with a group of disparate assets – such as those owned by energy networks– will often differ from the depreciation of any individual asset. He writes:

> "Moreover, what may be true on a case-by-case basis may not be true of an entire population of assets. If so, this has important implications for evaluating

[63] Triplett, Jack. (1998). *A Dictionary of Usage for Capital Measurement Issues*, presented at the Second Meeting of the Canberra Group on Capital Stock Statistics (OECD).

[64] C. Hulten (1990), "The Measurement of Capital" in *Fifty Years of Economic Measurement* eds. E.R. Berndt and J. Triplett, Studies in Income and Wealth, vol. 54, the National Bureau of Economic Research, Chicago: The University of Chicago Press, p. 124.

[65] C. Hulten and F. Wykoff (1981), "The Measurement of Economic Depreciation," in *Depreciation, Inflation and the Taxation of Income from Capital* ed. C. Hulten, Washington DC: The Urban Institute Press, 81-125.

[66] Hulten, Charles R & Wykoff, Frank C. (Jan 1996). Issues in the measurement of economic depreciation: Introductory remarks. *Economic Inquiry 34*(1), pp. 10-24.

econometric results, which typically reflect the average experience of whole populations and not individual units. For instance, it may well be true that every single asset in a group of 1000 assets depreciates as a one-hoss shay, but that the group as a whole experiences near-geometric depreciation. This fallacy of composition arises from the fact that different assets in the group are retired at different dates: some may last only a year or two, others ten to fifteen years. When the experience of the short-lived assets is averaged against the experience of the long-lived assets, and the average cohort experience is graphed, it will look nearly geometric if the 1000 assets have a retirement distribution of the sort used by the Bureau of Economic Analysis (i.e., one of the Winfrey distributions). Thus, the average asset (in the sense of an asset that embodies the experience of 1/1000 each of 1000 assets in the group) is not one hoss shay, but something that is much closer to the geometric pattern. This can easily be verified by performing this experiment using the parameters of the Bureau of Economic Analysis's capital stock program."[67]

This is a subtle but important point. It implies that whenever utilities within an industry have different retirement patterns, the "portfolio effect" described above will apply. If the portfolio effect does in fact apply, then even if *every* asset exhibits one hoss shay depreciation, the industry-wide depreciation pattern would **not** be consistent with one hoss shay depreciation. The existence of different asset retirement patterns across a utility industry is therefore *sufficient* to rule out the assumption of one-hoss shay depreciation as an industry-wide measure of depreciation. Furthermore, if one-hoss shay depreciation does not apply, then physical capital metrics cannot be used.

It must therefore be remembered that two necessary conditions that *must* be satisfied for physical capital metrics to be used to measure the industry's capital stock:

1. Every capital asset good measured by physical counts must exhibit one-hoss shay depreciation.
2. The portfolio effect – wherein firms in an industry have different asset retirement patterns – must not apply.

If latter is satisfied, this is sufficient to rule out one-hoss shay, which in turn rules out the use of physical capital metrics.

[67] Hulten, Charles R & Wykoff, Frank C. (January 1996). Issues in the measurement of economic depreciation: Introductory remarks. *Economic Inquiry 34*(1), pp. 10-24.

Some consultants have put forward empirical evidence which they claim supports the use of one-hoss shay depreciation. However, a careful examination shows that in literally *none* of the referenced studies is one-hoss shay depreciation actually used. For example, on pp. 54-57 of its December 2009 *Total Factor Productivity Index Specification Issues* Report to the Australian Energy Market Commission (AEMC), Economic Insights (EI) presents evidence on the actual depreciation patterns used in four applications which it claims supports the use of one-hoss shay depreciation: the US Bureau of Economic Analysis (BEA); the US Bureau of Labor Statistics (BLS); Statistics New Zealand (SNZ); and the Australian Bureau of Statistics (ABS). Of these four specific studies EI says BEA uses geometric depreciation, which is not one hoss shay depreciation. Regarding the SNZ and ABS depreciation treatments, EI writes:

> Importantly, both Statistics New Zealand and the Australian Bureau of Statistics have adopted the hyperbolic age–efficiency profile in their productivity studies. A key parameter in the hyperbolic age–efficiency profile can be set to influence the degree of curvature. A value of one for this parameter leads to a flat or one hoss shay profile while a value of zero would give equal deterioration each year (ie approximate straight line deterioration). Both SNZ and the ABS set this parameter at 0.5 for equipment and 0.75 for structures. That is, they are assuming closer to one hoss shay deterioration for structures. This is the complete opposite of the geometric deterioration profile advocated by PEG.[68]

EI also notes that in a hyperbolic age-efficiency profile, a value of a curvature parameter equal to one leads to one hoss shay depreciation. However, EI states both SNZ and the ABS set this curvature parameter at 0.5 for equipment and 0.75 for structures, which is flatly incompatible with the value needed for one hoss shay depreciation. EI never define what it means to be "closer to one hoss shay depreciation" for depreciation, but this would be irrelevant in any event, because one hoss shay depreciation will not apply whenever there is any physical deterioration of capital, and if this extreme assumption is not satisfied then physical capital counts cannot be used to measure capital. Finally, EI says that BLS uses a hyperbolic depreciation treatment but do not report the curvature parameter; however, it can be easily confirmed that the BLS also does not use one hoss shay depreciation. The factual record has therefore presented no evidence of any official, national statistical agency using one hoss shay depreciation.

[68] Economic Insights, *Total Factor Productivity Index Specification Issues*, December 2009, p. 56.

It should also be noted that PEG is aware of 42 separate instances where TFP information was used to set rate adjustments, and in every one of these instances monetary rather than physical metrics were used to measure capital.[69] The only known instance where physical capital metrics have been used in regulatory applications of TFP methodologies is in the two plans EI personnel have been involved in New Zealand. In the most recent New Zealand price control plan adopted in November 2009, the Commerce Commission said that it "based its decision on the long-run average productivity improvement rates as derived by the TFP analysis of both Economics Insights and PEG." [70] In this most recent proceeding, the New Zealand Commerce Commission therefore relied on TFP evidence from PEG (which used monetary capital measures) and its advisor EI when setting the X factor.[71]

In sum, it is not sufficient to note that assets are long-lived or that a depreciation treatment is "closer" to one hoss shay than to an alternative method. One hoss shay is an extreme depreciation assumption, since it is literally impossible to have less than *no* physical decay in an asset from the day it is installed until the day it is replaced. Nevertheless, this is what one hoss shay requires, and this condition *must* be satisfied if physical counts of assets from different years are to be added together to measure the capital stock. *Any* physical decay in an asset over time is sufficient to rule out the use of one hoss shay deprecation and physical asset metrics. Moreover, if there is any evidence of a portfolio effect (*i.e.* different patterns of asset retirement across the industry), then even if every asset displays one-hoss shay depreciation, the industry-wide depreciation rate will not and physical capital measures must

[69] These plans are for Southern California Gas, Southern California Edison, San Diego Gas and Electric – gas, San Diego Gas and Electric – electric, Pacificorp California (twice), Boston Gas (twice), Berkshire Gas, Bay State Gas, Union Gas, electricity distributors in Ontario Canada (twice), US oil pipelines (twice), AT&T, local exchange carriers subject to FCC jurisdiction (twice), US West-North Dakota, NYNEX-MA, NYNEX-PA, and Class I US railroads (a total of 21 times – original plan, plus 20 annual updates). It could be argued that the latter example constitutes a single plan, although in principle that is not the case, since the regulator is able to propose changes in the TFP specification at any time. Even if the more restrictive interpretation of this plan is accepted, however, there are at least 22 separate indexing plans – and almost certainly more – that have adopted a TFP specification that does not use either physical capital measures or infrastructure-based system capacity outputs.

[70] Commerce Commission, *Decisions Paper: Initial Reset of the DPP*, November 2009, p. 47.

[71] It is notable that the Commerce Commission relied on PEG's TFP evidence even though their Draft Decision echoed some of the points that EI was making about PEG's TFP approach, especially that it assumed a competitive market exists and capital was entirely fungible, which made our results irrelevant for regulated industries. If the Commerce Commission accepted these claims, it would have almost certainly placed no weight on our TFP results, which EI argued were irrelevant for electricity distributors. The fact that the Commerce Commission did use PEG's TFP evidence for its final decision is strong evidence that it did not accept EI's theoretical claims regarding PEG's TFP specification.

not be used in the TFP specification. Whether a portfolio effect exists among Ontario energy networks is an inherently empirical issue that can be investigated and, if it exists, this fact would be sufficient to rule out one hoss shay depreciation and, therefore, physical capital measures.

Appendix Three: Evaluation of Benchmarking Options

This appendix will evaluate the merits of four alternative approaches for benchmarking utility performance**.** A number of criteria can be used to evaluate the merits of benchmarking methods. We believe five criteria are most important for judging the advantages and disadvantages of alternative benchmarking techniques.

1. *Consistency with economic theory* A given benchmarking technique is preferred if it is consistent with the economic theory of production.

2. *Restrictions on the relationship between the performance measure and business conditions* Benchmarking inevitably uses models that relate a performance measure (e.g. costs) to business condition variables. Models can vary in terms of how restrictive the assumed relationship is between the performance measures and these business conditions. All else equal, a benchmarking approach will be more generally applicable, and therefore preferred, if it imposes fewer assumptions on this relationship.

3. *Ability to capture business conditions* A given benchmarking technique must be appropriate to the power distribution industry. As discussed in our previous report, power distribution has a number of unique characteristics which can affect network cost but are beyond company control. Not all benchmarking techniques may be well suited to measuring and capturing these conditions. A given benchmarking method is clearly preferred it is better able to reflect the business conditions that networks face.

4. *Data requirements* Some benchmarking approaches may require more information in order to be implemented reliably. This can limit the usefulness or applicability of a technique, especially if a limited amount of data are available. All else equal, benchmarking techniques that require less data to generate reliable results are preferred.

5. *Ability to deal with uncertainty* A number of factors that can affect network costs will either have a random element (*e.g.* weather which affects O&M costs) or

measured imperfectly at best (*e.g.* the difficulty of terrain). If these factors influence cost but are not reflected in specific business conditions in the analysis, measured performance will be distorted. All else equal, a benchmarking technique is preferred if it provides evidence on the certainty associated with benchmarking assessments.

We evaluate, in turn, index-based benchmarks; econometric benchmarking; stochastic frontier analysis (SFA); and data envelope analysis.

A.3.1 Index Based Benchmarks

A.3.1.1 Partial Factor Productivity (PFP)

There are few advantages with using PFP measures in benchmarking. Perhaps the most important is that they are simple to compute. Measures like labor productivity (total output per unit of labor) are also relatively intuitive and easy to understand. PFP also does control for some differences in operating conditions that utilities face. For example, PFP comparisons across companies and across time do control for differences in input prices.

However, there are many well-known problems with using PFP and other partial measures. One is that this is a non-statistical approach. It therefore does not allow evaluations of the uncertainty associated with the calculated benchmarks.

Another important disadvantage is that partial measures do not control for differences in most business conditions, including the network's own choices for other inputs. For example, it is widely recognized that a utility's operation and maintenance (O&M) expenses will be affected by its capital choices. Asset replacement and maintenance are often substitute activities, so companies face trade-offs regarding capital and O&M inputs. Benchmark measures that focus on only a single factor (such as O&M spending) can therefore provide a misleading indicator of overall performance.

A.3.1.2 Total Factor Productivity (TFP)

TFP has several advantages as a means for generating benchmarks. One advantage is its consistency with economic theory. Well-established economic theory and empirical methods can be used in index construction.

A TFP index also controls for some business conditions. For example, TFP indexes control for differences in input prices across companies. They also control to some extent for differences in the scale of operations and local demand conditions that may, for example, be affected in output growth.

Another important advantage is that TFP has a direct link to the competitive market paradigm that can be used to establish effective rate regulation. As previously noted, price trends in competitive markets can be decomposed into the trend in industry input prices minus the trend in industry TFP. Industry TFP is therefore a natural basis for benchmarks in CPI-X plans, particularly if CPI inflation is a good proxy for growth in the industry's input prices.

Many North American regulators have recognized this competitive market paradigm and used TFP in indexing plans. As documented in previous PEG research, industry TFP trends have been used to calibrate the X factors in North American indexing plans for over a decade. TFP measures have been the basis for X factors in plans for North American railroads, telecom companies, and gas and electric utilities.

In spite of these benefits, TFP does have limitations. Perhaps the most important is that one cannot evaluate the uncertainty associated with TFP-based benchmarks. TFP indexes are not derived using statistical techniques. Accordingly, there is no information on the statistical precision of a TFP index. This is less problematic when TFP trends are used as the basis for an X factor, since the competitive market paradigm establishes a direct link between the long-run TFP trend in an industry and industry prices. However, it is more problematic if a TFP level index is used to evaluate a utility's performance at a given point in time.

In addition, TFP does not control for as many business conditions as other benchmarking techniques. For example, TFP indexes will control imperfectly, at best, for differences in customer mix and customer density between utilities. These factors can significantly impact utility cost. Again, this is less problematic when examining TFP trends, but it is more of an issue when comparing TFP levels across companies.

For these reasons, TFP is an important external benchmark, but it may not be sufficient when implementing an external approach to utility regulation. Information on industry TFP trends is quite valuable as a calibration point for the X factor in CPI-X regulation. It may also be appropriate to set X factors equal to industry TFP trends for utilities that are superior performers. But additional information may be needed to establish consumer dividends, which will be appropriate for many utilities that are not superior performers.

A.3.2 Econometric Cost Functions

A.3.2.1 Advantages of Econometric Cost Functions

With econometric cost models, performance is measured by comparing a company's actual cost with the cost predicted by the model. The following comparison makes use of the point prediction of cost.

$$Estimated\ Cost\ Performance = C_{Network,t} - \hat{C}_{Network,t}$$

Here $C_{Network,t}$ refers to the network's actual cost in period *t*, while $\hat{C}_{Network,t}$ is the estimated network cost in that period. Econometric cost functions reflect the cost that would be expected for that firm given an average efficiency standard.

An important advantage of econometric benchmarking is that results can assess the precision of such "point" predictions. Precision is greater as the variance of the prediction error declines.[72] The estimated variance of the prediction error can be used in two ways to assess the model's precision. One is to calculate a t-statistic for a model's prediction. This statistic will decline as the estimated variance increases.

A second approach is to construct a confidence interval around the point prediction. This interval represents the range of cost figures that is apt to encompass the true cost value at a certain confidence level. The point prediction lies at the center of this interval. The confidence interval may be viewed as the full range of cost predictions that is consistent with the historical data. It is wider as the estimated variance of the prediction error increases. If a utility's actual cost is not within a confidence interval, we may conclude that a network's actual cost differs significantly from the model's prediction.

---

[72] Generally speaking, the precision of econometric cost models will increase as the size of the sample increases; the number of business condition variables in the model declines; the business conditions of sample companies become more heterogeneous; the business conditions of the company in question become closer to those of the typical firm in the sample; and the model is more successful in predicting the costs of the sampled companies.

Another advantage of this approach is that it can be sensitive to a wide range of utility business conditions. Econometric benchmarking does not require identification of a suitable peer group. Indeed, variation in sampled business conditions is actually welcomed in econometric benchmarking since it helps to make estimates of model parameters more accurate. Econometric benchmarks will be based on the exact business conditions that were faced by a utility.

Econometric benchmarking is also linked to the economic theory of production. Since total cost is the performance indicator, it possible to use the economic theory of cost to select business condition variables. The resultant benchmarking model therefore has a direct link to economic theory and is free of the accusation of being a "black box".

Econometric benchmarking also has properties that can make it valuable in regulatory applications. For example, performance evaluations equal to the difference between actual and predicted cost can be generated for all firms in the sample. This makes it possible to observe the entire distribution of performance over the sample. This can be valuable to regulators that are attempting to select performance targets that are close to but not actually at the best observed performance. As discussed in Chapter 4, it is not appropriate for regulated prices to reflect "frontier" performance standards for all firms since this would not allow a utility's returns to be commensurate with its performance.

A.3.2.2 Disadvantages of Econometric Cost Functions

There are four main criticisms of econometric cost functions. The first is that they do not compute the minimum total cost function but only an average or expected cost function. It is therefore purportedly less consistent with the economic theory of production, which is based on cost minimizing behavior, than frontier econometric methods like SFA. We believe this criticism is baseless. There is nothing theoretically suspect about estimating the average cost function for an industry. We also believe that econometric cost functions are appealing in terms of the competitive market paradigm, since competitive markets prices depend on the industry's average performance.

Second, econometric cost functions necessarily assume a functional form, which imposes some restrictions on the underlying cost and production relationships. It is true that econometric approaches must assume functional forms, but economists have identified a number of "flexible" functional forms that minimize these restrictions. A flexible cost function will be a good approximation to any underlying production structure. We therefore believe that, in practice, this is not a serious limitation.

However, flexible functional forms do tend to increase the number of variables used in econometric analysis. Adding new business condition variables to a flexible function cost model can lead to a more than proportional increase in the number of parameters that must be estimated. It is therefore true that there is a tradeoff between the extent to which an econometric functional form imposes few restrictions on the underlying production process and the amount of data that are needed to estimate that function reliably.

Third, it is sometimes said that econometric cost functions assume that any deviation from the predicted cost function is a measure of efficiency and/or inefficiency. This is deemed to be an invalid inference since the residual can contain both random error and an efficiency/inefficiency factor. Neither of these factors can be observed, but benchmarking methods should measure only the latter.

However, there are ways of discriminating between random error and inefficiency in non-frontier econometric cost models. These methods can be used to isolate the efficiency/inefficiency factor. One established method comes from a classic econometric paper by Mundlak. Under this approach, the error term is assumed to have a firm specific effect that is constant over the sample period and a random variable with a mean of zero whose value may vary from year to year.[73] The firm specific effect captures any persistent deviation in the cost of a company from that predicted by the business condition variables over the sample period. It reflects the net effect of a range of conditions, including differences in the efficiency of companies and in business conditions that were excluded from the model.

Following Mundlak, it can be assumed that the firm specific effect has a systematic and a non-systematic component. The systematic component depends on the mean values of the included business condition variables included in the econometric model. For example, the impact on cost of an excluded output quantity variable might very well be larger as the

---

[73] Yair Mundlak, "On the Pooling of Time Series and Cross Section Data", *Econometrica*, Vol. 46, pp. 69-85, 1978.

values of the included output quantity variables increase. Only the non-systematic effect will then reflect the cost inefficiency of the utility. This formulation therefore enables firm inefficiency to be isolated from random error.

A fourth potential disadvantage is that econometric cost functions have greater data requirements than some other methods. One aspect of this was discussed above with regard to flexible form cost functions. Another potentially problematic data requirement is that cost functions require input prices. Data on capital cost and capital input prices, in particular, are not always easy to obtain. While this is true, we do not believe that alternative input measures such as physical units of capital will ultimately be appropriate in utility benchmarking studies. We discuss this issue extensively in Appendix One.

A.3.3 Stochastic Frontier Analysis

A.3.3.1 Advantages of SFA

SFA is also an econometric approach, so it shares many of the advantages and disadvantages as econometric cost functions. In particular, SFA allows for analysis of the statistical precision of benchmarking assessments. SFA can also use tailored business conditions and benchmark networks subject to the actual conditions that they face.

In addition, SFA computes the minimum total cost of production and directly calculates the firm's inefficiency factor. It does so by specifying two components of the error term. The first is a purely random factor that can be either positive or negative for a firm at a given point in time. This implies that random factors can have either a positive or negative impact on any given cost observation. The second component of the error term is a one-sided inefficiency factor. In a cost function, this term can only have non-negative values. This implies that inefficiency can only raise cost above the minimum total cost. A firm that has cost equal to the minimum total cost will have an inefficiency factor of zero.

A.3.3.2 Disadvantages of SFA

Two of the disadvantages cited for SFA also apply to econometric cost functions. They are the specification of specific forms for the cost function and the need to obtain data measures like capital input prices that can be difficult to collect. Our comments above on these issues also apply here.

In addition, another disadvantage with SFA is that is typically involves specifying a statistical distribution for the inefficiency factor.[74] Since this factor cannot be observed, some type of distribution is assumed. Different assumptions on the distribution of inefficiency can affect the value for inefficiency that is computed. While there is no academic consensus on which distribution is most appropriate, a relatively small number of distributions have been used in most research.[75] SFA can be applied using each of these options, and the results can be examined to determine the sensitivity of estimated inefficiency to this assumption.

SFA can also have some problems from a regulatory standpoint. The most important may be in the interpretation of SFA results. SFA calculates "frontier" performance levels, and there is likely to be a temptation to use the estimated frontier as a performance standard for all firms in the industry. As we have emphasized, this is not an appropriate regulatory standard.

For technical reasons, it is also difficult to estimate multi-equation systems using SFA.[76] This is not typically true of econometric cost models, where cost share equations derived using economic theory are estimated simultaneously with the total cost model. Since econometric cost models can generate predictions for specific cost categories as well as total cost, they may thereby facilitate the transition to more comprehensive benchmarking more easily than SFA.

A.3.4 Data Envelope Analysis

A.3.4.1 Advantages of DEA

Several advantages are frequently cited for DEA. One is that since it is a non-parametric approach, there is no need to specify a functional form. This imposes fewer restrictions on the underlying production relationship

---

[74] With panel data (*i.e.* time series data for a cross section of firms), it may be possible to estimate firm-specific inefficiency, as in SFA, without specifying a distributional assumption on the inefficiency term. This is discussed in P. Schmidt and R. Sickles (1984), "Production Frontiers and Panel Data", *Journal of Business and Economic Statistics*, 367-374.

[75] The two most common assumptions on the statistical distribution of inefficiency are likely to be the half-normal distribution and the gamma distribution.

[76] This is sometimes referred to as "the Greene problem."

It is also sometimes argued that DEA has fewer data requirements. In particular, it is possible to use physical rather than financial input and output measures in DEA. Physical input data are sometimes easier to obtain, particularly for capital inputs.

The possibility of reduced data requirements was a primary reason why the energy regulator in the Netherlands recently chose to benchmark the country's networks using DEA rather than SFA or other econometric methods. There is currently only a single year's worth of data for the 20 networks in the nation. The regulator has written that DEA is therefore preferable because the small sample size "…makes meaningful regression analysis virtually impossible. After all, regression techniques that estimate relationships between costs and cost drivers (such as customer numbers or customer density) can produce misleading results in small sample sizes."[77]

A.3.4.2 Disadvantages with DEA

When applied to electricity networks, we believe that many of DEA's potential advantages are illusory. There are also numerous problems with this technique. Some of these problems have been noted generally, but few have examined the particular problems that arise when applying DEA to electricity distribution and transmission. We divide the disadvantages with DEA into four categories: data requirements and related problems; the ability to deal with uncertainty; assumptions regarding the production process; and problems with controlling for networks' business conditions. Some of these issues are interrelated so the problems will overlap somewhat between categories.

Data Measures and Requirements

Capital accounts for the dominant share of network costs, so its treatment in benchmarking models is critical. DEA typically uses physical rather than financial capital measures as inputs in network benchmarking studies. Examples include MVA of transformer capacity and km of distribution line. In addition to all the reasons discussed in Appendix One, we believe that this approach is problematic in several respects.

---

77 DTE, *Guidelines for Price Cap Regulation of the Dutch Electricity Sector in the Period from 2000 to 2003*.

One reason is that power distributors' capital is in fact extremely varied. For example, SCADA and related computer systems are increasingly important for monitoring and controlling distribution systems, but these cannot be measured in simple quantitative units.[78] Similarly, networks have sophisticated telephone call centers, customer information service systems for maintaining metering and billing databases, networks that link customer service and field service representatives, and many other types of equipment. These items account for sizeable shares of networks' capital stock, but they can only be measured in financial terms. It is therefore not possible to measure the scope of network capital accurately with a few simple physical measures.

In addition, physical capital units will not capture assets' age profile. This can be an important consideration, since older assets will typically entail greater maintenance expenses. If DEA inputs include higher O&M costs but do not reflect the age profile of the capital stock, results may be biased against firms with an older asset profile. In contrast, there are rigorous methods for constructing financial capital measures that appropriately reflect the age and effective services provided by a firm's capital assets. This should lead to more reliable benchmarking assessments.

Power distribution systems are also designed differently in different countries, and this can affect the relative amounts of physical assets. For example, the US delivers electricity to most end-users at 110V, while in Australia power is delivered to most end-users at 220/250V. This difference has implications for the design of distribution systems for most urban and suburban residential customers. In the US, there is usually one transformer per residential customer (usually 10 or 16 kVA) with little low voltage line. In Australia, there is usually one larger transformer for each 100 customers or so (usually a three phase 315 kVA) with an extensive low voltage network.

These design differences can affect the results from DEA benchmarking models, particularly when applied to international samples. DEA usually deals with physical quantities of inputs, so differences in relative amounts of inputs can affect DEA results. In general, US networks will have more MVA of transformer capacity and fewer km of line,

[78] SCADA stands for system control and data acquisition and refers to computer-based systems that are used for a variety of operations, including monitoring and controlling network components.

while Australian networks will have more km of line and fewer MVA of transformer capacity.[79] Different input proportions can distort which firms are selected as peers, since this choice depends on relative input proportions among sampled companies. Comparing a distributor to an inappropriate peer leads directly to inappropriate benchmarking results.[80] In contrast, distortions do not arise with econometric cost models that focus on total cost and financial capital measures. Differences in network design do not distort these measures since, given each system's history, the differences in design are most cost effective. Therefore total cost comparisons (as in econometric models) remain valid, while DEA results are distorted by differences in system design and the proportions of physical inputs.

Difficulties also arise in accounting for the transportation nature of energy delivery networks. Measures of energy transportation, such as km of distribution line, are sometimes treated as inputs in DEA studies. However, this is flawed in at least two ways. The first is that purely physical measures like km of line do not reflect the efficiency with which firms construct delivery networks. There is evidence that these differences can be substantial, particularly because of differences in work rules and other factors that affect the productivity of construction labor in different countries.[81] These factors will not be manifested in the physical km of line measure, but they will be reflected in the financial cost (and efficiency) of constructing distribution lines.

In addition, it is difficult to capture the transportation nature of power distribution services if km of line is treated as an input. Direct delivery of power to customers is an essential network *output* in cost performance studies. This output can be proxied by the total km of distribution line, since this is related to the physical location of customers in the

---

[79] For example, if there is one 16kVA transformer for each US customer, there will be 1600 kVA for each 100 customers, compared with 315 kVA for each 100 Australian customers. But consistent with using a higher voltage transformer, Australian networks have a greater reticulated low-voltage network compared with US firms.

[80] Put another way, differences in system design between US and Australian networks would lead to expected differences in the proportions of MVA capacity and km of line for two firms in the same countries that served the same customer mix. If an Australian and US firm were selected as peers because they used similar proportions of MVA capacity and km of line, this would imply that these firms actually served a different mix of customers.

[81] The *Richardson International Construction Cost Location Factors* provide evidence on the cost of constructing utility plant in different countries, as well as evidence on the factors that account for differences in construction costs. This data source estimates significant differences in labor productivity between US and Australian firms.

network's territory. But it is not possible to include km of line as an output in DEA models if it is already used as an input. However, if km of line is used as a DEA output rather than an input, then the model will not reflect the costs associated with the "lines and poles" needed to deliver power to customers.

In short, it is not possible to capture networks' essential service of delivering power directly to customers and the costs associated with this service by using a single variable such as km of line. The only sensible model must also include the financial costs associated with constructing these lines. It therefore does not appear to be practical to benchmark networks using only physical capital measures.

Data Issues and Uncertainty

DEA is not a statistical method, so it much less conducive to dealing with uncertainties regarding benchmarking measures. It is often not possible to test the statistical precision of benchmarks that are estimated through DEA. DEA also does not naturally lend itself to the construction of confidence intervals around benchmarks.

In fact, since DEA is not a statistical approach, the data themselves establish the cost and/or production frontier. This means that the constructed frontier, and therefore any firm's estimated inefficiency, is extremely sensitive to the quality of the sample data themselves. While it is important to use high quality data in any benchmarking study, the quality of the data becomes a paramount issue under DEA.

Data problems can directly affect efficiency measures. For example, estimated frontiers can result from sample "outliers." Firms may be outliers because of data errors, business condition variables that are omitted from the analysis, and a host of other reasons. DEA measures are also sensitive to the size of the sample. All else equal, larger samples will reduce a firm's efficiency score. The reason is that as the sample size increases, it becomes more likely that a firm will dominate the firm in question.[82] Again, this demonstrates that DEA benchmark measures can be affected by the performance of a single firm.

---

[82] This result has been demonstrated by Zhang and Bartels; see Y. Zhang and R. Bartels (1998), "The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand", *Journal of Productivity Analysis*, 9: 1877-204.

Data-related problems and the uncertainty of benchmark measures are likely to be greater with international samples. With international data, there is a higher probability that variables will be defined and measured differently across countries. Researchers must take great care to ensure that data are comparable in international benchmarking. Even the most conscientious researcher may have difficulty making data series entirely comparable between countries. Because of its nonparametric nature, non-comparable or otherwise erroneous data are likely to have a much bigger impact in DEA than in econometric studies.

In this regard, the recent decision by the Netherlands energy regulator to use DEA rather than statistical methods for benchmarking in that country is noteworthy. The regulator based this decision on the fact that there were limited data in the country (20 data points), and statistical methods are not precise with such small sample sizes. However, it is *possible* to obtain point estimates of cost function parameters using as few as 20 data points, but statistical analysis is also likely to show that these estimates are very imprecise.[83] DEA will not present information on the confidence associated with DEA-based benchmarks, but there is no *a priori* reason to believe that DEA uses a small number of data points to generate more precise benchmarks. Indeed, it is fair to say that with econometrics, the imprecision with small sample sizes is made plain, while this imprecision simply remains unknown under DEA.

The regulatory implications of data errors and uncertainty are also worth noting. With DEA, problematic data are more likely to lead to outliers that directly affect efficiency measures. Bad data can therefore be translated directly into incorrect inferences on efficiency and, ultimately, bad regulatory policy. With econometrics, "noise" in the data will likely lead to less precise estimated benchmarks. This will be reflected in wider confidence intervals around the benchmarks, which should make regulators less confident about adopting this benchmark as the basis for public policy. Hence, another disadvantage of DEA relative to econometric benchmarking is that its diminished ability to deal with uncertainty can lead to unfortunate policy decisions.

---

[83] However, even to estimate cost function parameters with such small sample sizes, it may be necessary to limit either the number of independent variables and/or restrict the form of the cost function.

Restrictions on Production Process

While DEA does not directly restrict the relationship between utility cost and business condition variables, it can involve other problems in terms of correctly specifying the production process. One is that you need *a priori* knowledge to categorize a variable as an input or an output in DEA models. This may be straightforward in many businesses, but it is not always the case for power distribution. One example of this, whether km of line is treated as an input or an output, has already been discussed. Such incomplete specifications necessarily reduce the quality of DEA results.

In addition, DEA results depend on the *number* as well as the choices for inputs and outputs. Increasing the number of variables in DEA studies generally makes it more difficult to identify peers for any individual firm. This can lead to artificially high efficiency measures.

DEA can overcome this problem through second stage regressions that relate DEA efficiency scores to other business conditions variables. These are typically Tobit regressions.[84] However, it is known that second stage Tobit regressions will lead to biased estimates for business condition parameters if these variables are correlated with the inputs used in DEA. Careful modeling may be able to reduce this problem, but there can still be significant correlations between inputs used in DEA models and business conditions used in Tobit regressions. Two possible examples are km of line (input) and population density (business condition), and O&M costs (input) and percent of kWh sales to residential customers (business condition).

A second stage Tobit will also impose a functional relationship between the efficiency measure and the business conditions. This appears to undercut one of DEA's advantages, that there is no need to specify a functional form for the cost or production relationship. A functional relationship appears to be implicit when a function is specified that relates efficiency to business condition variables, since the efficiency measure is itself derived from DEA's input-output analysis. This relationship may be even more ad hoc than flexible form

[84] The assumptions needed to implement simpler regression methods, such as generalized least squares, are not satisfied when DEA scores are used as the dependent variable in a regression.

cost functions, which are disciplined by economic theory and place a minimum of restrictions on the underlying production relation.

Problems with controlling for Differences in Business Conditions

DEA may not control for differences in business conditions as well as econometric methods. Some reasons are suggested above. DEA must often limit the number of business conditions considered, and second stage regressions may yield biased estimates of business condition parameters. Also, because DEA is a non-statistical approach, it may be more difficult to select the right set of business conditions. With econometric methods, one can test the statistical significance of different business conditions on utility cost. This provides a straightforward criterion for judging whether a given business condition should be included in the analysis.

The treatment of service quality represents a particularly nettlesome business condition for utilities. There are clear cost-quality tradeoffs in electricity distribution. Network managers make inter-related decisions about optimizing cost and reliability. This optimization process is influenced by other business conditions that the utility faces. In other words, the cost-quality tradeoff confronting a distributor will vary depending on its other business conditions. Rural utilities, in particular, face circumstances that tend both to raise the cost and reduce the quality of their service.

It is not clear that DEA is a subtle enough benchmarking tool to model these relationships. Indeed, simply adding a service quality output to a DEA model may further bias results. For example, suppose a rural distributor has a low DEA efficiency score relative to urban distributors because it requires more inputs to provide the same level of output. If service quality is added as an output, the rural distributor's performance is likely to look even worse. The DEA model will now show the rural utility is providing fewer units of the quality output relative to urban networks. All else equal, this further reduces the DEA score. This is not a reasonable result, since rural operating conditions *per se* will tend both to raise costs and reduce quality (at a given level of cost).

In principle, econometric cost functions may be able to capture this inter-relationship. For example, econometrics can model distributor behavior so that it involves simultaneous decisions on cost and quality levels. Higher quality can only be obtained at higher cost, with the cost-quality tradeoff itself influenced by other business condition variables. This optimization problem can be solved for equilibrium cost and quality levels as a function of exogenous business conditions, and these equations can then be estimated simultaneously. While this is a complex problem, it reflects distributors' real behavior and thus should be explored in benchmarking analysis. To be honest, this has not been the case to date, and econometric benchmarking studies have relied on much simpler models of utility behavior. Nevertheless, it is possible to see how econometric models can reflect these complexities, but it is not clear how it can be done in DEA. This is an important issue, since managing the complex relationships between cost and service quality is central to the power distribution business and thus should be reflected in network benchmarking.

A.3.5 Some Evidence from Other Markets

To provide further evidence on the merits of benchmarking approaches, it would be desirable to check the robustness and reasonableness of specific benchmarking results. A recent paper by Bauer et. al has proposed some criteria for evaluating the robustness and reasonableness of benchmarking results.[85] We believe these criteria are also relevant for utility benchmarking and discuss them below. These criteria are:

1. efficiency scores generated by different approaches should have comparable means, standard deviations and other distributional properties\
2. different approaches should rank institutions in approximately the same order
3. different approaches should identify mostly the same institutions as either "best practice" and "worst practice"
4. approaches should demonstrate reasonable stability over time; that is, approaches should tend to identify the same institution as relatively efficient or inefficient in different years, rather than varying markedly from year to year

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

5. efficiency scores generated by different approaches should be reasonably consistent with competitive conditions in the market
6. measured efficiencies should be reasonably consistent with standard non-frontier performance measures, such as return on assets or the cost/revenue ratio

The authors say that the first three conditions can be thought of as measuring the degree to which different approaches are mutually consistent. This can also be viewed as measuring the robustness of different benchmarking approaches. If one benchmarking technique produces results that are inconsistent with those from other methods, this may be evidence that this benchmarking approach does not generate accurate or reliable efficiency measures.

The authors view the last three criteria as measuring the degree to which different benchmarking approaches are consistent with reality or are believable. For example, one would not generally expect a company's efficiency to swing wildly from year to year. There are a number of reasons for this, including the fact that managers and management practices turn over slowly and capital equipment is often adjusted gradually. These factors should produce relative stability in efficiency measures in closely related time periods.

Similarly, one would expect some correlation between a firm's measured efficiency and its financial performance. This is natural since greater cost efficiency leads directly to higher returns. It would also be surprising if firms' efficiency rankings differed substantially from their standing in the marketplace since greater efficiency gives firms an edge over their market rivals. Benchmarking assessments that do not comply with these criteria become less believable. The authors summarize the criteria by saying that "the former (*i.e.* first three criteria) are more helpful in determining whether the different approaches will give the same answers to regulatory policy questions or other queries, and the latter (three criteria) are more helpful in determining whether these answers are likely to be correct."[86]

However, Bauer et al's last two criteria may have less relevance when applied to benchmarking results for networks. A network's financial performance can be affected as

[86] *Op cit*, p. 87.

much by regulatory decisions as the firm's own performance. Indeed, regulators in different jurisdictions often reach different conclusions on allowed rates of return for the companies they regulate.[87] Nevertheless, these remain important criteria for evaluating the general reasonableness of benchmarking techniques. If certain methods tend to yield results that are not believable when applied in competitive markets, the results may be similarly unbelievable when applied to networks.

There are few examples of utility benchmarking that can be judged according to the Bauer et. al criteria discussed above. However, these authors do apply the six criteria they establish to a common dataset. Surprisingly, few past researchers have done this.[88]

Bauer et. al analyze a dataset of 638 US banks over the 1977-1988 period. They applied several frontier estimation models to each bank in this dataset and compared the results. The primary benchmarking alternatives they considered were SFA, two related econometric methods, and DEA. Thus while this paper does not provide evidence on every external benchmark discussed here, it is useful for evaluating the reasonableness of benchmarking results using econometric methods and DEA.

The first criterion was whether the benchmarking methods yielded similar results in terms of the means and standard deviations of efficiency scores. They find that the three econometric approaches yield broadly similar and consistent results. However, there is a significant discrepancy between the econometric and DEA approaches. The mean of the DEA-based efficiency scores for financial institutions is significantly lower than that using SFA (0.30 versus 0.83, respectively, on a scale from zero to one). DEA also produces more variability in efficiency scores across firms, with a standard deviation of 0.14 versus 0.06 using econometric methods. Efficiency scores derived through DEA therefore display twice as much variability across firms as those derived from SFA.

The second criterion is the rank order of efficiency scores using different methods. These results are similar to those above. There is a strong correspondence in efficiency

[87] These determinations affect prices in both explicit rate of return regimes and under "building block" approaches to CPI-X regulation.

[88] Bauer et al review past comparative benchmarking approaches that had been applied to common datasets for financial institutions. They identify three such studies. These studies had mixed results, and none evaluated benchmarking alternatives on the basis of all six criteria. *Op cit*.

rankings between the econometric approaches but a weak correlation between the DEA and econometric approaches.[89] The results were similar on the third criterion of identifying the best and worst-practice firms

In this sample, these results suggest that benchmarking using DEA or SFA tends to produce significantly different results. In general, DEA leads to lower and more variable efficiency estimates than those resulting from SFA. DEA and SFA also lead to different rankings of firms according to their estimated efficiency, and to different identifications of best- and worst-practice firms. The authors conclude that "DEA and parametric models cannot be relied upon to generally rank the banks in the same order, and so may give conflicting results when evaluating important regulatory questions."[90] They further find that "..the two types of approaches were not consistent in their identification of the best-practice and worst-practice firms. As a result, regulatory policies targeted at either efficient or inefficient firms would hit different targets, depending upon which set of frontier efficiency approaches was used to frame the policy."[91]

Since econometric and DEA approaches tend to yield internally inconsistent results, the next issue is which set of results tends to be more reasonable. We examine this with respect to criteria four through six. The authors state that these criteria should be used to evaluate the consistency of benchmarking results with "reality."

The fourth criterion is how stable efficiency scores are over time. The authors find that all of the methods tend to produce relatively stable efficiency scores over time. Hence all of these benchmarking techniques tend to believable according to this criterion.

This is not the case for the last two criteria. The fifth criterion is consistency with market conditions. Here, the authors find the econometric results to be much more plausible than the DEA results. For example, using the DEA approach, over 90% of banks have measured efficiency of 30% or less. The SFA-based efficiency scores of most banks have measured efficiencies of 90% or more. The authors write:

---

[89] Rank order correlation of only 0.098 across the parametric and non-parametric models. There were also significant differences in rank order correlations in ten of 14 models.

[90] *Op cit*, p. 104.

[91] *Op cit*, p. 106. It should be noted that the same rank ordering and identification of best and worst practice firms would occur under both SFA and econometric cost functions, therefore these approaches are internally consistent. They differ only in terms of magnitudes of estimated inefficiencies.

> it seems fairly clear that the parametric approaches are generally more consistent with what are generally believed to be the competitive conditions in the banking industry. The relatively high efficiencies for the vast majority of banks seem consistent with a reasonably competitive industry in local markets which allowed entry by branch banking…moreover, all of these firms survived branching competition over at least a 12-year period of economic turbulence in the industry, which would be difficult to achieve for firms which consumed many more inputs than the best practice banks.
>
> In contrast, the DEA result that the vast majority of firms have measured efficiency of less than 30% does not seem to be consistent with competitive conditions in this industry. One potential explanation of this finding is that DEA does not take account of random error as the parametric approaches do.[92]

This finding was buttressed by the results on the sixth criterion. This criterion was the consistency of the efficiency measures with financial measures. The authors find "the parametric-based efficiencies were generally consistent with the standard (financial) performance measures, but the DEA-based efficiencies were much less so."[93]

Overall, these results suggest that DEA and econometric methods yield much different efficiency measures, but only the econometric measures tend to be believable and consistent with reality. The authors rightly caution that this only a single study, and it should not be used to draw general conclusions about the desirability of alternative benchmarking techniques. That warning is certainly relevant here, for these results were based on benchmarking applied to financial firms rather than networks. Nevertheless, we believe that both this research approach and the authors' findings are valuable. The Bauer et. al paper presents a well-developed framework for analyzing the results from different benchmarking methods, and this framework can be usefully applied to energy networks as well.

---

[92] *Op cit*, p. 107.

[93] Op cit, p. 109.

Appendix Four: Reviews of Performance Measures and Standards in Different Jurisdictions

This appendix will discuss performance measures, techniques and standards, and associated incentive mechanisms, that have been applied in a number of different jurisdictions. We begin with the UK, which has a long history with incentive regulation. This history for electricity distributors up to, but not including RIIO, is briefly summarized in this section, but special emphasis is placed on two examples of performance benchmarking and associated standards: 1) the econometric benchmarking of O&M costs; and 2) the "information quality incentive," which combines capital expenditures benchmarking with a type of menu approach.

Following the UK, we discuss the (unsuccessful, but instructive) comprehensive benchmarking of power distributors in the Netherlands; the benchmark-based earnings sharing mechanism, called the "Performance Evaluation Plan" or PEP, that Mississippi Power has operated under for nearly a quarter century; the targeted benchmarking studies (against the company's own history, and national and international peers) that have been used to set rates for gas distributors and natural gas pipelines in Mexico; and the service reliability benchmark mechanism established for electricity distributors in Victoria, Australia.

A.4.1 UK

Utilities in the UK have been subject to incentive regulation since the early 1980s. Most British utilities were formerly public enterprises and were subject to privatization and formal regulation beginning in 1984 with British Telecom (BT). Since then, privatization has extended to the nation's electric, gas, water, airport and rail utilities.

The decision to use rate indexing in British utility regulation was strongly influenced by the recommendations of Stephen Littlechild of the University of Birmingham, in a report released in 1983.[94] He proposed to adjust BT's rates using an index with an "RPI-X" formula. The RPI term is the inflation in the Retail Price Index (RPI). A specific value for X

[94] Stephen Littlechild, *Regulation of British Telecommunications' Profitability: Report to the Secretary of State*, February 1983.

was not recommended, nor was there significant discussion in Littlechild's paper of the appropriate framework to be used to determine X. Rather, the value for X was described vaguely as "a number to be negotiated."

Following its application to BT in 1984, RPI-X regulation was first applied to the gas industry in 1986 and to the electric utility industry in 1990. The electricity industry in England and Wales was unbundled into a separate power transmission firm (National Grid) and twelve distribution network operators (DNOs) when industry restructuring was completed in 1990. The two DNOs serving Scotland were originally part of vertically-integrated firms. The gas utility industry was initially served by a single regulated firm, British Gas, which also had gas production and other interests. In the mid 1990s, the gas transmission and distribution operations of British Gas were functionally unbundled into a firm called Transco. UK gas distribution operations were later formally unbundled into eight separate local gas distributors, four of which were retained by the original entity (which had since merged with National Grid) and four of which became stand-alone utilities. The first price review for the UK's unbundled gas distributors was recently completed in 2007.

RPI-X regulation for UK energy distributors has employed a "building block" approach that calibrates the terms of the indexing formula based on forward-looking revenue requirements of each regulated firm over the term of the price controls. The earliest energy price reviews were rather opaque and did not provide much detail on the regulators' specific determinations on particular "building block" elements. Over time, however, UK regulatory reviews have become more transparent and followed a more clearly defined and organized process.

The first fully articulated statement of the British approach towards price cap regulation is contained in the 1997 price cap plan for Transco. To determine the price controls for Transco, the regulator took as a "starting point" a long term net present value (NPV) calculation.[95] This calculation determined "a level of revenue which, when set against expected expenditure (over the term of the controls) and discounted at the company's cost of

---

[95] There were separate regulators of the gas and electricity industries until 1999, when the regulatory agencies were merged to form the Office of Gas and Electricity Markets (Ofgem).

capital, would produce a net present value (NPV) of zero".[96] In other words, price controls were based on a projected forward-looking revenue requirement that just recovers the sum of opex and capital costs (return on and depreciation of existing assets plus costs of new capital expenditures) for the price cap period. More specifically, the basic components of the basic building method are:

1. Defining the regulatory asset base (RAB). The approach that ultimately developed was based essentially on the (conventional) historic cost of assets.
2. Estimating depreciation of the RAB
3. Assessing future capital expenditure (capex) and its depreciation
4. Estimating the weighted average cost of capital (WACC).
5. Determining a reasonable level of future operating expenditure (opex)

New price controls are almost always affected via two price adjustments: an initial price (P_0) change in the first year of the plan; and an X factor that applies during the subsequent plan years, when index-based price changes take effect. The building block approach used in the UK can lead to any number of initial price adjustment-X factor combinations for a company that are consistent with that company's allowed revenue adjustment over the term of the controls. Any revenue neutral reallocation between initial price adjustments and X factors (*i.e.* any change between the P_0 and X factor that does not affect the NPV of the company's expected revenues over the term of the price control) is consistent with the regulator's building block computations.

The UK incentive regulation experience is extremely rich and diverse, but the most relevant precedents in the context of 3rd Generation IRM in Ontario are the plans that have been approved for the UK power distribution industry. Five-year price cap plans were instituted for the DNOs upon their privatization in 1990. Initial rates were set at the levels charged by the companies just before privatization, even though these rates presumably reflected inefficiencies under state ownership. Different X-factors were established for each DNO, ranging from 0 to -2.5% with an average value of -1.3%. Therefore, DNOs' distribution prices were allowed to *increase* by an average of 1.3% per annum in real terms during the five years of the first price cap plan. The reasons for allowing real price increases

[96] Office of Gas Supply, *Price Control Review, British Gas' Transportation and Storage: A Consultation Document*, June 1995, p. 22.

were not made explicit. However, the companies were being sold to private investors. The terms of the indexing plans were likely set, in part, to spur investor interest and extend share ownership.

DNO price controls were first reviewed in 1994. This review focused on four considerations when re-setting allowed revenues over the upcoming price control term: operating expenses, planned capital expenditures, the valuation of the capital stock used in power distribution, and the allowed return on that capital stock. The Office of Electricity Regulation (Offer) reviewed these factors by analyzing the DNOs' cost and sales data and by soliciting independent evaluations of REC operations. For example, consultants provided opinions on "best practice" for different distribution functions, and outside analysts estimated the costs of network expansions given projected changes in the number and location of customers. Statistical benchmarking studies were undertaken to estimate the efficient levels of operating costs for individual DNOs given various factors beyond management control. These included the number of customers served, volumes distributed at low and high voltage, and customer density within the territory served. The results of these benchmarking studies were not made public, nor did the regulator detail how the benchmarking results specifically affected the final X factors.

The outcome of this review was an initial price cut for each of the DNOs and a common X-factor of 2%. Distribution rates were cut either 11%, 14%, or 17% in the initial year of the new plan, depending on what the benchmarking and other analyses indicated were efficient cost levels for the company. Revenue reductions were divided between an initial rate cut and a higher X because it was believed that both customers and utilities preferred this approach.

The new price cap plan took effect in April 1995 and was widely viewed as too generous for the Companies. Public dissatisfaction was heightened when outside investor groups responded to the new price controls with takeover bids for several DNOs, allegedly because the new price controls offered the opportunity for unexpected profits. Only one month after the distribution price cap plan went into effect, the Director General (DG) re-opened the plan, which led to an additional, up-front price cut of 9% and an increase in the X factor to 3%.

The DNOs distribution price control was updated again in 2000. This led to another initial price cut that varied between 19%, and 33% between companies. The X factor in the other four years remained at 3%. The methods used to update the control were similar to those used in 1995.

The 2005 update of DNOs distribution prices included an initial price increase that averaged about 1% per company and an X factor of 0 for the remaining four years of the control. Unlike the earlier power distribution price reviews, prices did not decline in real terms as a result of this review. The main reason was that Ofgem allowed substantial increases in capital spending for many of the distributors.

Over time, benchmarking has played an increasingly important role in the regulation of opex in UK RPI-X plans. Ofgem has primarily relied on econometric benchmarking in its price reviews. Its econometric benchmarking approach is a variant of corrected ordinary least squares (COLS). For price controls taking effect in both 2000 and 2005, Ofgem regressed a "normalized" measure of opex on what it called a "comprehensive scale variable" (CSV). Distributors' opex data were normalized by ensuring that these data were defined and collected comparably across all DNOs. The CSV was based on each DNO's number of customers served, kWh distributed, and network length. The weights applied to these variables in developing each DNO's CSV were 25%, 25%, and 50%, respectively. These weights differed from those used in the 1999 COLS study, which were 50% for customers served, 25% for kWh and 25% for network length. These weights were considered roughly proportional to the impact of each scale measure as a "driver" of distribution opex.

In two dimensional space, COLS is normally applied by running an OLS regression and shifting the intercept of that regression until the line passes through the minimum observation. Any gap between a DNO's opex and this COLS line would therefore reflect that DNO's inefficiency, or the excess of its opex costs over the observed minimum regression line. For the 2000 review, however, Ofgem's COLS benchmarking was done by shifting the *slope* of the estimated function and not the intercept. The slope was shifted until the line passed through the *second* lowest observation. This approach was taken because Ofgem believed a conventional COLS application would have led to implausible results. That is, the intercept from a regression of (normalized) opex on the CSV could be interpreted as the fixed operating costs of a DNO, independent of the size of its operations. In the 2000 review, Ofgem believed that if the intercept was shifted as in a typical COLS procedure, it would have produced a fixed opex cost estimate that was implausibly low from an engineering perspective, so Ofgem shifted the slope as an alternative.

For the 2005 review, Ofgem did shift the intercept in its COLS application as is typically the case. However, the intercept was shifted so that the line passed through the upper quartile opex performance rather than minimum performance. Upper quartile performance was effectively determined as the midpoint between the third and fourth lowest opex cost observation of the 14 DNOs.

In the 2000 review, Ofgem set opex targets by assuming that companies would catch-up to the opex target determined by the COLS procedure by closing 75% of the gap between their (normalized) operating cost and the normalized opex of the second most efficient firm in the UK by the second year of the price review.[97] In the 2005 review, each REC's allowed opex is based on an upper quartile benchmark within the UK. Ofgem's rationale for this decision is that an "upper quartile benchmark…provides a more robust and sustainable benchmark than a frontier based on one company."[98] The 2005 review also undertook some data envelope analysis (DEA) as a "cross check" on the econometric results. However, Ofgem concluded that the DEA results "are not plausible so it (DEA) has not been incorporated directly."[99]

The regulation of capex has also changed considerably since the initial RPI-X controls but has evolved in a different direction. In the 2005 price review, Ofgem applied a sliding scale mechanism to the UK distribution companies' capital expenditures. A similar type of mechanism was applied in the most recent energy price control review for the gas distributors but was called an "information quality incentive." These mechanisms were motivated by Ofgem's view that the distributors have incentives to inflate their forecast capex during the next price control period but then "underspend" once an allowed capex is used to set the

[97] Office of Gas and Electricity Markets, *Electricity Distribution Price Control Review: Initial Proposals*, June 2004, p. 66. "Normalized" cost here refers to costs that are adjusted for scale of output and other factors that are quantified through econometric benchmarking.

[98] Ofgem, *op cit*, p. 67.

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

value of X. Ofgem believes some utilities have actually behaved in this way, although others have not. The aims of the sliding scale mechanism are to:

- retain incentives for efficient capital spending during all years of the control
- reduce the emphasis on Ofgem's or its consultant's view of the appropriate level of capex
- reduce the perceived risk that the price control causes under-investment
- allow but not encourage expenditure in excess of the allowance
- reduce the possibility that companies submitting high capex projections will make very high returns from underspending
- reward companies making "low" capex forecasts
- avoid incentives to underspend in ways that reduce service quality or create service quality problems in subsequent years

The sliding scale mechanism essentially gives companies a choice between:

- a lower allowance for capex reflected in the controls, but with a higher- powered incentive that allows them to retain a greater share of "underspend" relative to the allowance and collect a greater share of "overspend"; or
- a higher allowance for capex in the controls, but with a lower-powered incentive that lets companies keep a lower share of "underspend" and collect a lower share of "overspend."

Companies also get an additional reward if they do choose the lower allowed capex option, but do not receive this reward if they select higher allowed capex. If the sliding scale mechanism is designed correctly, it is "incentive compatible" and removes incentives for the company to inflate its projected capex. The mechanics of Ofgem's proposed sliding scale mechanism are as follows:

- Ofgem determines a benchmark level of projected capex over the price control period for each DNO; in the 2005 distribution price review, these benchmarks were determined by the engineering consulting firm PB Power
- Each REC presents its actual capex projections over the price control period
- Ofgem determines a capex *allowance rate*, *additional income* and a capex *incentive rate* depending on the relationship between benchmark and forecast capex. The allowance rate is the total amount of capex that will be allowed in the controls; this number is specified as a multiple over the benchmark level.

The additional income term is an addition to the distributor's allowed revenue. The incentive rate is equal to the portion of capital "underspend" the company is allowed to retain. The allowance rate, additional income and incentive rate each increase as the company's forecast gets closer to the benchmark level, and vice versa. This approach therefore rewards companies for keeping their capex forecasts low.

For example, if a company's projects its capex to be 140% of the PB Power benchmark, their capex allowance rate is 115% of the PB Power forecast value. If they over- or underspend relative to this forecast, they get to keep or bear 20% of the difference *i.e.* the marginal incentive rate is 20%. Alternatively, for companies whose capex forecasts are equal to or less than the PB Power benchmarks, their allowance is set at 105% of the PB Power capex forecast. Companies keep or bear 40% of any over- or under-spend relative to the allowed capex level, so their marginal incentive rate is 40%.

Ofgem established the sliding scale mechanism as a matrix which displays the values of the key parameters and how they vary with the forecast/benchmark relationship. The table below captures the main features of the sliding scale matrix.

Forecast (F)/ Bench (B)	Δ	Allowance Rate (AR)	Δ	Incentive Rate (IR)	Δ	Additional Income (AI)	Δ
100		105		0.4		2.5	
105	5	106.25	1.25	0.38	-0.02	2.1	-0.4
110	5	107.5	1.25	0.35	-0.03	1.6	-0.5
115	5	108.75	1.25	0.33	-0.02	1.1	-0.5
120	5	110	1.25	0.3	-0.03	0.6	-0.5
125	5	111.25	1.25	0.28	-0.02	-0.1	-0.7
130	5	112.5	1.25	0.25	-0.03	-0.8	-0.7
135	5	113.75	1.25	0.23	-0.02	-1.6	-0.8
140	5	115	1.25	0.2	-0.03	-2.4	-0.8

The first column shows the ratio between forecast and benchmark capex (in percentage terms). The second column (the "delta") presents the change in the forecast/benchmark ratio from the row above. The third column presents the allowance rate (AR, also in percentage terms) associated with a given forecast/benchmark ratio; this allowance rate is multiplied by the benchmark capex value, and the product determines allowed capex. The fourth column presents the change in the AR from the row above. The fifth column presents the incentive rate (IR) for a given forecast/benchmark ratio; this incentive rate is multiplied by the difference between allowed and actual capex value. The sixth column presents the change in the IR from the row above. The seventh column presents the additional income (AI) associated with a given forecast/benchmark ratio. The eighth column presents the change in the AI from the row above.

In some ways, the UK approach to incentive regulation must be seen as a success. It is indisputable that price cap regulation in the UK has delivered considerable benefits to British consumers. There have been substantial declines in prices for all regulated utility services in Britain (except water, where there has been substantial new investment to comply with EU water quality standards) since RPI-X controls took effect. The British "building block approach" to price cap regulation can create some incentives for firms to pursue efficiency gains and, over time, these efficiency gains have been distributed to customers in the form of price reductions.

Other aspects of the British approach are also appealing. The sliding scale mechanism that is being applied to capex should help to diminish the incentives to game capex forecasts. Developments regarding the actual operation of this scheme merit attention.

The econometric approach to benchmarking opex has also worked reasonably well, although the econometric models and methods have been extremely simple because of the regulator's decision to rely only on data from the limited sample of UK DNOs. Richer econometric specifications (for both opex and total distribution cost) can be estimated using the much more ample data from North America. The upper quartile benchmarking standard that was applied in the 2005 distribution price review is also appealing and generally consistent with a competitive market paradigm. It is not reasonable for regulators to expect all firms in their industry to be performing at frontier levels, or to set the terms of price controls so that firms earn their cost of capital only by achieving frontier performance standards. In competitive markets, firms that are on the frontier earn above average returns. If regulation is designed to emulate the operation of competitive markets, then the appropriate performance standards must also be set at less than the frontier. Equivalently, firms must have "room" to outperform the standards reflected in the price controls for them to have incentives to boost their efficiency and thereby earn more than their weighted average cost of capital. The upper quartile standard chosen by Ofgem is ultimately based on judgment, but it is generally consistent with this competitive market paradigm.

There are also disadvantages associated with UK, building block regulation. One is that the building block model is susceptible to gaming on the part of companies. Prices are based on a company's projected costs. Companies therefore clearly have incentives to game the estimates of their projected costs that they present at the outset of the regulatory process. Regulators must attempt to "de-game" these forecasts and ascertain the "truth" about how much costs are actually expected to increase over the term of the controls. This is an inherently imprecise exercise which necessarily exposes regulators to the well-known "information asymmetry" problem, since regulators will know far less about the company's actual and projected costs than the companies themselves. Ironically, economists have long believed that information asymmetries are at the heart of problems with cost of service regulation. Incentive regulation is therefore designed to create regulatory institutions that encourage companies to use their superior information in a socially beneficial manner; it should not allow companies to profit by gaming this information through other channels. The UK has created elaborate sliding scale or information quality incentive mechanisms to counter this problem, but developing and implementing such mechanisms is likely to be difficult and costly in Ontario, particularly since separate capex benchmarks would need to be developed for more than 80 distributors.

This reflects a more fundamental concern, which is the information-intensiveness and regulatory burdens of the building block approach. Building block regulation requires detailed cost information, on both a historical and prospective basis, for each regulated company. Implementing this approach for a large number of regulated energy networks could place considerable burdens on the regulatory process and increase the cost and complexity of regulation for all parties involved (companies, regulatory staff and intervenors). The costs of a UK-type approach to incentive regulation are therefore considerably higher than a North American-style approach, and these incremental administrative and regulatory costs would likely outweigh the incremental benefits of implementing a full, building block methodology in Ontario.

A.4.2 Netherlands

The 1998 Electricity Act created a new regulatory framework for the electricity supply industry in the Netherlands. This Act created the Director of Dienst uitvoering en toezicht Energie (DTe) as the regulatory authority responsible for determining regulated tariffs. It also required that tariffs after 2000 would be updated using a CPI-X indexing formula.

The first price control review for the Dutch power distribution utilities and TenneT (the national transmission system operator) was completed in September 2000, and set the terms of the CPI-X indexing formula for thee years (from 2001 through 2003). There were two components of the X factor. The first was a general component, which was termed "frontier shift." The value of the frontier shift was estimated to be 2% in the first review. The second component was a "catching up" factor, which varied by company depending on their estimated initial cost inefficiencies.

Benchmarking played a critical role in setting the "catching up" or company specific factor in the first price review. The reason was the DTE believed that benchmarking "can be used to predict what this efficient level of costs should be." DTe believed that prices initially differed significantly among distributors due to differences in management efficiency and tariff setting practices (*e.g.* the extent to which tariffs reflected commercial returns or, for some utilities, lower returns in order to achieve broader social objectives). Benchmarking for the first price control was designed to set tighter price controls for companies that were either less efficient or had historically set higher tariffs in order to generate higher returns.

The benchmarking method used by the DTe in the first price review was data envelope analysis (DEA). However, the DTE stated that:

> "It is important to stress that DTe is not proposing to adopt a mechanistic approach to the determination of X factors for the companies, based solely on efficiency scores from a benchmarking exercise. The values of the X factors will be based on a number of important and interrelated issues, including the valuation of the regulatory asset base and acceptable thresholds for a number of financial ratios."

In spite of this policy, the process used to set the X factors in fact proved to be highly mechanistic. A 'raw' X factor was computed that would essentially move companies from their initial (2000) "controllable" revenues to "efficient" revenues over the three year term of the plan. “Controllable” revenues were determined by taking initial revenues and subtracting what the DTe deemed to be “excess” returns. “Efficient” revenues were essentially determined by the company’s DEA score on the preferred DEA model. For example, if a given company had a DEA score of 0.60, this would imply that its costs were 40% below the efficient frontier, and X would be set to reduce revenues by a cumulative 40% (excluding adjustments for inflation) over the three year plan. However, X factors were capped at 8% maximum per year, regardless of the company’s DEA score. Moreover, several companies pointed out material errors in the data used to compute their initially proposed X factors, and this led to revisions of several utilities’ X factors.

Nevertheless, after the DTe’s decision, all the power utilities appealed the tariffs to the Director of the Nederlandse Mededingingsautoriteit (NMa), the Competition Authority. In September 2001 NMa published its review of the price controls of the distribution network businesses. It endorsed the main technical aspects of DTe’s approach including, importantly, the choice of DEA for benchmarking. However, the NMa made a number of adjustments related to the definition of the regulatory asset base and cost of capital and some changes to the model itself. These changes significantly impacted the X factors for many of the companies, as illustrated below.

	DTe’s analysis in 2000	NMa’s analysis in 2001
	X factor (%)	X factor (%)
ENECO Group		
ENECO	8.1	7.8
Delfland	9.6	6.7
EMH	6.4	3.3
Weert	4.3	-.6
EZK	9.8	7.9
Essent Group		
Brabant	-3.5	-3.5
Friesland	9.7	7.4
Limburg	-1.5	-1.5
Noord	1.9	1.9
Mosane	6.3	6.3
NUON	7.7	7.2
REMU	8.4	6.5

The revised X factors after the NMa review still represent significant real price reductions for most of the companies, in large part because of the results of a benchmarking model. It should also be noted that these price reductions for many companies occurred even after a legal review of the regulator's decision. Finally, it is noteworthy that both the DTe and NMa used DEA modelling when determining X factor recommendations yet the outcomes of these DEA models differed significantly. This outcome suggests that DEA is not a robust benchmarking technique and can sometimes lead to unreasonable inferences on a company's cost efficiency.

During its initial price review, DTe also stated that it eventually wanted to move to "yardstick regulation", which it defined as allowing the annual change in prices for each distributor to follow the average change in costs of the industry as a whole. While DTe did not believe it was possible to start with this approach in 2001 because of the varied tariff levels and perceived efficiencies among the distributors, the current regulatory approach to setting CPI-X controls for Netherlands power distributors is essentially an example of what the DTe calls "yardstick regulation." The current regulatory period began in 2007 and will extend until 2010. As discussed in the previous chapter, beginning in 2007, all Dutch distributors will be subject to a common X factor reflecting the industry's TFP trend. This value has been calculated to be 1.3%. However, the regulator has also allowed tariff adjustments for two Dutch distributors to reflect "regional differences" in costs that are not otherwise reflected in the value of current tariffs or the industrywide X factor.

A.4.3 Mississippi "PEP"

Mississippi Power Company (MPC) has operated under a Performance Evaluation Plan (PEP) since 1986. PEP ran from 1986 through 1990, and PEP-1 was in effect from 1990 until 1993. Both provided for quarterly adjustments of MPC's rates and allowed returns depending on the company's performance in a number of areas. Better performance increased MPC's allowed rate of return, while allowed returns declined when performance deteriorated. In many ways, PEP-1 is the most general of the PEP plans, so we focus our discussion on the features of this plan before turning to recent plan modifications.

The features of PEP-1 were determined through a collaborative process between MPC, Commission Staff and major intervener groups. The discussions focused on choosing a comprehensive set of performance indicators. To promote comprehensiveness, PEP-1 included several performance variables.

Operationally, PEP-1 was administered through a five-step procedure:

1. MPC's Earned Return on Equity (EROE) was computed.

2. A Benchmark Return on Equity (BROE) was calculated. The BROE was determined as the average of three financial models that are used to calculate expected utility returns.

3. MPC's performance rating was calculated from data on its performance in the targeted areas.

4. The BROE and performance rating were used as inputs to the so-called PEP-1 matrix. This matrix determined the allowed range of returns for MPC. The midpoint of allowed returns was equal to the BROE. Higher performance ratings increased the allowed earnings range by up to 100 basis points above the BROE; lower performance ratings reduced the allowed earnings range by as much as 100 basis points below the BROE.

5. The EROE was compared to the allowed range of returns given by the PEP-1 matrix and required revenue adjustments were made.

Seven performance indicators were specified in PEP-1: customer price, customer satisfaction, service reliability, equivalent availability, construction performance, contribution to load factor, and employee safety. The first three indicators were directly linked to customer welfare. The customer price indicator evaluated the level of MPC's retail prices compared with the retail prices of utilities in the Southeastern Electric Exchange (SEE). The customer satisfaction indicator measured the public's perception of MPC's service. The service reliability index measured the reliability of the MPC's power supply.

The next three indicators were focused on the efficiency of MPC's operations. The equivalent availability indicator measured the availability of generating units to produce electricity. The construction performance indicator measured the correspondence between MPC's budgeted and actual expenditures on construction projects. The contribution to load factor indicator measured the ability of MPC to utilize its capacity.

The final indicator was focused on the welfare of MPC's employees. The employee safety indicator was a measure of the safety of the MPC workplace. Safety was measured by the amount of employee time lost because of accidents.

Each indicator was evaluated against a benchmark level of performance. These benchmarks were determined through negotiation between the parties. A score between 0 and 10 was awarded for each indicator, with higher scores indicative of better performance. In most cases, the value of the indicator was based on MPC's operations over the 12-month period that ended in the last month of the quarter of the evaluation period.[100] Individual performance indicators were then weighted to arrive at a comprehensive performance rating.[101]

MPC was placed into one of five performance categories based on its performance rating. The performance category determined the allowed earnings range for the company. The relationship between performance ratings, performance categories and the allowed earnings ranges was as follows:

Performance Rating	Performance Category	Midpt Allowed Range
0.0 - 2.0	I	BROE less 100 bp
2.1 - 4.0	II	BROE less 50 bp
4.1 - 6.0	III	BROE
6.1 - 8.0	IV	BROE plus 50 bp
8.1 - 10.0	V	BROE plus 100 bp

Revenue changes took place based on the relationship between actual MPC earnings and the allowed earnings range. Four types of revenue adjustments could result from this comparison:

1. If EROE was above the allowed range and MPC was in performance categories I, II, III or IV, then revenues would be reduced until returns were at the midpoint of the allowed range.

[100] One exception was for the customer service indicator, which was based on the results of a semi-annual survey of customer satisfaction.

[101] The following weights were applied: customer price, 0.2; customer satisfaction, .15; service reliability, .16; equivalent reliability, .16; construction performance, .11; contribution to load factor, .11; safety, .11.

2. If EROE was above the allowed range and MPC was in performance category V, then revenues would be reduced until returns were halfway between the EROE and the midpoint of the allowed range.

3. If EROE was below the allowed range and MPC was in performance categories II, III IV, or V, then revenues would be increased until returns were at the midpoint of the allowed range.

4. If EROE was below the allowed range and MPC was in performance category I, then revenues would be increased until returns were halfway between the EROE and the midpoint of the allowed range.

From these rules for revenue adjustments, it is clear that PEP-1 could influence MPC's earnings in two ways. First, higher overall performance ratings would tend to increase the allowed range of earnings. In addition, MPC's returns could be outside of the allowed range if the company was in either the highest or lowest performance categories. The last feature removed the upper bound (*i.e.* at the midpoint of the allowed range) on MPC's potential earnings and therefore accentuated the incentives for efficient performance. Consumers would also benefit from performance gains since rates would be reduced automatically anytime earnings exceeded the allowed range.

The operation of the PEP-1 matrix can perhaps be clarified through a few numerical examples. Using actual MPC data, in the fourth quarter of 1990 the BROE was calculated to be 12.79%. The company's performance rating was 8.7. From the PEP-1 matrix, this performance rating placed the company in performance category V. When MPC is in this category, the *midpoint* of MPC's allowed earnings is100 basis points above the BROE, so the midpoint of the company's allowed earnings was increased to 13.79%. Since there was a deadband of 100 basis points on either side of the adjusted BROE, earnings between 12.79% and 14.79% did not require revenue adjustments. The company's actual earnings rate (EROE) in the fourth quarter of 1990 was 13.476%. Since this return was within the allowed range, MPC was not compelled to adjust its rates.

Suppose now that in the fourth quarter of 1990, Mississippi Power still had a performance rating of 8.7 but registered an EROE of 16.79%. In this case, the Company's rates would be adjusted until returns were halfway between the midpoint of the allowed range (13.79%) and the EROE of 16.79%. Returns after sharing would therefore be 15.29%.

Now suppose the company's EROE remains 16.79% but its performance rating is 7.0 instead of 8.7. From the PEP-1 matrix, this performance rating placed the company in performance category IV. When MPC is in this category, the *midpoint* of MPC's allowed earnings is 50 basis points above the BROE, so the midpoint of the company's allowed earnings is 13.29%. EROE is still above the allowed range and, since the company is performance category IV, returns are adjusted until they are equal to the midpoint of the allowed range, or until they are equal 13.29%.

These examples demonstrate that the PEP plan is designed, in part, to stabilize the company's earned returns. However, the amount of the stabilization depends on the firm's performance on the benchmark mechanism. Good performance improves the allowed return and tends to "stabilize" earnings at a higher level, reflecting the performance gains. Bad performance reduces the allowed return and tends to "stabilize" earnings at somewhat lower levels.

In July 1993, several changes were made to the PEP-1 plan. For our purposes, two such changes are noteworthy. First, the amount of sharing would be related to MPC's retail prices relative to those of proximate utilities. Customers would be subject to a greater fraction of revenue adjustments when MPC prices were low relative to those of electric utilities in the southeastern U.S. The opposite would be true if relative MPC prices were high. More precisely, the company proposed that customers' share of rate adjustments be given by the formula:

$$Customer\ Share\ =\ \frac{(Avg.\ Price\ per\ kWh\ )_{SEE}\ \ -\ \ (Avg.\ Price\ per\ kWh\ )_{MPC}}{(Avg.\ Price\ per\ kWh\ )_{SEE}\ \ -\ \ (Avg.\ Price\ per\ \ kWh\ )_{LowCost}}$$

Here, the subscripts SEE, MPC and LowCost refer to the Southeastern Electric Exchange, Mississippi Power Company and the lowest cost SEE Utility, respectively. Based on the actual price data for these groups in June 1993, the customer share would have been equal to .785. Therefore, if actual earnings were less than the lower end of the Range of No Change, revenues would be increased by 78.5% of the difference between EROE and the bottom of this range. Alternatively, if a rate reduction were required, revenues would be reduced by 21.5% of the difference between EROE and the top of the Range of No Change. This application would incent MPC to reduce its prices relative to those of other utilities in the region. In fact, if MPC had the lowest prices in the SEE, it would be able to retain 100%

of earnings outside of the allowed range and could fully recover earnings below the allowed range through revenue increases.

Another change was that the number of performance indicators was reduced. The plan eliminated the construction performance, contribution to load factor and employee safety measures. The Commission found that the four remaining indicators more closely reflected the concerns of customers. These indicators were re-weighted so that the customer price indicator now contributed 40% to the performance rating. Customer service, service reliability and equivalent availability would receive weights of 20% each.

These and a few other administrative changes were approved to the plan, and the modified plan was called PEP-2. The Mississippi Commission approved this plan in October 1993. It has been in effect to the present day, although there have been a few additional modifications. For example, "PEP-2A" reduced the number of indicators to three (the generation availability indicator was eliminated). PEP-3 also led to some relatively minor adjustments in some of the empirical parameters of the plan.

The most substantial adjustments since PEP-1 occurred in the PEP-4 plan approved in May 2004. The biggest change in PEP-4 compared with earlier versions of the plan is that performance is evaluated based on a future test year instead of a historical test year. The Commission staff agreed with this change after it was persuaded that MPC has appropriate and budgeting process and budgetary controls in place to make a projected test year just and reasonable. The process for implementing the forward-looking test year also allows for greater on-going Staff scrutiny and understanding of MPC's budget process, operations, and significant budget variances. PEP-4 was also changed to allow for annual rather than semi-annual rate adjustments. This was considered a logical consequence of the switch to a forward-looking test year and is also designed to reduce administrative burdens on both the company and Commission. PEP-4 also further refined the measurement and benchmarks of the reliability, price and customer satisfaction indicators. The weights applied to these indicators for computing the performance rating were also revised to be 40%, 40% and 20%, respectively.

A.4.4 Mexico

Mexico has extensive natural gas reserves, but a lack of available capital has historically prevented gas distribution infrastructure from being developed. In 1996, however, Mexico passed natural gas legislation that was intended to jump-start the natural gas industry. The Mexican Energy Regulator operates subject to a regulatory Directive (the *Directiva sobre la Determinacion de Precios y Tarifas para las Actividades Reguladas en Materia de Gas Natural*). This Directive contains specific requirements that the regulator must satisfy when updating X factors. Specifically, Section 6.4.2 of the Directive says that when the CRE sets a new efficiency factor for a concession winner it will take into account the following:

- The expected gains in operating efficiency for the next five years
- Other factors that effect the firm's unit costs, such as investment programs

On considering the expected efficiency gains of the firm, Section 6.43 of the Directive says the regulator must consider

- The company's own historical efficiency trends
- International efficiency standards in the industry
- Long run TFP indexes
- Economies of scale
- Comparisons with other firms established in Mexico

The regulatory framework in Mexico therefore directs the regulator to consider both domestic and international evidence when re-setting X factors. In addition, the regulator must consider the company's own efficiency and international efficiency standards and/or benchmarks when re-setting the value of X. As a result, the regulatory methods used in Mexico draw upon a range of diverse information sources and bring this information to bear through interesting "hybrid" approaches towards setting X factors.

The incentive regulation model in Mexico is similar in many respects to that used in Britain. Like the British building block approach, updated prices depend in large part on the expected costs of the regulated company itself. However, Mexican regulation also makes extensive use of benchmarking analyses when setting a company's allowed operating expenditures (opex).

We briefly review three sets of X factor reviews in Mexico. The first is for the gas distribution companies. The second is for the natural gas pipeline Gasoductos de Chihuahua. The third is for the natural gas pipelines Energia Mayakan and Sistema Nacional de Gasoductos (SNG).

These reviews relied heavily on two sets of benchmarking analyses for opex. The first examined the "intrinsic efficiency" of the regulated company in their initial incentive regulation plan, which was called the first quinquenium (Q1). This was essentially used to adjust the starting point for the company's opex at the outset of the next controls. The regulator also examined the "relative efficiency" of regulated company during Q1. Both analyses were used to set X factors for each distributor for the upcoming incentive regulation plan, termed the second quinquenium (Q2).

The most elaborate benchmarking work in Mexico took place during the gas distribution review. For the intrinsic efficiency analysis, the energy regulator (the CRE) undertook some statistical analysis on the relationship between opex and various opex cost "drivers" for its sample of Mexican distributors. Some of the main cost drivers identified were total customers, total energy delivered, peak gas demand, and the total km of gas distribution main. After identifying these as the main outputs, DEA analysis was performed on the CRE's sample of Mexican gas distributors. The sample consisted of five annual observations for 12 distributors. For each distributor, the CRE would identify which of the five years during Q1 yielded the company's highest DEA score. The company's opex in that year would then be taken as the initial value of opex at the outset of Q2.

The CRE then examined the relative efficiency of the distributors. Here, the emphasis was not identifying the company's best historical performance but rather examining how each distributor performed, on average, relative to the frontier for the Mexican industry. A DEA score of 1 indicates a company is on the frontier. Any value less than 1 measures the relative gap between the company's efficiency level, as computed by DEA, and the frontier. For example, a DEA score of 0.8 would indicate a company had an efficiency level equal to 80% of frontier efficiency or, equivalently, efficiency 20% below the performance frontier. CRE estimated six different DEA models and obtained average DEA scores for each Mexican distributor subject to the review. The Commission then assumed that each company would "catch up" to the opex frontier over the five-year term of Q2. Thus in the previous example, if a company had a DEA score of 0.8, it was assumed that there would be a 20% opex reduction in Q2, which translates into about a 4% per year.

The gas distribution review translated each company's opex target directly into an X factor. Opex accounts for only a share of the total costs subject to CPI-X adjustments, so the final X factors would be a fraction of the total targeted annual reduction in opex costs. CRE multiplied each annual target reduction by 20%, roughly corresponding to the share of opex in gas distributors' total cost. Hence in the example above, the 4% target opex reduction is equivalent to an X factor of 0.2 * 4% = 0.8%.

The X factor review for Gasoductos de Chihuahua (GDC) took a similar approach but applied different benchmarking methods and data than the gas distributors' review. Both reviews included analysis of the utility's own intrinsic efficiency and relative efficiency. One difference between the reviews is that relative efficiency review for the gas distributors was determined with reference to the national gas distribution industry. For GDC, relative efficiency was evaluated with reference to a sample of 15 US gas transmission pipelines. Different benchmarking techniques were also applied. DEA was used in the gas distribution review, along with some auxiliary statistical analysis, while the gas transmission review relied on unit costs benchmarks computed from the international dataset.

It was initially proposed that the analysis of GDC's intrinsic efficiency focus on two unit cost measures. These metrics were opex per unit of comprehensive output and capex per unit of comprehensive output. While this was the initial proposal, the CRE ultimately applied the proposed methodology only to setting an X factor for allowed opex. Allowed capex was determined in a manner very similar to what was done for the gas distributors. That is, the ultimate "volume" of investment needed was negotiated and agreed with GDC, and the indexing formula contained a separate factor to reflect new capital investment needs.

The comprehensive output measure was computed as the product of GDC's pipeline capacity (measured in million cubic feet per day, or MMCFPD) and total length of pipe (measured in km). The unit cost benchmarks were therefore opex/(MMCFPD*km) and capex/(MMCFPD*km). GDC's intrinsic opex efficiency was given by the lowest observed value of opex/(MMCFPD*km) during the five years of Q1. Similarly, GDC's intrinsic capex efficiency was given by the lowest observed value of capex/(MMCFPD*km) during the five years of Q1.

The initial proposal was that the X factor consistent with the company's intrinsic efficiency was to be determined through the following process. First, the unit cost benchmarks were multiplied by GDC's projected (MMCFPD*km) for Q2; this latter value was determined through extensive review of the company's Q2 business plan. Multiplying the unit cost benchmarks by projected (MMCFPD*km) yielded total values for opex and new capex for Q2. The capex was then added to the capital stock determined to exist at the outset of Q2. The capital costs associated with this capital stock were computed by multiplying this capital stock by what the consultants and CRE determined to be a reasonable rate of return and depreciation rate, respectively. The revenue requirement over Q2 was simply equal to the sum of the intrinsically efficient opex plus intrinsically efficient capital cost. Finally, the X factor was determined by examining the relationship between GDC's costs at the outset of Q2 compared to these Q2 "intrinsic efficiency" costs.[102] "X" is essentially the slope associated with moving from the costs at the outset of the control with the intrinsically efficient costs over the term of Q2.

The proposed evaluation of relative efficiency was similar. The unit cost benchmarks were still opex/(MMCFPD*km) and capex/(MMCFPD*km), and the process for translating these benchmarks into an X factor is the same as described above. In this case, however, the benchmark values computed for opex/(MMCFPD*km) and capex/(MMCFPD*km) were not based on GDC's own best observed performance during Q1 but rather on the values observed for a sample of 15 US pipelines. The benchmarks were computed as a weighted average of these pipelines' opex/(MMCFPD*km) and capex/(MMCFPD*km). The weights were equal to each US pipelines' share of MMCFPD*KM within the US sample.

For GDC, the proposed efficiency analyses yielded two different X factors. We will denote X1 as the X factor consistent with GDC's intrinsic efficiency and X2 as the X factor consistent with relative efficiency. The final X factor was equal to the maximum value of X1, X2, or 0. That is, if X1 was greater than X2 and also greater than 0, then X was based on intrinsic efficiency. If X2 was greater than X1 and also greater than 0, then X was based on relative efficiency. If both X1 and X2 were less than 0, then X was equal to 0. As with the

[102]Costs at the outset of Q2 were determined by examining end of Q1 costs and adjusting these to eliminate costs that were either not documented, not systematic, not related to providing service, or otherwise deemed to be inappropriate or imprudent. This step is clearly important for determining overall Q2 revenues for GDC but only incidental to the calculation of the X factor itself.

gas distribution review, the final X factor applied to opex only, and was therefore multiplied by the share of opex in GDC's overall cost to set the final X factor.

The reviews for Energia Mayakan and SNG used identical benchmarking methods as that applied to GDC. The application was to opex only. The Commission was interested in perhaps using a variant of UK's "sliding scale mechanism" for setting allowed capex but decided against it because of a lack of benchmark capital expenditures data.

The most unique element of Mexican regulation is the way it has integrated assessments of intrinsic and relative efficiency into the computation of the X factor. In the case of gas transmission pipelines, these assessments have also brought together domestic and international benchmarking metrics. While the Mexican approach is intriguing in some respects, the unit cost benchmarks established in the pipeline plans are fairly crude and will not control for many differences in business conditions that can affect pipeline costs.

A.4.5 Victoria, Australia

Service quality regulation in Victoria has evolved considerably. In 1996, Victoria set minimum reliability standards for its distributors' short and long feeders. The regulator also published detailed reports on service quality for each distributor covering average performance and identifying problem feeders. While the minimum standards were generally considered to be too low, publication of service quality performance was designed to increase public awareness and put pressure on poor performing distributors to improve their performance.

The 2001–2005 determination introduced a formal service quality incentive scheme.[103] The service adjustment, S_t, that applied in year *t* for a particular distributor was calculated as a percentage according to the following formula:

$$S_t = \sum_{r,n} s_{r,n} (GAP_{t-2}^{r,n} - GAP_{t-3}^{r,n})$$

where:

[103] In addition, this determination led to compensation payments to customers affected by poor reliability. Distributors are required to make guaranteed service level payments of $80 to urban customers who experience more than 9 supply interruptions in a year and to rural customers who experience more than 15 interruptions. Urban and rural customers both also receive payments of $80 if their power is off for more than 12 hours at any one time.

R Refers to the following indicators:

Unplanned interruption frequency (SAIFI)

Unplanned interruption duration (CAIDI)

Planned minutes off supply (SAIDI)

N Refers to the following customer categories:

Central Business District (CBD)

Urban

Rural

$s_{r,n}$ is the incentive rate for indicator *r* and customer category *n.*

$GAP_{t-2}^{r,n}$ is the performance gap for indicator *r* and customer category *n* in calendar year *t-2*; that is the difference in performance between target and actual performance ($GAP_{t-2}^{r,n} = TAR_{t-2}^{r,n} - ACT_{t-2}^{r,n}$).

$TAR_{t-2}^{r,n}$ is the distributor's performance target for indicator *r* and customer category *n* in calendar year *t-2.*

$ACT_{t-2}^{r,n}$ is the distributor's actual performance for indicator *r* and customer category *n* in calendar year *t-2*, not including the impact of excluded events.

$GAP_{t-3}^{r,n}$ is the performance gap for indicator *r* and customer category *n* in calendar year *t-3*; that is the difference in performance between target and actual performance ($GAP_{t-3}^{r,n} = TAR_{t-3}^{r,n} - ACT_{t-3}^{r,n}$).

$TAR_{t-3}^{r,n}$ is the distributor's performance target for indicator *r* and customer category *n* in calendar year *t-3.*

$ACT_{t-3}^{r,n}$ is the distributor's actual performance for indicator *r* and customer category *n* in calendar year *t-3*, not including the impact of excluded events.

The incentive plan is symmetric, with the same reward rate for performance above the target as for penalties for performance below the target. The incentive rates are based on estimates of the distributors' marginal costs of improving reliability rather than estimates of customer valuation. These marginal cost estimates (in terms of A$ per kWh of unserved energy) were $3.96 for AGL, $7.05 for CitiPower, $6.47 for Eastern Energy (TXU), $4.55 for Powercor, and $4.67 for United Energy (Australian currency with each $A approximately equal to 1.33 US$). Total rewards and penalties were not capped under the plan, since the regulator believed that the Victorian data were sufficiently robust that caps were not necessary.

The basic elements of this scheme were retained in the 2005 Determination, but with three significant modifications. The first is that indicators for momentary interruptions (MAIFI) and telephone answering performance (percent of calls answered within 30 seconds) were added to the plan. Second, an “S bank” was added so that price adjustments for rewards/penalties were not implemented annually but could be smoothed over the term of the price controls. Finally, penalty/reward rates were changed from the (marginal) cost-based estimates to estimates of customer valuations. The latter rates were significantly greater.

References

Application of Southern California Edison to adopt a Performance Based Rate Making Mechanism Effective January 1, 1995, Alternate Order of Commissioners Fessler and Duque, July 21, 1996.

Bauer, P., A. Berger, G. Ferrier, and D. Humphrey (1998), "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods," *Journal of Economics and Business*, 50.

Breusch, T. and A. R. Pagan (1980), "The LaGrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47 pages 239-54.

Buse, A. (1982), "The Likelihood Ratio, Wald and LaGrange Multiplier Tests: An Expository Note," *The American Statistician*, 62, pages 153-7.

California Public Utility Commission Decision 97-07-054, *In the Matter of the Application of Southern California Gas Company to Adopt Performance Based Regulation for Base Rates*, July 16, 1997

Denny, M., M. Fuss and L. Waverman (1981), "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications". In *Productivity Measurement in Regulated Industries*, edited by T.G. Cowing and R.E. Stevenson. Academic Press.

Dhrymes, P. J. (1971), "Equivalence of Iterative Aitkin and Maximum Likelihood Estimators for a System of Regression Equations, *Australian Economic Papers*, 10, pages 20-4.

DTE, *Guidelines for Price Cap Regulation of the Dutch Electricity Sector in the Period from 2000 to 2003*.

Economic Insights (2009), *Total Factor Index Specification Issues*, Report to the Australian Energy Market Commission.

Federal Energy Regulatory Commission, Order 679, Docket No. RM06-4-000, July 20, 2006.

Griliches, Z. (1963), "Capital Stock in Investment Functions: Some Problems of Concept and Measurement", in *Measurement in Economics,* C. Christ et. al (editors), Stanford: Stanford University Press; reprinted in Z. Griliches (1988), *Technology, Education, and Productivity*, New York: Basil Blackwell.

Guilkey, et. al. (1983), "A Comparison of Three Flexible Functional Forms," *International Economic Review*, 24, pages 591-616.

Hall, R. and D. W. Jorgensen (1967), "Tax Policy and Investment Behavior", *American Economic Review*, 57, 391-414.

Handy-Whitman Index of Public Utility Construction Costs, (1993), Baltimore, Whitman, Requardt and Associates.

Hulten, C. and F.C. Wykoff. (1996). Issues in the measurement of economic depreciation: Introductory remarks. *Economic Inquiry 34*(1).

Hulten, C. and F.C. Wykoff (1981), "The Measurement of Economic Depreciation" in *Depreciation, Inflation, and the Taxation of Income From Capital*, C. Hulten ed., Washington, D.C., Urban Institute.

Hulten, C. R. "The Measurement of Capital," pp. 119-158 in *Fifty Years of Economic Measurement*, E.R. Berndt and J.D. Triplett (eds.), Chicago: the University of Chicago Press.

Joskow, P. and R. Schmalensee (1986), "Incentive Regulation for Electric Utilities"*, Yale Journal of Regulation*.

Kaufmann, L., *et al* (2010), *System Reliability Regulation: A Jurisdictional Survey*, Report Prepared for the Ontario Energy Board, May 2010.

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

Littlechild, S. (1983) *Regulation of British Telecommunications' Profitability: Report to the Secretary of State*.

Lowry, M.N., *et al* (2010), *Review of Distribution Revenue Decoupling Mechanisms*, Report Prepared for the Ontario Energy Board, March 2010.

Lowry, M.N., *et al* (2008), *Benchmarking the Costs of Ontario Power Distributors*, Report Prepared for the Ontario Energy Board, March 20, 2008.

Lowry, M.N., *et al* (2007), *Rate Adjustment Indexes for Ontario's Natural Gas Utilities*, Report Prepared for the Ontario Energy Board, November 20, 2007.

Lowry, M.N., *et al* (2006), *Second Generation Incentive Regulation for Ontario Power Distributors*, Report Prepared for the Ontario Energy Board, June 13, 2006.

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

Massachusetts DPU, February 24, 1995 Order in Docket 94-158, Investigation by the Department of Public Utilities on its own motion into the theory and implementation of Incentive Regulation for electric and gas companies under its jurisdiction.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, pages 69-85.

New Zealand Commerce Commission, *Decisions Paper: Initial Reset of the DPP*, November 2009.

Oberhofer, W. and Kmenta, J. (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models", *Econometrica*, 42, pages 579-90.

OECD Manual. (2001). Measuring Capital – Measurement of Capital Stocks, Consumption of Fixed Capital and Capital Services.

Office of Gas Supply (1995), *Price Control Review, British Gas' Transportation and Storage: A Consultation Document*, June 1995.

Office of Gas and Electricity Markets (2010), *Handbook for Implementing the RIIO Model*, October 2010.

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

Office of Gas and Electricity Markets (2004a), *Electricity Distribution Price Control Review: Initial Proposals*, June 2004.

Ontario Energy Board (2011b), Staff Report to the Board, *Electricity Distribution System Reliability Standards,* EB-2010-0249, March 31, 2011.

Ontario Energy Board (2011a), *Draft 2011-2014 Business Plan*, January 4, 2011.

Ontario Energy Board (2010). *Amendments to the Distribution System Code and Affiliate Relationships Code for Electricity Distributors and Transmitters*, Board File No EB-2009-411.

Ontario Energy Board (2009), *Staff Discussion Paper on the Regulatory Treatment of Infrastructure Investment for Ontario's Electricity Transmitters and Distributors*, EB-2009-0152.

Ontario Energy Board (2008c), *Supplemental Report of the Board on 3rd Generation Incentive Regulation for Ontario's Electricity Distributors*, September 17, 2008.

Ontario Energy Board (2008b), *Report of the Board on 3rd Generation Incentive Regulation for Ontario's Electricity Distributors*, July 14, 2008.

Ontario Energy Board (2008a), *Staff Discussion Paper on 3rd Generation Incentive Regulation for Ontario's Electricity Distributors*, February 28, 2008.

Ontario Energy Board (2006), *Report of the Board on Cost of Capital and 2nd Generation Incentive Regulation for Ontario's Electricity Distributors,* December 20, 2006.

Ontario Energy Board (2005), *Natural Gas Regulation in Ontario: A Renewed Policy Framework*, A Report on the Ontario Energy Board Natural Gas Forum, March 30, 2005.

Schwalbach, L., U. Grabhoff, and T. Mahmood, "The Dynamics of Corporate Profits," *European Economic Review*, October 1989, 1625-1639.

Schmidt, P. and R. Sickles (1984), "Production Frontiers and Panel Data", *Journal of Business and Economic Statistics.*

Pacific Economics Group, *et al* (2004), *Discussion Paper on Rate Regulation in Ontario*, Report Prepared for the Ontario Energy Board.

Stevenson, R. (1980), "Measuring Technological Bias", *American Economic Review*, vol. 70, 162-173.

Triplett, J. (1998). *A Dictionary of Usage for Capital Measurement Issues*, presented at the Second Meeting of the Canberra Group on Capital Stock Statistics (OECD).

U.S. Department of Commerce (1994), *Statistical Abstract of the United States*.

U.S. Department of Commerce, *Survey of Current Business*, various issues.

U.S. Department of Commerce, unpublished data on the stocks and service lives of the capital of Local Distribution Companies.

Varian, H. (1984), *Microeconomic Analysis*, Norton and Company.

Victorian Bushfires Royal Commission (2009), *Volume II: Fire Preparation, Response and Recovery.*

Waters, W.G. and J. Street (1998), "Monitoring the Performance of Government Trading Enterprises," *The Australian Economic Review*, Vol. 31, no. 4.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias," *Journal of the American Statistical Association*, 57, pages 348-68.

Y. Zhang and R. Bartels (1998), "The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand", *Journal of Productivity Analysis*, 9.