External Benchmarks, Benchmarking Methods, and Electricity Distribution Network Regulation: A Critical Evaluation

July 2001

Lawrence Kaufmann, Ph.D.
Partner, Pacific Economics Group

Margaret Beardow
Principal, Benchmark Economics

PACIFIC ECONOMICS GROUP
22 E. Mifflin St., #302
Madison, WI 53703
United States
1.608.257.1522

BENCHMARK ECONOMICS
21 Cogan Place
Lane Cove, NSW 2066
Australia
61.2.9428.2079

Table of Contents

PROJECT OVERVIEW

Pacific Economics Group (PEG) and Benchmark Economics have been retained by the National Electricity Distributors Forum to undertake research on the cost structure of power distributors and the role of benchmarking in utility regulation. The first two reports in this project were completed in April 2000. The first Report, *The Cost Structure of Power Distribution*, was a detailed examination of the cost structure of power distribution and related services. Attention was paid to the roles that economies of scale, economies of scope, and transaction costs play in determining optimal industry structure. The second, *A Survey of Performance-Based Regulations Plan and Benchmarking*, was a survey of alternative, "performance-based" regulatory mechanisms that have been approved in Australia, the United States, and the United Kingdom.

The second stage of our work began in early 2001 and yields three additional reports. Report three[1], *External Benchmarks, Benchmarking Methods, and Electricity Distribution Network Regulation*: *A Critical Evaluation* evaluates different methods for benchmarking the performance of power distributors. We compare four different benchmarking approaches: index-based methods, econometric cost functions, stochastic frontier analysis (SFA), and data envelope analysis (DEA). When applied to power distribution networks, we find that the two econometric methods (econometric cost functions and SFA) have notable advantages relative to DEA. The full and final version of this report follows below.

The fourth report, *Electricity Distribution Network Cost Structures: Cost Driver Analysis* will investigate the network production process to identify major cost drivers and the implications for benchmarking models and performance comparisons. Concluding this stage, the fifth report draws on the findings of the cost driver analysis to assess the appropriateness of indicators selected by regulators for performance comparisons.

[1] For ease of identification, the reports from each stage of the cost structure project are numbered according to their sequence in the overall project.

1. INTRODUCTION

In most Australian states, policymakers have made a conscious decision to regulate the prices of power distribution businesses (DBs) using CPI-X formulas rather than explicit rate of return regimes. This decision was motivated by two primary factors. One is that CPI-X regulation would create stronger incentives for DBs to operate efficiently. Second, CPI-X regulation was considered to be more compatible with competition, which was being introduced for various power-related services. Ultimately, it was believed that competition and stronger incentives to provide regulated services efficiently would increase benefits to all stakeholders.[2]

This policy direction has prompted a considerable debate about how CPI-X regulation should be implemented. In many applications, regulators have set the terms of CPI-X formulas so that each utility's revenues just equal its projected cost of service over the next regulatory period. These forward-looking revenue requirements include a target rate of return on capital. This approach to CPI-X regulation bears an undeniable relationship to rate of return regulation, and some economists believe that using cost of service methods to calibrate CPI-X formulas have caused these two regulatory systems to converge.[3] In cases where this is true, CPI-X regulation arguably represents little improvement over rate of return regulation.

An alternative approach is to calibrate CPI-X formulas using "external" performance measures. Such performance measures are often termed benchmarks. In a

[2] The problems with cost of service/rate of return regulation are well known and have been discussed extensively in Australia. The compatibility between CPI-X regulation of regulated services and effective competition in contestable services is less familiar but, again, has also been addressed many times in Australia. In the interests of brevity, we will not reprise these arguments. Interested readers can find discussions of these points in *Incentive Regulation, Benchmarking, and Utility Performance: CitiPower's Response to the Utility Regulators' Forum Discussion Paper*, March 2001, and *Updating Victoria's Price Controls: Analysis and Options*, L. Kaufmann and M.N. Lowry, September 1997.

[3] This has especially been noted with the practice of CPI-X regulation in the United Kingdom. For example, Dieter Helm of Oxford University writes "The British RPI-X regulatory system was designed to improve upon the perceived failures of the US (rate of return) approach. Over time, however, the two have increasingly converged"; D. Helm (1994), "British Utility Regulation: Theory, Practice, and Reform," *Oxford Review of Economic Policy*, vol. 10, no.3, 17-39. This view is also supported by Armstrong, M., S. Cowan, and J. Vickers (1994), *Regulatory Reform: Economic Analysis and British Experience*, MIT Press, Cambridge, MA and London.

regulatory context, a benchmark will be external to a utility if the utility's own actions cannot influence the value of the benchmark. External benchmarks would therefore not use the utility's own costs or expected costs to set the terms of CPI-X price controls.

Benchmarking is one specific means of establishing external benchmarks. In utility regulation, benchmarking involves comparing one or more utility performance measures to external performance standards. These external standards are computed using various benchmarking techniques. 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. Using benchmarking *per se* in CPI-X regulation implies that the terms of the CPI-X formula depend on direct comparisons between the utility's performance and selected external performance standards.

External benchmarks can 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, allow for greater flexibility in utility operations, and can reduce regulatory burdens.[4] Indeed, if CPI-X regulation is to represent a fundamental break with cost of service methods, it is necessary to use external benchmarks to set rates since the alternative, by definition, is the utility's own costs.

The role of benchmarking *per se* in promoting effective regulation is less certain. Thoughtful and rigorous benchmarking studies can, in principle, be used to establish objective performance standards that strengthen incentives and increase the potential benefits from utility services. But inappropriate benchmarking may not be beneficial to either customers or shareholders, particularly in the long run. "Bad" benchmarking applications can create some of the same problems and undesirable outcomes of rate of

[4] The strong performance incentives result from the fact that a company's prices are insensitive to its costs, so it has optimal incentives to reduce its unit cost of operations. Enhanced operational flexibility can take many forms, including having greater discretion in pricing and in using utility inputs in competitive markets (*e.g.* cost-shifting concerns are eliminated since prices do not depend directly on allocated costs). Regulatory burdens can decline since the need for regular, detailed examinations of company cost is avoided.

return regulation. Benchmarking is therefore a double-edged sword, and its impact depends on the understanding and care with which it is wielded in practice.[5]

In light of the possibilities for positive or negative outcomes, policymakers need a better understanding of external benchmarks and benchmarking. This is particularly urgent given the growing interest in benchmarking from regulators throughout the world.[6] We believe that for benchmarking to realize its potential in power distribution regulation, two issues deserve further examination.

The first is defining more precisely the role of external benchmarks and benchmarking in utility regulation. Regulators need direction on the role that external benchmarks and benchmarking can play in setting just and reasonable rates. This depends on setting reasonable benchmark standards and in understanding alternative paths that can be taken towards achieving those standards. In addition, principles should be established for the process of developing and using benchmarking results in regulatory reviews.

The second issue is the appropriateness of different benchmarking methods for DBs. Many techniques can be used to benchmark utility performance. While several authors examine the merits of benchmarking alternatives in a general way, few have addressed the appropriateness of these methods for energy delivery networks.[7] We

---

[5] Under some applications of CPI-X regulation, regulators ask utilities to submit projections of their forward-looking cost of service but then commission benchmarking studies as a kind of check on the reasonableness of those projections. While this is a somewhat hybrid approach, where benchmarking evidence is used to supplement fundamentally cost-based regulation, in practice it resembles cost of service regulation more than external regulation. We discuss this point further in Section 3.4.

[6] This is apparent in the recent survey article by T. Jamash and M. Pollitt, “Benchmarking and Regulation of Electricity Transmission and Distribution Utilities: Lessons from International Experience,” Working Paper, December 2000, University of Cambridge. This article surveyed electricity regulators in 21 nations, and at least 12 of these countries have already used benchmarking in electricity regulation. Benchmarking is either planned or under consideration in every one of the other countries. Seventeen of the 21 nations were members of the Organization for Economic Cooperation and Development (OECD), while four were non-OECD members. A recent report by Britain’s airport regulator also examined the UK’s past application of benchmarking and found “the use of benchmarking to inform setting price caps is well established in economic regulation;” see the Civil Aviation Authority, *The Use of Benchmarking In the Airport Reviews: Consultation Paper*, December 2000, p. 2.

[7] For examples of general discussions of benchmarking techniques, see M. Pollitt, *Ownership and Performance in Electric Utilities: The International Evidence on Privatization and Efficiency*, Oxford University Press, 1995; London Economics, *Efficiency and Benchmarking Study of the NSW Distribution Businesses*, Report to the Independent Pricing and Regulatory Tribunal Panel (IPART), February 1999;

believe this issue is crucial. As discussed in our report *The Cost Structure of Power Distribution*, power distribution differs from most other businesses in several respects. Benchmarking methods should be firmly rooted in an understanding of the industry's cost structure and must appropriately reflect its unique conditions.[8]

This report is intended to address these issues. We examine the relationship between external benchmarks, benchmarking and effective regulation. We then develop some criteria that should inform the application of benchmarking methods in regulation.

We also evaluate the main techniques for computing external benchmarks. The alternatives considered are index-based methods, econometric cost functions, stochastic frontier analysis (SFA), and data envelope analysis (DEA). In analyzing the merits of each technique, we emphasize its appropriateness for DBs.

One of our main conclusions is that, when benchmarking DB performance, econometric techniques have significant advantages over DEA. This is not meant to disparage DEA, which is a well-established method and has been applied by researchers numerous times. However, we believe that DEA is not well suited for power distribution networks. Ironically, some of the main attractions of DEA (e.g. reliance on physical rather than financial capital measures) are the factors that limit its usefulness for power distribution. The power distribution industry also has unique features that may be difficult to capture in DEA models.

This result has important policy implications. DEA is apparently the most common method of benchmarking power distribution, particularly in studies commissioned by regulators.[9] While this inclination may be understandable, if our analysis is correct, it will ultimately prove to be unfruitful. This could be unfortunate, for the specific problems of using DEA to benchmark DB performance may tend to undermine the role of external benchmarks more generally in utility regulation.

IPART, *Benchmarking the Efficiency of Australian Gas Distributors*, 2000; and the Civil Aviation Authority, *The Use of Benchmarking In the Airport Reviews: Consultation Paper*, December 2000.

[8] Of course, for any given benchmarking method, the actual application of this method (*e.g.* definitions of outputs, choices for business conditions, etc.) should also be firmly grounded in an understanding of industry structure. We deal with these issues in our second report.

[9] For example, see the Tamash and Pollitt survey, *op cit*.

This report is organized as follows. Section 2 discusses the role of external benchmarks and benchmarking in economic regulation. Section 3 describes the main alternatives for establishing external benchmarks. Section 4 evaluates these alternatives. Section 5 discusses directions for further research.

2. The Role of External Benchmarks and Benchmarking in Economic Regulation

This chapter will address the role that external benchmarks and benchmarking can play in promoting effective utility regulation. We begin by establishing some criteria for "just and reasonable" rates and the implications for the role of external benchmarks and benchmarking. Next, we discuss how the application of benchmarking can play either a positive or negative role in achieving effective regulatory outcomes. We then discuss some guideposts for the process of undertaking benchmarking and integrating it into rate regulation.

2.1 "Just and Reasonable" Rates, Benefit-Sharing, and Utility Returns

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.[10] For these and related reasons, a "competitive market paradigm" is useful for establishing effective regulatory arrangements. Below we consider how competitive markets operate and the implications for economic 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

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

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.[11]

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 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.

Another implication of the competitive market paradigm bears a direct relationship to the calibration of CPI-X formulas. As noted above, in the long run, competitive market prices grow at the same rate as the industry trend in unit cost. Industry unit cost trends can be decomposed into the trend in the industry's input prices minus the trend in industry total factor productivity (TFP). Thus if the selected inflation measure is approximately equal to the growth in the industry's input prices, the first step in implementing the competitive market paradigm is to calibrate the X factor using the industry's long-run TFP trend.[12]

---

[11] 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"; *op cit*, p. 11.

[12] We have detailed the algebra that decomposes industry unit cost trends into industry input price and industry TFP trends in various reports presented in Australia. This algebra also shows that if economy-wide inflation measures such as the CPI or GDP-PI are used, the X factor can also include an inflation

However, some will argue that utilities are likely to display greater cost inefficiency on average than firms in competitive markets. This is because utilities have historically not operated under the competitive market pressures that naturally create incentives to operate efficiently. In addition, traditional, cost of service regulation has not promoted efficient utility behavior. Therefore if economic regulation is designed to strengthen performance incentives, it should encourage utilities to increase their efficiency above historical norms in the industry. It is also reasonable for these performance gains to be shared with customers, since CPI-X regulation is designed to produce "win-win" outcomes for customers and shareholders.

There are many ways to share performance gains with customers in ways that enable shareholder returns to remain commensurate with performance. We discuss three options below.[13] These alternatives are arrayed in descending order of the risks that they create.

The riskiest 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.

---

differential which is equal to the differences in input price inflation for the industry and the economy. We explain this term in detail in Kaufmann and Lowry (1997), *op cit*.

[13] Other benefit-sharing options, such as a rolling X factor, can also be considered and may be particularly valuable when there is limited information on the industry's long-run TFP trend.

While this approach to benefit-sharing 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 and therefore puts utilities at risk if these gains do not materialize.

A more moderate approach is for X to include a "stretch factor." The X factor would then have two components: the industry TFP trend, which reflects long-run industry performance; and a stretch factor that shares short-run performance gains with customers. A reasonable stretch factor may be between 0.5% and 1.0%.[14]

Benchmarking can be used to 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. 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.

The least risky option is for the CPI-X plan to include an earnings-sharing mechanism (ESM). ESMs have well-known drawbacks in terms of reducing efficiency incentives and introducing cost-shifting concerns. They also make allowed prices depend in part on the company's performance rather than strictly external factors. But in spite of these disadvantages, ESMs have a number of benefits. They share actual efficiency gains as they occur. They also tend to offset uncertainties regarding industry input price trends (such as changes in the cost of capital) and the calibration of CPI-X formulas. These

[14] In North American CPI-X plans, where this approach has been applied, average stretch factors for energy utilities typically fall at the lower end of this range.

factors can lead to unexpected changes in utility returns.   Accordingly, ESMs tend to keep utility earnings within politically "acceptable" bounds, which can bolster the long-term credibility of the regulatory regime. Provided the same earnings-sharing formula is applied to all firms in the industry, ESMs also enable a utility's relative returns to be commensurate with its performance in the industry.

These options show that many paths can be taken towards implementing the competitive market paradigm. Regulators should be aware of the diversity of available approaches and how they differ in terms of risk and benchmarking emphasis. 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.[15] In all cases, however, several factors must be kept in mind in 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.[16] 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.

---

[15] An example of an institutional constraint may be that ESMs are not compatible with some DBs' current legal frameworks. For example, Victoria's laws regarding reviews of price controls rule out the use of rate of return regulation, which is arguably a component of ESMs. Data issues will likely affect the feasibility and quality of benchmarking studies.

[16] These are known as sunk assets, defined as assets whose value is significantly lower in the next best use. Many power distribution assets are literally "sunk" into a particular location and thus have far less value outside their particular location and dedicated uses. By way of contrast, consider the airline industry, which is similarly capital intensive but whose primary assets (airplanes) can be readily resold to competing firms.

This can perhaps be clarified with an example. Suppose a DB can achieve best-observed practice in the industry by installing a new piece of capital equipment, but the DB's current capital equipment has many more years of operation left. Since the secondary market value of this equipment is well below its economic value to the DB, there are significant costs of changing operations and discarding this capital. If these costs are greater than the (discounted present value of the) benefits associated with "best practice" assets, it will not be cost effective to adapt these methods immediately. It will only be cost effective to make such capital adjustments later on, after the capital equipment has been further depreciated. The installation of new capital equipment would take place sooner if the existing asset could be readily re-sold, since this would effectively reduce the marginal cost of the investment.

It should also be remembered that in competitive markets, firms with superior performance earn above average returns. This is true even in the long run.[17] 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

[17] 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.

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.

2.2 Potential Costs and Benefits of Benchmarking

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.

Moreover, as some observers have noted, if incentive regulation is to be something other than cost of service regulation with a regulatory lag, then external benchmarks must play a role.[18] If external benchmarks either do not exist or regulators do not have confidence in them, they will invariably focus on whether company returns are "reasonable." This will cause CPI-X regulation to converge to rate of return regulation, as it has in some applications. As one method for generating external benchmarks, benchmarking therefore has the potential to be part of a regulatory regime that is truly more objective.

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

[18] For example, see T. Coelli, "Some Scattered Thoughts on Performance Measurement for Regulation of a Natural Monopoly Network Industry," Working Paper, August 2000, University of New South Wales.

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. In Chapter 4 we will discuss criteria that can be used to evaluate the quality of alternative benchmarking methods for power distribution and the specific results that emerge from benchmarking studies. Below we deal with the more limited, but still important issue of the process of implementing benchmarking in regulation.

2.3 Applying Benchmarking in Regulation

Many of the criteria for how to apply benchmarking in regulation are relevant to any regulatory system, so we hope that our points in this section are straightforward and not controversial. However, one should be especially cognizant of these factors when applying benchmarking in regulation. One reason is that benchmarking is a relatively new approach. It is therefore important to build consensus and confidence among various stakeholders, and this is more likely if the regulatory process obeys certain criteria. Relatedly, benchmarking often involves sophisticated empirical techniques that are not widely understood. Understanding and agreement on appropriate regulatory methods are again more likely if the regulatory process has certain properties.

We begin by examining some criteria recently put forward by the Australian Competition and Consumer Commission (ACCC) on "best practice regulation."[19] The ACCC criteria are that the regulatory process must be consultative, predictable, consistent, accountable and transparent. With respect to the specific point of integrating benchmarking into regulation, we agree with these criteria and would add only minor elaborations.

A consultative regulatory process would be one where due process is respected and all interested parties can provide input. This is currently the case in Australia. One important aspect of an accountable regulatory process is that regulatory decisions be subject to judicial review. This is a critical safeguard to ensure that regulators are exercising their authority in accordance with the underlying legislation. This type of accountability exists some States in Australia.

We believe a transparent regulatory process is one where the basis for all regulatory decisions is clearly articulated and supported. Decisions cannot be based on assertions that have no factual basis. With respect to benchmarking, this implies that the empirical basis for every benchmark that is established must be explicit and clear. This "empirical basis" applies to both the benchmarking methods and the specific benchmarking results used by regulators.

Regulatory transparency is also related to verifiability. Results should be subject to verification, at least by certain designated parties and subject to defined rules. This criterion is likely to rule out using confidential datasets as the basis for the benchmarks in regulation. Benchmarks based on confidential data are not transparent, since the data themselves are not visible. Using confidential data also undermines the objective of a consultative regulatory process, for parties cannot respond knowledgeably to benchmarking studies that use non-verifiable data.[20] Confidential data also clearly reduce the accountability of the regulatory process.

---

[19] Presentation by Joe Dimasi, Executive General Manager Regulatory Affairs at the ACCC, *The Role of Benchmarks: Australian Experience and Future Directions*, March 13, 2001, Sydney.

[20] In some benchmarking studies commissioned by Australian regulators, reviews by companies and outside parties revealed significant data errors; this was particularly true in the IPART study for power distribution.

It should also be noted that we believe the principles of transparency and verifiability apply to all parties in a proceeding. If outside critiques of benchmarking studies are to be credible and given weight by regulators, they must be held to the same standards as regulator-commissioned studies. Regulators can have little confidence in companies' benchmarking studies if those studies utilize confidential data or do not clearly articulate and support their results.

The predictability and consistency of benchmarking depend in large part on the details of the benchmarking studies themselves. We will discuss this issue in detail in Chapter 4. However, it is worth noting that predictability and consistency can also be promoted at the outset of the regulatory process by having a clear understanding of benchmarking alternatives. An "up front" investment in evaluating the merits of benchmarking options and their appropriateness for the industry in question can make any subsequent application of benchmarking more effective. For example, this can reduce the "trial and error" aspect of undertaking (perhaps several) benchmarking studies and then attempting to select some appropriate benchmark measure. The lack of pre-established selection criteria or understanding of the sensibility of benchmarking methods and results increases the unpredictability and inconsistency of such a process. This is not to suggest that any regulatory process can ever be made completely predictable. But greater understanding at the outset of the process is likely to lead to better choices for benchmarking methods and impose discipline on procedures for evaluating and implementing the results of benchmarking models.

3. External Benchmark Alternatives

This chapter will briefly describe four approaches towards generating external benchmarks for utility performance. These approaches are 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.[21] Moreover, as should become apparent in our discussions, the last three of these methods can be properly described as benchmarking *per se*, as this term was defined in Chapter 1. Our description of these methods is intentionally non-technical.[22]

3.1 Index-Based External Benchmarks

This approach is sometimes not mentioned in discussions of benchmarking techniques, particularly in articles describing the measurement of production frontiers. However, index-based approaches are an important method for calculating external benchmarks. This method has also been applied numerous times in utility regulation.

There are two main types of indexes. A total factor productivity (TFP) index is a comprehensive 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 indexes are designed to compare the overall efficiency with which enterprises use capital, labor, and other production inputs to produce goods and services.

---

[21] 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.

[22] Greater technical detail on these benchmarking alternatives has been presented in other Australian benchmarking studies. For example, see the *Technical Annex: Efficiency and Benchmarking Study of the NSW Distribution Businesses*, London Economics, February 1999.

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.

Each TFP index is the ratio of an output quantity index to an input quantity index. An output quantity index is a summary measure of the amounts of goods and services produced. An input quantity index aggregates all of the quantities of inputs used in production. A TFP level index will increase when input quantity levels decline relative to output quantity levels. For example, suppose that utility A produces the same amount of output as utility B with 10% less input. The TFP of utility A is then 11% above that of utility B.

Aggregating inputs and outputs involves choices for index forms. Economists have identified certain index forms as being particularly attractive for productivity measurement. These are sometimes known as “superlative” indexes.[23] The two superlative index forms are the Fisher Ideal and Tornqvist indexes. Nearly all TFP studies use one of these indexes.[24] These indexes use data on cost shares to weight input quantities when they are aggregated into a comprehensive input quantity index. Data on revenue shares are used to weight output quantities when they are aggregated into a comprehensive output quantity index.

Some data on the revenues associated with DB outputs may not be readily available. For example, the revenues stemming from customer numbers (*e.g.* the per-customer charge) and kWh distributed are often not publicly reported by North American utilities. When such data are not available, information from econometric cost functions can be used to generate appropriate output weights.[25]

---

[23] W.E. Diewert (1976), “Exact and Superlative Index Numbers,” *Journal of Econometrics*, 115-145.

[24] Diewert and Fox have recently emphasized the value of the Fisher Ideal index in regulatory applications; see W. E. Diewert and K. Fox (2000), “Incentive Indexes for Regulated Industries,” *Journal of Regulatory Economics*, 17:1, 5-24.

[25] In such cases, it can be shown that an appropriate output weight for each output is equal to the estimated elasticity of cost with respect to that output divided by the sum of the cost elasticities for all outputs.

3.2 Econometric Benchmarking

3.2.1 Econometric Cost functions

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 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.[26] 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, DBs confront other operating conditions due to their special circumstances. Unlike firms in competitive industries, power distributors 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 DB 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

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

customers. Even more importantly, different types of customers have different levels and temporal patterns of demand and different load factors.[27]

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 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.

3.2.2 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.[28] The actual

---

[27] These differences can affect the incremental and average costs associated with a given set of utility assets. As we noted in our report *The Cost Structure of Power Distribution*, the relationship between average costs, incremental costs, and the size of the market can have important implications for scale and scope economies and the minimum efficient scale of operations associated with providing a given utility service.

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 Chapter 4.

3.3 Data Envelope Analysis

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

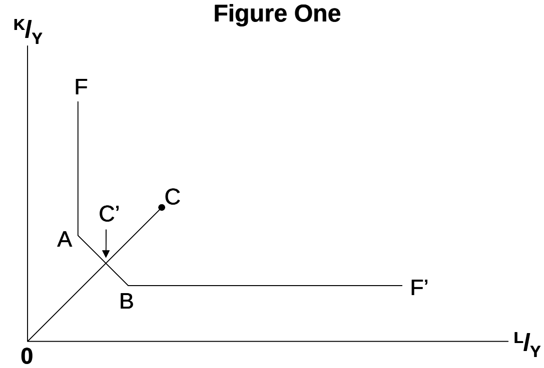
In a basic input-oriented DEA model, the relative efficiency of a firm is determined by assigning weights to firm inputs and outputs such that the ratio of aggregated outputs to aggregated inputs is maximized. This linear programming problem is subject to the constraint that the efficiency score cannot 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.

---

[28] Alternatively, SFA can be focused on estimating maximum production frontiers.

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 various components.[29]

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

[29] 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.

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.

It is important to point out that DEA can be conducted using only physical input and output measures. It is not necessary to compute the financial costs or input prices associated with various inputs.[30] This is sometimes considered to be a significant advantage, for these measures are often not readily available and can require significant data to calculate. This is particularly true for capital inputs, which account for the largest share of power distribution cost.

3.4 External Benchmarks and Benchmarking in Energy Distribution Regulation

It may be useful to review some of the previous applications of external benchmarks and benchmarking in energy distribution regulation. This discussion is not intended to be comprehensive, particularly with regard to the most recent applications. A more detailed survey of the most recent uses of benchmarking *per se* in energy regulation is contained in Jamash and Pollitt (2000).[31]

As noted in our previous research, North American regulators routinely calibrate CPI-X plans on the basis of industry unit cost trends rather than a utility's own costs.[32] TFP trends are therefore important components of the index-based regulation plans that have been approved in North America. Most of these plans are in the telecom sector. However, TFP evidence has been considered in approved CPI-X plans for the power distribution services of San Diego Gas and Electric, Southern California Edison and

---

[30] However, input prices are required to calculate allocative efficiency, for this measures the extent to which the input mix is optimal given the relative input prices facing the firm.

[31] See Jamash and Pollitt, *op cit*. However, this article does not discuss index-based benchmarks in detail, and is in fact deficient in its discussion of external benchmarks and benchmarking in North America.

[32] See Kaufmann and Lowry, *op cit*, for a discussion of these plans.

Central Maine Power and the gas distribution services of San Diego Gas and Electric, Boston Gas, and Southern California Gas. Prior to the restructuring of the electric power industry, TFP evidence was also considered in indexing plans for Pacificorp-California and Central Maine Power.

Econometric cost benchmarking has been employed in several jurisdictions. In North America, Boston Gas, Southern California Gas, Louisville Gas and Electric, and Kentucky Utilities have all offered econometric evidence on their comprehensive cost performance. In Australia, the Victorian DBs presented similar econometric evidence in the review of their price controls.

Several regulators have undertaken econometric benchmarking of operation and maintenance (O&M) costs. The Queensland Competition Authority relied on O&M benchmarking studies that were derived from a model of comprehensive DB cost. In the United Kingdom, more *ad hoc* models were used to benchmark the O&M costs of the Regional Electricity Companies (RECs).

The Independent Pricing and Regulatory Tribunal (IPART) in New South Wales has used both SFA and DEA benchmarking when reviewing price controls for the state's power and gas distributors. These methods proved to be tremendously controversial and were ultimately given little weight in the reviews. According to several analyses, one reason for the unreliability of these studies is that the benchmarking methods showed little understanding of, and were not appropriate for, the energy distribution industries to which they were applied. DEA studies have also been undertaken for the regulation of power distributors in Norway and the Netherlands. Evidently, DEA is also under consideration by several other regulators in Europe and Latin America.[33]

There are other purported applications of benchmarking in Australian regulation, but it is difficult either to identify the benchmarking methods or results in the Regulatory decisions. A salient example is the recent update of power distribution price controls in Victoria. In this review, every element of each DB's cost is referred to as a "benchmark," but it is impossible to ascertain the empirical bases for these benchmarks from the Determination. However, it is clear that each line item of DB cost has been

[33] See Jamash and Pollitt, *op cit.*

carefully scrutinized and often adjusted on the basis of either staff or outside party judgments. Apparently, this process of scrutinizing company costs and submitting them to outside analysis has been viewed as a form of "benchmarking."

These methods are more reminiscent of cost of service regulation than Australian observers may realize. Regulators have long recognized that a pure "cost plus" approach towards rate of return regulation creates few incentives for managers to contain unit costs. As a result, regulators never take a utility's submitted costs at face value.[34] "Rate cases" that establish cost-based rates submit every utility cost item to extensive review and analysis by Commission staff and outside consultants. These efforts attempt to determine that company costs that form the basis for prices are "reasonable." Such determinations often involve judgments about the efficiency of firm operations.[35] At times, this leads to explicit disallowances of expenses that are not "prudently incurred" or capital investments that are not "used and useful." Nearly always, rate cases lead to smaller rate increases (or larger rate decreases) than the company requested. Although the terminology may differ, Victoria appears to have employed a nearly identical process that focuses on detailed examinations of company-specific costs, rather than external benchmarks. We therefore would not categorize this as an application of either external benchmarks or benchmarking *per se*.

---

[34] A good description of rate of return regulation regulatory procedures can be found in P. Joskow and R. Schmalensee (1986), "Incentive Regulation for Electric Utilities," *Yale Journal of Regulation*, volume 4, no.1, 1-50.

[35] Indeed, rate cases would not be so burdensome and contentious unless, at some level, regulators believed that their duty to implement "just and reasonable" rates involved at least implicit evaluations of utility operations. Detailed cost examinations and prudence reviews can be viewed as heavy-handed attempts to promote efficiency and customer benefit using "sticks" rather than "carrots." In contrast, the philosophy underlying incentive regulation is that customers ultimately benefit more from well-designed rules that motivate utilities to pursue efficiency gains. Such rules utilize the carrot of potentially higher returns.

4. Evaluation of Benchmarking Alternatives

This section evaluates the benchmarking options described in Chapter 3. We begin by presenting criteria for evaluating benchmarking methods as well as the robustness and reasonability of benchmarking results. These latter criteria can only be applied to the actual outcomes of benchmarking studies, while the former can be used to analyze the appropriateness of different methods of benchmarking DB performance. In light of these criteria, we assess the advantages and disadvantages of the four main benchmarking methods discussed in Chapter 3. We then evaluate the reasonableness of some actual benchmarking results using evidence from financial markets. Although not based directly on power distribution studies, this evidence is still valuable in terms of judging the reasonableness of results derived from DEA and econometric methods.

4.1 Criteria for Evaluating Benchmarking

4.1.1 Evaluating Benchmarking Methods

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 DB cost but are beyond company control.[36] 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 DBs 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 DB 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.

4.1.2 Evaluating Benchmarking Results

A recent paper by Bauer et. al has proposed some criteria for evaluating the robustness and reasonableness of benchmarking results.[37] 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

[36] For example, see *The Cost Structure of Power Distribution* by L. Kaufmann and M.N. Lowry.

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
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

---

[37] 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.

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."[38]

However, Bauer et al's last two criteria may have less relevance when applied to benchmarking results for DBs. A DB's financial performance can be affected as 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.[39] 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 DBs.

4.2 Index Based Benchmarks

4.2.1 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.

---

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

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

Another important disadvantage is that partial measures do not control for differences in most business conditions, including the DB's own choices for other inputs.[40] 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.

However, in spite of their well-known problems, partial productivity measures remain far more common in Australia. This tendency is inherently tied to regulators' preference for the "building block" approach to regulation. This approach generally eschews comprehensive performance measures and treats O&M and capital spending separately. In an attempt to identify "efficient" levels for these inputs, this inevitably leads to benchmarking O&M and capital separately in spite of the difficulties of doing so.

These difficulties have recently been manifested in Victorian regulation. The Office of the Regulator General (ORG) employed a building block approach that included an efficiency carry-over where the differences between actual and projected costs were phased out over the next price control period. Separate efficiency carry-overs were specified for O&M and capital expenditures. Under this approach, the amounts for carry-overs are inherently sensitive to how O&M and capital expenditures are defined. Victorian DBs employed different approaches towards capitalizing and expensing costs and different cost allocation policies, and this inevitably affected the magnitudes for the carry-overs. Several DBs objected to the ORG's methods, and an Appeal Panel agreed with many of their concerns. Largely in response, the ORG was required to issue an amended version of its Price Determination.

We believe there is a natural link between the philosophical approach taken towards regulation and the performance measures that are employed. The building block approach is more naturally associated with partial measures. External regulation that creates stronger performance incentives is inevitably linked to comprehensive measures.

[40] This point has recently been highlighted in a Discussion Paper for the Utility Regulator's Forum, R. Albon, *Incentive Regulation, Benchmarking and Utility Performance*.

We believe this linkage is critical and should be recognized by regulators. It also follows that, since building block regulation is naturally linked to performance measures which are themselves often unreliable, this supports the move to external regulation.

4.2.2 Total Factor Productivity (TFP)

4.2.2.1 Advantages of 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 our previous research, industry TFP trends have been used to calibrate the X factors in North American indexing plans for over a decade.[41] TFP measures have been the basis for X factors in plans for North American railroads, telecom companies, and gas and electric utilities.

There has also been considerable consensus about the magnitudes of TFP trends in some of these proceedings. For example, in appraising an indexing plan for Southern California Gas (SoCalGas), the California Public Utilities Commission (CPUC) said that the industry TFP growth trend proposed by the company "elicited little criticism from the

---

[41] For example, see L. Kaufmann and M.N. Lowry, *Updating Price Controls in Victoria: Analysis and Options*, June 1997.

parties."[42] We believe that this statement compares favorably with the discussion of benchmarking in most Australian proceedings and demonstrates that TFP can be widely accepted in practice.

4.2.2.2 Disadvantages of TFP

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.[43]

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.

---

[42] 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.

[43] The Utility Regulators' Forum Discussion Paper mentioned above (R. Albon, *op cit*.) cites several other potential problems with TFP, but these are without merit. For an evaluation of these points, see *CitiPower's Response to the Utility Regulators Forum Discussion Paper*, March 2001, pp. 6-7.

4.3 Econometric Cost Functions

4.3.1 Advantage 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.

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

Here $C_{DB,t}$ refers to the DB's actual cost in period t, while $\hat{C}_{DB,t}$ is the estimated DB 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.[44] 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 DB's actual cost differs significantly from the model's prediction.

Another advantage of this approach is that it can be sensitive to a wide range of DB business conditions. Econometric benchmarking does not require identification of a suitable peer group. Indeed, variation in sampled business conditions is actually

---

[44] 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.

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 DB.

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 2, 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.

4.3.2 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.[45] 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

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

very well be larger as the 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 DB benchmarking studies. We discuss this further in section 4.5.2.

4.4 Stochastic Frontier Analysis

4.4.1 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 DBs 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.

4.4.2 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.[46] 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.[47] 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.[48] 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. This may be an advantage in Australian regulation, where the focus has been on partial rather than comprehensive benchmarking. Since econometric cost models can generate predictions for specific cost categories as well as total cost, they may be more compatible

---

[46] 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.

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

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

with current Australian practice. Econometric cost models may thereby facilitate the transition to more comprehensive benchmarking more easily than SFA.[49]

4.5 Data Envelope Analysis

4.5.1 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

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 DBs using DEA rather than SFA or other econometric methods. There is currently only a single year's worth of data for the 20 DBs 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."[50]

4.5.2 Disadvantages with DEA

When applied to DBs, 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 power distribution. We divide the disadvantages with DEA into

---

[49] A related reason is that econometric cost models may be better able to handle the complex inter-relationships between cost and service quality than SFA. These interrelationships may involve the specification and simultaneous estimation of multiple equation systems. This is a frontier area for research, which we discuss further in Section 4.5.2 and Chapter Five.

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

four categories: data requirements and related problems; the ability to deal with uncertainty; assumptions regarding the production process; and problems with controlling for DBs' business conditions. Some of these issues are interrelated so the problems will overlap somewhat between categories.

4.5.2.1 Data Measures and Requirements

Capital accounts for the dominant share of DB costs, so its treatment in benchmarking models is critical. DEA typically uses physical rather than financial capital measures as inputs in DB benchmarking studies. Examples include MVA of transformer capacity and km of distribution line. We believe that this approach is problematic in several respects.

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.[51] Similarly, DBs 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 DBs' capital stock, but they can only be measured in financial terms. It is therefore not possible to measure the scope of DB 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

[51] 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.

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. DEA usually deals with physical quantities of inputs, so differences in relative amounts of inputs can affect DEA results. In general, US DBs will have more MVA of transformer capacity and fewer km of line, while Australian DBs will have more km of line and fewer MVA of transformer capacity.[52] Different input proportions can distort which firms are selected as peers, since this choice depends on relative input proportions among sampled companies. Comparing an Australian DB to an inappropriate peer leads directly to inappropriate benchmarking results.[53] 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 between US and Australian DBs, 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

[52] 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 DBs have a greater reticulated low-voltage network compared with US firms.

[53] Put another way, differences in system design between US and Australian DBs 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.

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.[54] 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 DB *output*. This output can be proxied by the total km of distribution line, since this is related to the physical location of customers in the DB'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 DBs' 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 DBs using only physical capital measures.

4.5.2.2 Data Issues and Uncertainty

DEA is not a statistical method, so it much less conducive to dealing with uncertainties regarding benchmarking measures. It is generally 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

[54] 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.

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.[55] Again, this demonstrates that DEA benchmark measures can be affected by the performance of a single firm.

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. This issue is also likely to be relevant in Australia for some time, since the small number of Australian DBs and short data series suggest that benchmarking is likely to rely on international samples for the foreseeable future.

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

[55] 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.

very imprecise.[56] 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.

4.5.2.3 Restrictions on Production Process

While DEA does not directly restrict the relationship between DB 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

[56] 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.

regressions.[57] 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 cost functions, which are disciplined by economic theory and place a minimum of restrictions on the underlying production relation.

4.5.2.4 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 DB 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 DBs. There are clear cost-quality tradeoffs in power distribution.[58] DB managers make inter-related decisions about optimizing cost and reliability. This

---

[57] 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.

[58] Cost-quality tradeoffs, and their relationship to differences in business conditions, were discussed in *The Cost Structure of Power Distribution*, *op cit*, pp. 21-22.

optimization process is influenced by other business conditions that the utility faces. In other words, the cost-quality tradeoff confronting a DB will vary depending on its other business conditions. Rural DBs, 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 DB has a low DEA efficiency score relative to urban DBs because it requires more inputs to provide the same level of output. If service quality is added as an output, the rural DB's performance is likely to look even worse. The DEA model will now show the rural DB is providing fewer units of the quality output relative to urban DBs. 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 DB 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 DBs 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 DB 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 DB benchmarking.

4.6 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. There are few examples of utility benchmarking that can be judged according to the Bauer et. al criteria listed at the outset of this chapter. However, these authors do apply the six criteria they establish to a common dataset. Surprisingly, few past researchers have done this.[59]

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.[60] 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 rankings between the econometric approaches but a weak correlation between

[59] 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.*

[60] The two other econometric methods are "thick frontier analysis" and "distribution free analysis," these are similar to SFA in that they are both employ econometric techniques to estimate production frontiers. However, as previously discussed, they are not used extensively in utility benchmarking and so have not been discussed explicitly in this paper.

the DEA and econometric approaches.[61] 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."[62] 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."[63]

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

---

[61] 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.

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

[63] *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.

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:

> 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.[64]

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."[65]

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 DBs. 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 DBs.

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

[65] Op cit, p. 109.

5. Conclusion and Directions for Further Research

5.1 Conclusion

This report had two main aims. The first was to examine the role of external benchmarks and benchmarking in utility regulation. The second was to evaluate alternative methods for benchmarking performance.

On the first issue, we believe that external benchmarks can play either a positive or negative role in regulation depending on how they are applied. In principle, external benchmarks and benchmarking can help set objective performance standards that create stronger incentives. Stronger incentives can lead to greater long-run benefits for customers and shareholders. However, inappropriate benchmarking can create unrealistic performance standards. Such standards may lead to prices that do not fully compensate utilities for their costs. This can be damaging to both customers and shareholders, particularly in the long run, as unrealistically low prices lead to less investment in utility industries.

Criteria should be specified that support appropriate benchmarking applications. We believe a competitive market paradigm should be used for integrating external benchmarks and benchmarking in regulation. While there is some flexibility in how this paradigm can be applied, it should allow superior performers to earn superior returns, permit gradual adjustments to performance targets, and recognize that considerable uncertainty exists about "frontier" performance levels in any industry. This latter point implies a more cautious stance regarding the adoption of performance targets.

The process of applying benchmarking in regulation should also obey certain properties. We believe this process should be consultative, predictable, consistent over time and across companies, accountable, and transparent. Satisfying some of these criteria will depend on the details of the benchmarking studies themselves.

On the second issue, we began by specifying criteria that can be used to evaluate both benchmarking methods and the results of specific benchmarking studies. Ideally, benchmarking methods should be consistent with economic theory, impose minimal assumptions or restrictions on the underlying production process, be able to capture the impact of business condition variables on company performance, have low data requirements, and recognize the uncertainty stemming from random and/or unmeasured factors on utility performance. Our criteria for evaluating the reasonableness of benchmarking results were proposed in a recent paper by Baer et al.

We examined the merits of four different benchmarking options according to these criteria. Overall, we support the use of comprehensive as opposed to partial benchmarks. Partial benchmarks can lead to incomplete and distorted performance assessments. However, some comprehensive measures can be decomposed into a consistent set of partial measures. This can be done to provide a greater range of information, but the partial measures should be interpreted in the broader context.

Index-based methods can be valuable means of setting external benchmarks. Most importantly, the competitive market paradigm establishes a direct link between price trends and industry TFP trends. This suggests that industry TFP indexes can play an important role in promoting effective utility regulation. At the same time, index-based methods have limitations, including an inability to quantify the uncertainty associated with performance assessments or reflect a full range of business conditions that can affect performance. Index-based methods may therefore not be sufficient for evaluating differences in performance levels between utilities.

Some of these problems can be dealt with using econometric methods. Both econometric cost functions and stochastic frontier analysis (SFA) can be used to measure company performance in ways that control for many factors beyond company control and reflect the statistical precision of performance measures. These features are crucial if benchmarking is to be both practical and valuable in utility regulation. However, there is a tradeoff between data requirements and the richness of econometric specifications. Econometric functions that place fewer restrictions on the underlying production process and can better capture the impact of a wide range of business conditions on company performance generally have greater data requirements.

While not disparaging the technique in general, we believe that Data Envelope Analysis (DEA) is unlikely to be an appropriate method for benchmarking the performance of electricity networks. This is particularly true in countries like Australia, where there is relatively little data on domestic energy networks. There are a number of problems associated with implementing DEA for energy networks, including problems with capital measurement, dealing with distance and service quality, and the impact of non-comparable variables in international datasets. These problems have generally not been recognized to date, and they augment the more widely-recognized limitations of DEA as a benchmarking tool.

5.2 Directions for Further Research

While our analysis suggests that certain benchmarking approaches are more valuable than others, we also recognize that the use of benchmarking in utility regulation is in its infancy and more work is needed. We therefore suggest a number of directions for further research that may be promising.

One identifiable research project is to apply the Bauer et al methodology to DB benchmarking studies. This would require undertaking DEA and econometric analyses for a single sample of utility companies and seeing how the results compared with their six specified criteria. To our knowledge, such a study has never been conducted for utility companies, and it would have to be done carefully, particularly in evaluating the relationship between efficiency measures and broader market conditions. This relationship is not as clear-cut for DBs as in competitive markets because of the role that regulators play in setting allowed returns. Nevertheless, the Bauer et al paper represents a rigorous basis for determining which benchmarking approaches are most reasonable, and it should be extended to power distribution.

Another important issue is the relationships between cost, service quality and other DB business conditions. As we have stated, these are inherently inter-related issues that have received almost no theoretical or empirical attention. Theory can be directed towards appropriately modeling these relationships. Empirical research can focus on gathering the best service quality data that are available, ensuring that these data are defined and measured consistently, and exploring econometric techniques for integrating quality into a comprehensive benchmarking analysis.

Another research area may be developing larger samples, particularly for rural-based utilities. Even though econometrics can quantify a wide array of differences in business conditions, confidence in benchmark predictions will decline for companies that diverge from the mean in the sample that is used to estimate the model. Australia has a significant number of DBs that operate under very rural conditions. Few investor-owned DBs in the US or Europe operate under similarly rural territories. Benchmarking predictions for Australia's rural DBs may be enhanced if econometric models are estimated using samples with a greater number of rural DBs. It may therefore be useful to begin developing databases on rural power distributors in other developed countries. Examples may include US electric cooperatives and some New Zealand DBs.

There are also some potentially valuable research areas in terms of how benchmarking should be applied in regulation. One issue may be what constitutes an appropriate long-run external standard for utility industries. Information on this topic may be gathered from examining the relationship between average and superior performance levels in competitive industries. As discussed in Chapter 2, this relationship can be useful for evaluating the impact of competitive market discipline on the performance of a typical firm. This data can then be used to implement a very strong form of the competitive market paradigm.

Another research topic may be how long it is expected to take for DBs to attain superior performance levels. As noted, it typically takes time for efficiency gains to be realized. Some regulators may ask how long this process can be expected to be for DBs. Some information on this topic may be available by examining the experience of competitive industries, particularly those (like power distributors) that have a relatively high percentage of "sunk" capital.