SUMICSID

Benchmarking for Regulation

PREPROJECT 4 – FINAL REPORT

Per Agrell
Peter Bogetoft

2003-02-07

Disclaimer

This is the final report on a pre-project on the regulatory benchmarking process, commissioned by the Norwegian Water Resources and Energy Directorate (*NVE*), delivered 2003-02-10 by the authors, professors Per AGRELL and Peter BOGETOFT for SUMICSID AB.

The contents has been subject only to a brief review from the Commissionee and expresses only the viewpoint of the authors, who exclusively bear the responsibility for any possible errors.

Table of Contents

Summary

Benchmarking has become the most cost-effective mean by principals and regulators to deal with the information asymmetry problem in subcontracting and regulation. Not only does it provide information on costs and attained performance for comparable firms, it constitutes in itself a principal element in the regulatory regime. Norway is one of the very first countries to operationalize the integration of an advanced benchmarking approach into a performance based regulation scheme. This report discusses the NVE benchmarking model from three perspectives.

First, any benchmarking method must be viewed in relation to the purpose and context. We makes this precise by three fundamental questions: Why? What? Whom? The objectives for the benchmarking element may have orientations towards learning, i.e. information dissemination to firms on best practice, motivation, i.e. provision of rewards for firms owners or managers, or coordination, i.e. information to facilitate reallocation of resources within the sector. The choice of scope (What?) and reference set (Whom?) follow the first answer and the structure of the industry.

Second, the benchmarking methodology progresses rapidly. To evaluate the Norwegian experience, we therefore include at non-technical state-of-the-art survey with a particular view to the different regulatory uses of benchmarking. We introduce a 2x2 taxonomy of methods and discuss the essential procedural characteristics of good benchmarking. Based on this, we conclude that the Norwegian model seems well-executed although surprisingly simple and without too much discriminatory power. We identify several routes for further progress, viz. the introduction of more detailed models by combining with partial weight and (industrial) norms, the support of learning via interactive benchmarking, the support of motivation via more detailed analysis of the courses and values of inefficiency and the support of more advanced coordination via extended structural analysis.

Third, the rich international empirical evidence from DEA studies witness about a high diversity of models that is nothing but a reflection of the evolutionary and institutional diversity in regulation. Countries have chosen different positioning with respect to benchmarking and its subsequent integration in the regulation schemes.

Based on these three perspectives and the challenges posed by the change of regulatory mode in FP2 and the non-grid technologies in FP5, we find that the NVE benchmarking process has a high value for the regulation. The high data quality and the smooth integration are excellent basis for a further review to improve its positioning towards either higher discretionary power or higher informational contents. The report contains proposals for further studies that could support NVE in this endeavor.

1. Introduction

Background

1.01 The Norwegian Water Resources and Energy Directorate (NVE) is appointed regulator for the electricity distribution and transmission sectors in Norway. Currently, NVE operates an individualized revenue-cap system for electricity distribution concessionaires with five-year regulation periods. The regulatory regime will be unconditionally revised effective from 2007, which means that the regulator NVE on behalf of the Oil- and Energy Department (OED) of the Government will investigate alternative regimes until 2004, when they have to be settled. To anchor the potential reforms, the investigations are to be intensified during 2003. The Oil- and Energy Department has commissioned a study by SNF on the principles of network regulation (von der Fehr *et al*., 2002), which will guide the further work where applicable.

1.02 Based on individual reflection and the SNF report, NVE has defined five pre-projects to be concluded in 2002 and early 2003:

1) Degrees of freedom in the NVE choice of regime? ECON
2) Ex-post vs. ex-ante regulation. SUMICSID
3) Survey of existing evaluations of the current regime. PWCoopers
4) Efficiency analysis and benchmarking in regulation. SUMICSID
5) Incentives for non-grid technological innovation in regulation. ECON, Sefas, SUMICSID.

1.03 The project committees of pre-projects 1 thru 4 consist of NVE, OED and representatives for industry and consumer organizations, for pre-project 5 the responsibility is shared by NVE and ENOVA. After the reporting of all pre-projects, the project committees may decide to launch full-scale projects in one or more areas as in 1.02.

1.04 This report is the final report on pre-project 4, commissioned by the Norwegian Water Resources and Energy Directorate (NVE) and is authored by senior associates, professors Per Agrell and Peter Bogetoft from SUMICSID AB. The report has profited by feedback on a previous draft by senior advisors Kari Ekelund Thørud and Eva Naess Karlsen, NVE, and by consultant Erik Dugstad, ECON. The principal ideas have been presented to the working groups at a seminar 2003-02-03.

Objectives

1.05 The objectives of this study are to discuss criteria and issues for consideration when evaluating the NVE benchmarking model for electricity distribution in anticipation of future regulatory changes. The evaluation should rely equally on international empirical evidence, benchmarking state-of-the-art and theoretical support from information and regulation economics.

1.06 The outcome of this preproject should also bring the formulation of one or more projects to improve upon the NVE benchmarking procedure, with explicit consideration to other ongoing and suggested projects in the regulatory reform period.

Outline

1.07 The principal outline of the report is illustrated in Figure 1.1 below. A brief summary of the Norwegian benchmarking models is given in Chapter 2. A conceptual framework for benchmarking in general and a common thread for the report are provided in Chapter 3. Methodological comments to benchmarking are the subject of Chapter 4, followed by empirical comments in Chapter 5 and a note on information economics related to benchmarking in Chapter 6. The pre-report concludes with some crucial questions and challenges for the current performance assessment models in Chapter 7 and some proposed further projects to address these in Chapter 8. The bibliography offers further reading on the empirical and theoretical topics of this study.

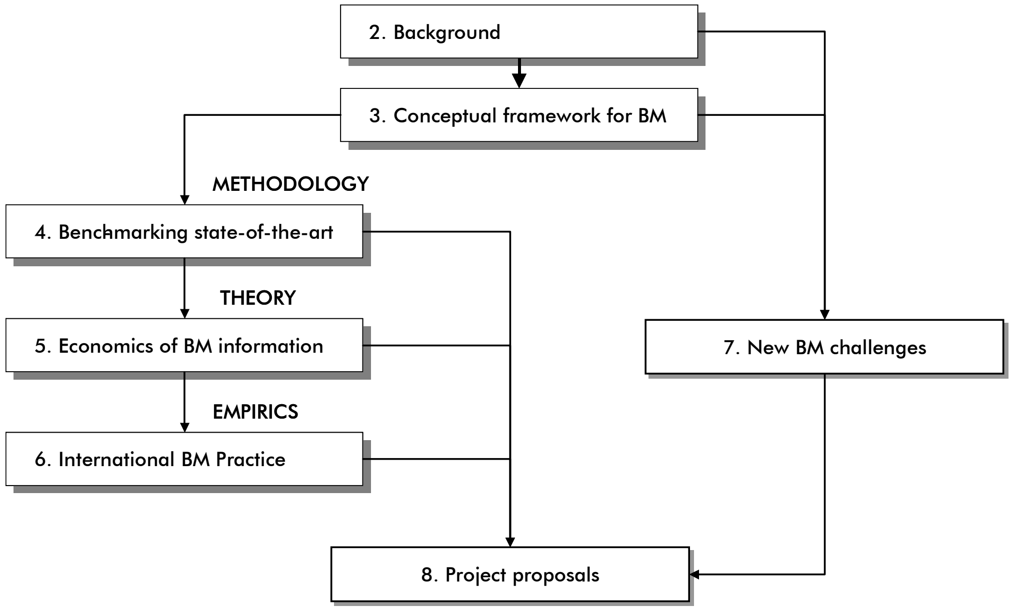


Figure 1.1 Report outline.

2. Background

2.01 In this chapter, we briefly summarize the current structure of the Norwegian benchmarking for regulation of electricity distribution utilities. The chapter will then serve as a base for further analysis and comparison.

The NVE benchmarking models

2.02 Ever since the 1991 Energy Act, NVE initiated limited benchmarking exercises using key performance ratios to monitor and motivate efficiency improvements in the incumbent cost-plus regime. The culmination of this predecessor to the DEA regulation model was probably the NVE (23/1997) benchmarking software tool that was publicly distributed. However, not before the efficiency requirement was individualized did NVE synthesize the benchmarking model.

2.03 The Data Envelopment Analysis (DEA) benchmarking model of NVE has been documented in Kittelsen (1993, 1994, 1996), Kittelsen and Torgersen (1993) and NVE (1994, 1995, 1996). In this context we do not give any general recollection of the DEA method, for an introduction see, e.g., Agrell and Bogetoft (2001). The model is based on classical activity analysis and production theory, using annually aggregated figures for inputs, outputs and environmental factors. The structure is given in Figure 2.1 below.

2.04 The input operating costs (opex) is calculated from staffhours (fulltime equivalents), physical netlosses, costs for materials and services and the regulatory asset base in accounting and replacement value. The conversion was made for 1994/95 using a set of a priori prices for all but capital costs, where an individual calculation has been made. Accounting unbundling for activated staff costs has been made by the firms, as well as the allocation of joint costs for distribution and regional transmission grids. In the 1999 run, actual staff costs were used, leaving the net losses as the only item with estimated price. Compared to Kittelsen (1994), the disaggregated model had also the number of netstations as an input. However, as no technical efficiency is calculated, this change is offset by the inclusion of the asset register in the calculation of the replacement capital.

2.05 The outputs represent the fixed connection charge (number of connection) and the variable costs, primarily the losses in the grid (energy distributed). This choice corresponds to Kittelsen (1994), but is more narrow than the complementary model suggested by the author where outputs were subdivided into customer segments.

2.06 The exogenous delivery conditions are represented with the sole variable length of network (km). The advantage with the current variable is of course its availability and well-defined character. The disadvantage is that it rests a decision variable in the long run. Kittelsen (1994) reports in-depth econometric studies of three other proxies: roadlength, corrosion index and climate index. Finally, the added complexity and methodological problems were considered larger than the marginal benefit from a tighter modelling of the operating conditions.

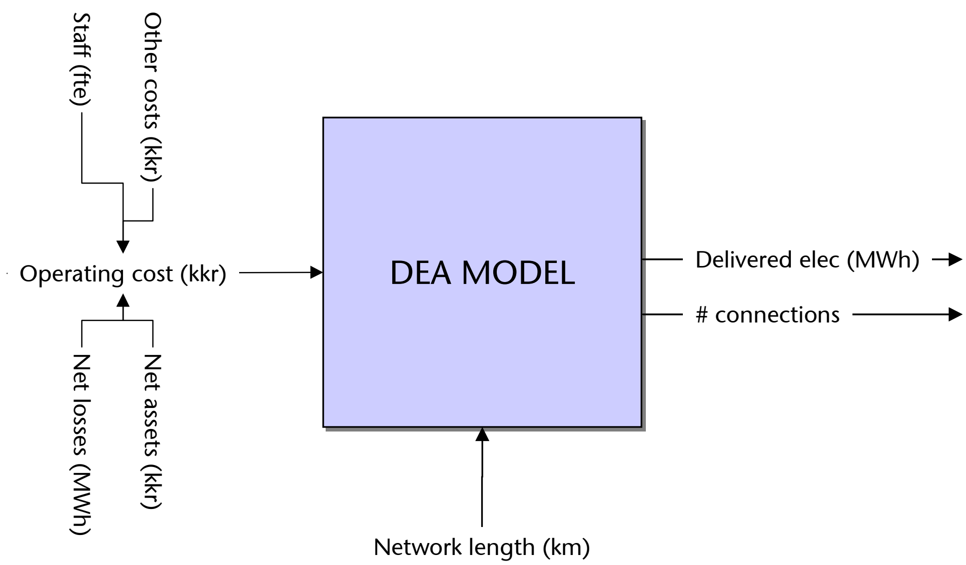


Figure 2.1. Inputs, outputs and environmental factors for the NVE DEA model.

Parameters

2.07 The original NVE model in Figure 2.1 is an input-minimizing costefficiency model, run under variable returns to scale assumption. The environmental variable is assumed continuous and non-categorical, thus technically it is treated as an additional output. The raw data is screened using ratio indicators before runs, eliminating firms with unusually high or low values on key variables. We did not find methodological comments to this data validation process and cannot evaluate to which extent the pre-screening intervenes in the regulatory application of the benchmarking. The final score was calculated as the maximum of the scores with accounting capital and replacement capital.

2.08 Two revised models for distribution and regional transmission were presented in NVE(2001). The distribution model, illustrated in Figure 2.2 below, includes the *a priori* estimated cost of non-delivered energy, KILE, to account for quality differences among firms. The actual cost of non-delivered energy is added to the operating cost, whereas the anticipated cost is added to the exogenous variables as an indicator of operating quality. The problem of the underlying stochasticity was addressed by using a four year average. The model for regional transmission (Figure 2.3) draws on the 1996/99 distribution model for inputs and environmental factors, but is fundamentally different in its output definition. Except for the peak power variable, the two other outputs are essentially weighted assets indices for transforming equipment and central grid installations, respectively. The indices are determined using the 1994/95 weights.

2.09 A specific model (ECON, 1998) using pairwise ranking of operating and construction costs was used for the determination of the individual efficiency target for the TSO Statnett. This model and the TSO regulation in general are subject to particular attention in other projects (cf. Charter of Accountability, Agrell and Bogetoft, 2002) and not specifically analyzed in this report.

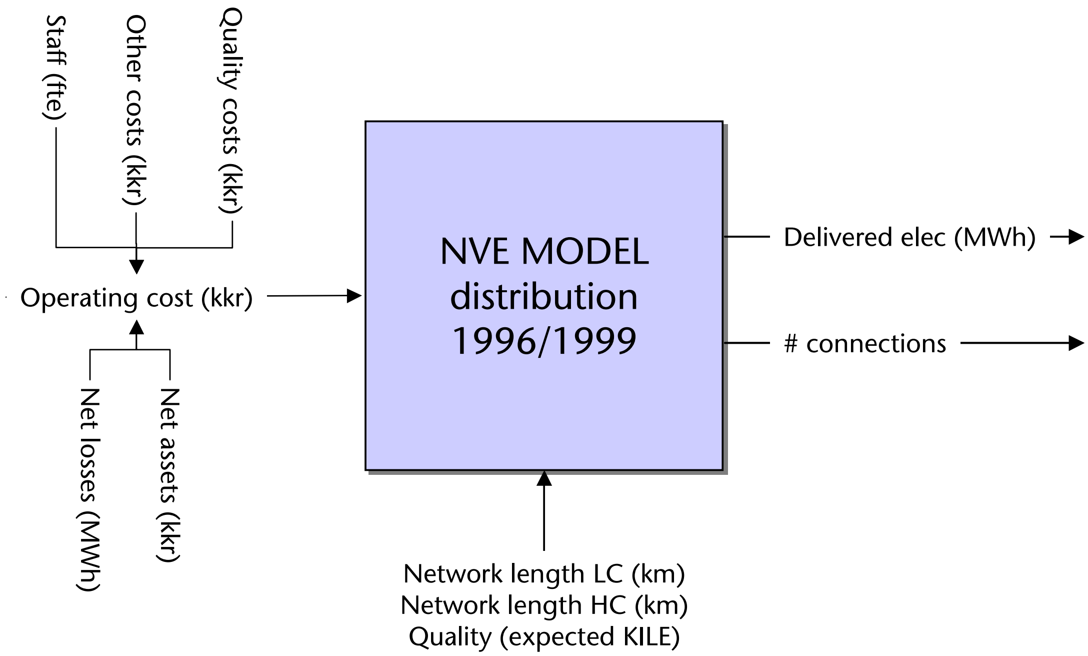


Figure 2.2 NVE efficiency model 1996/99 for local distribution firms.

Runs

2.10 The first regulation period beginning from 1997 introduced individual efficiency targets for distribution nets from 1998, based on DEA production runs using average values for 1994/1995. The reference set was formed of 197 firms after elimination of 38 regional grids and 11 distribution utilities. The average score for the distribution utilities was 86% and 34 units were ranked as fully efficient.

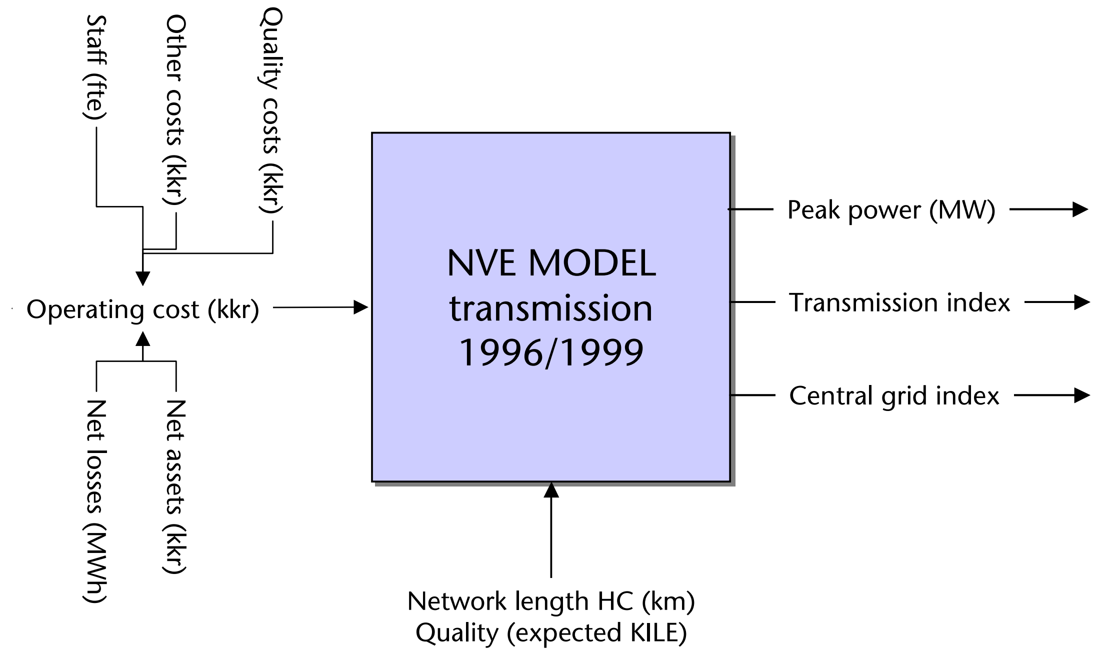


Figure 2.3 NVE efficiency model 1996/99 for regional transmission grids.

2.11 The subsequent run on averaged 1996-1999 data to set 2002-2007 targets was made on 171 (6 excluded) distribution utilities and 83 (21 excluded) regional grid companies and resulted in average cost efficiencies of 90% and 95% for distribution and regional network operations, respectively. The total number of efficient units were 43 (25% of reference set) and 48 (58% of reference set) for distribution and regional network operations, respectively. The results were implemented for regional grids in 1999.

2.12 A dynamic Malmquist analysis was made on 1995/98 data to provide data for the general productivity development parameter X.

Capping

2.13 A particular feature in the NVE implementation of scores is the floor at 70% (1994/95) and 50% (1996/99) to which all score inferior to the floor are truncated. The determination of the floor was based on regulatory discretion, likely to avoid outliers from being overly penalized. The lowered floor can then be interpreted as a tougher policy under improved data quality.

Reorientation

2.14 Although consultants have used the NVE model, no initiative has been made to facilitate the individual firm's apprehension of the performance assessment. In the context of the new regulation regime, this may be perceived as a reorientation from the 1991 learning-based initiative. We will return to the implications of the chosen positioning for the effectiveness of the approach.

3. Conceptual Framework for Benchmarking

3.01 Evaluations of benchmarking methods easily end up in detailed and incommensurate critique of structural and parametrical choices. Part of the problem is the lack of a conceptual framework for benchmarking that clarifies the objectives, scope and means of the exercise in order to enable proper comparisons. This chapter is an attempt to address three crucial issues in the formulation of performance assessment models for regulatory use. The concepts briefly introduced below will then be subject to further and more rigorous analysis from a methodological (chapter 4), theoretical (chapter 5) and empirical (chapter 6) viewpoint.

The Three Fundamental Questions of Benchmarking

3.02 The conception of any benchmarking model and its effectiveness should primarily be judged against the responses to three fundamental questions: Benchmarking – *Why? What? Whom?*

3.03 W*hy*? The objectives behind the model and its intended usage is certainly of importance for its optimal structure. Is the model to inform decisions, to promote learning, or to influence client decision-making? Are the data and results public, restricted or private to the regulator? The resources allocated to its development, updating and operation are also to be considered. Finally the time frame of application influences the potential performance in a given industry. The discussion on this question provides the regulation economics frame for the approach.

3.04 *What*? The definitions of the allowable inputs, the controlled exogenous conditions and the promoted outputs provide the methodological frame for the approach. Which factors are measured, and which factors are controllable during the horizon of measurement?

3.05 *Whom*? The choice of initial and allowable future decision making units for evaluation relates to the sector scope of the assessment exercise and provides the industrial organization perspective o benchmarking. Which units are included in the reference set, how are they selected and how do they change with industrial structural changes?

3.06 Below, we develop these three perspectives before positioning the Norwegian approach in its context.

Why?

3.07 Benchmarking is one of several possible instruments for a regulator (or a firm) to assess the relative performance standard compared to a set of selected observations. The use of a formalized performance assessment model, such as DEA, permits the economic interpretation and usage of the results in an objective and systematic fashion. However, the compilation and presentation of the results entirely depend on the intended usage.

3.08 In broad terms, the objectives of regulators employing benchmarking can be related to one or more of the categories *learning, motivation* and *coordination*. Although the preliminaries of performance assessment exercises normally contain arguments from all three categories, the design and execution of the model reveals the importance weights associated to each issue.

Learning

3.09 If the model is initiated to support learning and efficiency improvements by regulated firms, the approach is characterized more by information openness towards to firms than towards the market. Industry benchmarking projects are usually characterized by a reciprocity principle for data sharing, stating that data and results are restricted to participants only. DEA provides in this context particular strengths, as the peers (in convex technologies) or the dominating firms (in nonconvex cases) provide valuable and concrete information for performance improvement targets. However, such information can merely select potential best practice firms, the actual operational changes will necessitate in-depth process benchmarking that may, or may not, be promoted by the participating firms. When learning, it is sufficient if the results are correlated with best practice, they do not necessarily need to be adjusted to the exact level. Consequently, purely exploratory models may increase the information contents for forward-looking action by not controlling meticulously for incumbent capital structures and sunk investments. If the question of *reimbursement* is separated from the question of *efficiency*, firm-level and firm-regulator cooperation may be facilitated in the model development phase.

3.10 The regulatory administration of a benchmarking model for industry learning demands that two conditions are fulfilled. First, there should be some informational advantages of the regulatory, rather than industry, administration of the model. In a light-handed regime, the firms might actually independently develop and administer a model for organizational. The voluntary development, investments and communication of the common accounting system EL-BAS for the Swedish electricity distributors is an example where the industry preempted regulatory intervention by internal coordination. However, the more important advantages lies in the ownership of the model, that prevents capture by incumbent industry to block innovation, and the information gained by the regulator to be used for other purposes below. Second, to be effective, the regulation regime should enable, or at least not discourage, organizational learning and efficiency improvements. Whereas certain low-powered regimes give disincentives to individual learning, certain high-powered yardstick regimes create disincentives to disseminate learning.

Motivation

3.11 As noted in the last paragraph, the issue of motivation or incentive provision is intimately related to performance assessments. Several approaches are possible. The model may be used directly to determine the entire reimbursement, as in a yardstick regime, to calculate an individual efficiency-related element or addition to the reimbursement, or to estimate an industry-wide efficiency parameter, such as the X in the CPI-X regimes. The impact on benchmarking design is primarily linked to (i) the economic importance of the model outcome for the individual firm, (ii) the individual vs. collective implementation of its results, and (iii) the treatment of sunk costs.

3.12 Trivially, the higher the incentive power assigned to the efficiency result, the higher the motivation to report high performance. However, for the regulator this also implies steep requirements on the model with regard to its representativeness, comprehensiveness and stability. In addition, As discussed in pre-project 2 on ex-post regulation, the firms' demand for regulatory commitment before undertaking some investments prompt for careful consideration of the model definition no to cause technology lock-in. Also, firms with predominately municipal and public ownership may be less motivated by the economic rent than the public disclosure of information that shows excellent performance. Electricity distribution, with strong engineering traditions and pride, is furthermore a business where the managerial influence often is stronger than the 'business'-perspective of private owners. Here too, the motivation effect is less costly attained using frequent disclosure of managerial performance, appropriated as rent by the firms' managers, than by firm-level rents that potentially are redistributed to e.g. clients through lower tariffs.

3.13 The outcome of an efficiency analysis may be interpreted either as a sign of *individual* inefficiency or as *industrial* (*sectorwide*) efficiency potential. Motivation can be proved be either individually determined incentives (such as a yardstick regime) or a collective regime (e.g, a revenue cap CPI-X). In the former, firms have strong individual incentives to score well and also to question whether the model captures their individual operating conditions. Depending on the level of detail of the model, this interpretation may lead to a pedagogical as well as judicial problem if high-powered regimes are used. In the latter, firms have *collective incentives* to shown low catch-up capacity. Although this may inspire strong industry associations to collude upon the specification of the model (cf. Agrell and Bogetoft, 2003), such a direct negotiation may also reveal information on environmental factors that might be inconsistently presented by firms under individual regimes. E.g., the industry association in Sweden under the project work for the DEA model chose not the introduce parameters for snow, forest and landscape profile, since these effects largely are captured by a climate proxy.

3.14 Finally, an efficiency analysis has to take some stand on the treatment of sunk cost, especially sunk investments. Book valued capital is usually readily available, but may introduce systematic noise in the case of mixed ownership firm panels. E.g., public firms in earlier cost-plus regimes may have leveled tariffs by varying the depreciation length of some investments or the classification of renovations vs. new investments. Combined with an artificially low capital cost, the book value of such networks may be considerably below the capital required to finance and rebuild the net in a reformed system. NVE (e.g., Grasto, 1995) has taken notice of the problems and chosen a solution where both the accounting value and the replacement value are collected and used in the benchmarking. The firms get the benefit of the doubt when rankings differ between capital bases. Different methods have been chosen to correct for this problem, for which an ideal solution does not exist. Dte in the Netherlands chose to correct retroactively for subnormal capital costs, which necessitated a fairly intricate correction of the capital base. The Danish regulator also recreated the regulatory asset base using pre-defined prices. Belgium, Finland and Sweden chose solutions based on accounting value, but where the results are based on non-monetary proxies (Sweden) or used with discretion (Finland). However, no solution has been forwarded for the fundamental problem of separating the *valuation of assets* from the *implicit revenue generation* of the regulation. That is, two identical assets in a network may have very different true values in a merger depending on whether they are used and the revenue that the purchaser may accrue on them. Thus, replacement value analyses ignore that some investments have been made with equipment and methods that would not have been chosen under the current regime. This *initial asset valuation* problem in benchmarking has been proposed for further attention in the FP4 report. Here, we limit the discussion by noting that unit-invariant methods such as DEA are less sensitive to systematic errors in initial asset valuation with respect to the score and the ranking of firms.

Coordination

3.15 When the objectives of the benchmarking explicitly address the restructuring of the industry or operational changes in management of firms, we say that *coordination* is at hand. In private and industry-specific contexts, benchmarks, tournaments and bidding are extensively used to coordinate operations at optimal cost and performance. A manager of franchises may use benchmarks of operations, not only to motivate local managers, but also to allocate resources and staff according to their profile. It goes without saying that this form of benchmarking requires a careful analysis of the model and the reference set to assure useful results. Managerial discretion plays here an important role to limit the scope of the model to the relevant operation and the reference set to provide meaningful and feasible information. Nevertheless, the early NVE communication, as well as other regulatory benchmarks, alluded to coordination objectives in the presentation of a model, predominately designed for motivation.

3.16 The strength of the coordination benchmarks, their level of detail and relevance, comes with significant disadvantages in regulatory use. First, the information requirements for coordination largely surpass the capacities of the regulator. Second, the benchmarking process involved includes discretionary steps in model design that do not conform to the criteria of jurisprudence and administrative procedure that a regulator must follow. Regulatory models may, if designed with this objective in mind, be used as framework for coordination benchmarking by regulated firms.

What?

3.17 Once the objectives are clearly articulated, the scope of the benchmarking exercise can be outlined. Activity analysis would suggest that all activities under the *control* of the evaluated would be measured. (Note that this logically poses a problem for the definition of inefficiency, as non-activity or surplus costs would be counted as activities. An athlete resting to drink during a marathon is perhaps "inefficient" from an hourly evaluation, but may win the race when his "efficient", albeit dehydrated, peer falls in the finishing race). Production theory defines the inputs as resources expended and outputs as the outcome of the process that has an external value. In theory, this would lead to a unique black-box description of the production process. The operational problem with this approach is that small heterogeneities between process technologies and offerings may render the units incomparable. If the objectives are biased towards learning and coordination, this may be less of a drawback. For motivation, however, the model needs to make some tradeoff between its differentiating capacity and its accuracy in predicting actual operating cost.

3.18 Two methodological approaches have been proposed to this problem. First, the analyst may perform an activity-based costing exercise where the most important controllable activities are represented in the model, say up to an acceptable level of unaccounted costs. Second, using technological information, there may be a minimal number of independent variables that have different valuation across the panel. E.g., electricity distribution corresponds technically to three principal services: capacity (peak and installed power), coverage (network length and number of connections) and transportation (network losses and some variable costs). Any aggregated beyond the three performance dimensions makes some implicit assumptions about their relative valuation.

3.19 *Controllability* as a governing principle for the choice of variables merits a short comment. A model that is aiming for self-improvement through learning has little to offer if the variables are not fully controllable. This is less obvious than it may seem in a regulatory context when more than half the cost is dominated by sunk investments and the output is governed by an exogenous demand for electricity. A total (operational + capital) cost benchmark that provides motivation for regulatory use does not necessarily produce a ranking of good current management. In coordination, finally, the interest of controllability depends on our commitment to continued service. A gas station in a remote area may be doing well with respect to its market potential, but benchmarking revenue per staff hour and investment is still a rational view for the private enterprise. Once again, we face a fundamental problem within benchmarking design which forces us to clarify for which purposes the model primarily has been conceived.

Whom?

3.20 The definition of the unit for benchmarking, or the *decision making unit* (DMU) is a decision that is intertwined with the two preceding. The larger the reference set, the higher is the possibility to differentiate among the units. However, the risk of noisy observations dominating the frontier also increases. The frontier can be composed of data from the same time period, or from multiple time periods. Foreign, constructed or non-ranked observations may be used to span the frontier. In the quest to preserve information, the regulator may even request reporting for predefined concession areas, irrespective of structural changes, or follow the changes by redefining the frontier periodically.

3.21 As we have seen above, a direction towards learning calls for a more disaggregated model with controllable inputs and outputs and a meaningful reference set. In technically more advanced methods, such endeavor would require a substantial reference set over time to give some interesting information. However, these models would then be fairly vulnerable for strategic action on behalf of the firms, mergers and reallocations. Thus, an orientation towards motivation would call for a more limited aggregated set of variables and a stable reference set. In particular for vertically integrated firms, such as the distribution and regional grid operators, there seem to be ample of strategic reflection on the weighting of scores across categories. Since the number of firms is not constant, a firm may reallocate fixed costs to maximize the overall score, given varying "competitive pressure" in the two reference sets.

3.22 The inclusion of hypothetical or foreign observations in the reference set in a model for coordination or motivation, may seem "unfair" from a learning perspective, but the actual economic outcome could be adjusted to an arbitrary level by the incentive power. Mathematically, the introduction of a priori prices or weight restrictions (which is routinely done in many internal learning benchmarks) is equivalent to the introduction of hypothetical observations in a frontier analysis. What is done depends on whether the mechanics of the method, the peers and the point of comparison need to make sense to the regulator, the unit or the public.

4. Benchmarking State of the Art

4.01 In this part, we will discuss state-of-the-art of frontier models and their implementation using DEA or SFA. The aim is to summarize what seems to be particularly important factors contributing to successful use of the different methods, e.g. careful outlier analysis and supplementary sensitivity analysis, stochastic DEA, statistical significance tests theory etc.

4.02 Frontier analysis in general and DEA methods in particular is developing rapidly in theory as well as in practice. There are by now more than 1000 scientific papers and numerous text books focusing on frontier models, c.f. the bibliography on www.deazone.com. This prohibits a balanced and comprehensive coverage of benchmarking approaches within any pre- or main- project. Instead, we offer a discussion of some of the factors that we consider to be of particular importance in regulatory applications.

4.03 We first give a brief overview of the different *types of benchmarking models* and we summarize their pros and cons. We cover parametric and non-parametric models as well as stochastic and non-stochastic models. We also comment briefly on other approaches.

4.04 We next discuss the different *steps in a benchmarking study* and a few more practical aspects related to the *implementation of a benchmarking model*. A good procedural approach is critical but the literature is not too informative about several aspects of this. We therefore offer a survey based on our own understanding and drawing from other lines of literature.

4.05 We close by *a series of recent advances* that could be included in subsequent studies to ensure that the Norwegian analyses continue to be leading in an international context. We discuss i) how the models supports a value for money perspective, ii) how to cope with differences in the operating environments using partial weight information and combinations with engineering models, iii) how to support structural adjustments, iv) how to support learning and self-regulation using flexible internet based benchmarking, and v) how to improve incentive provision and support menu of contracts for alternative ownerships etc by making improved inference about the organizational uses of slack.

TYPES OF BENCHMARKING MODELS

4.06 We first discuss the different types of benchmarking models and we briefly summarize their pros and cons. At a general level, one can distinguish between parametric and non-parametric models on the one hand and between stochastic and non-stochastic models on the other.

Parametric versus non-parametric

4.07 In the modern benchmarking literature (as opposed to traditional statistics), parametric models are characterized by being defined a priori except for a finite set of unknown parameters that are estimated from data. The parameters may refer to the relative importance of different cost drivers or to the parameters in the possibly random noise and efficiency distributions. Non-parametric models are characterized by being much less restricted a priori. Only a broad class of functions – or even production sets – are fixed a priori and data is used to estimate one of these. The classes are so broad as to prohibit a parameterization in terms of a limited number of parameters.

Deterministic versus stochastic models

4.08 In stochastic models, one make a priori allowance for the fact that the individual observation may be somewhat affected by random noise, and tries to identify the underlying mean structure stripped from the impact of the random elements. In non-stochastic elements, the possible noise is suppressed and any variation in data Is considered to contain significant information about the performance of the unit and the shape of the technology.

Taxonomy

4.09 The two dimensions leads to a 2x2 taxonomy of methods as illustrated in table 4.1 below. A few original key references are included.

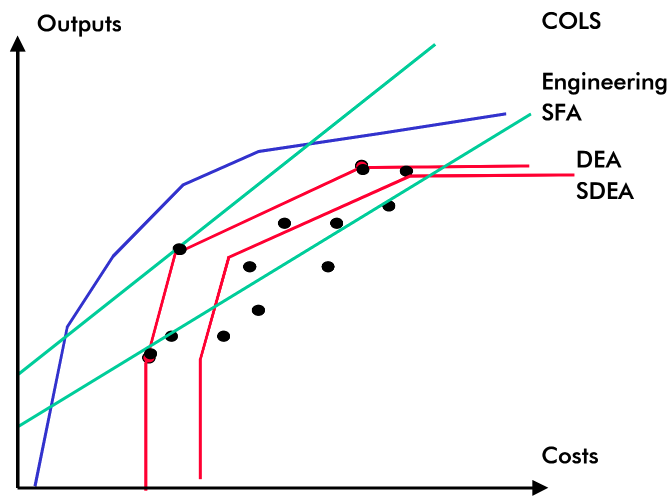
Tabel 4.1 Model taxonomy

	Deterministic	*Stochastic*
Parametric	Corrected Ordinary Least Square (COLS) Greene(1997), Lovell(1993), Aigner and Chu(1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovel and Schmidt (1977), Batesee and Coelli (1992), Coelli, Rao and Battesee (1998)
Non-Parametric	Data Envelopment Analysis (DEA) Charnes, Cooper and Rhodes(1978), Deprins, Simar and Tulkens(1984)	Stochastic Data Envelop-ment Analysis (SDEA) Land, Lovell and Thore(1993), Olesen and Petersen (1995), Weyman-Jones(2001)

4.10 We emphasize that for each class of model, there exist a large set of model variants corresponding to different assumptions about the production technology, the distribution of the noise terms etc. We will discuss the key assumption below. Here, we simply stress that the non-parametric models are the most flexible in terms of the production economic properties that can be invoked while the stochastic models of course are the most flexible in terms of the assumptions one can make about data quality etc.

4.11 We presume a basic knowledge of these models here and are not going to explain them in any details. We simply recall the differences in a simple cost modeling context. The setting then is that we seek to model the costs that results when best practice is used to produce one or more outputs. We have data from a set of production units as indicated in figure 4.1 below. Now, COLS corresponds to estimating an ordinary regression model and than making a parallel shift to make all units be above the minimal cost line. SFA on the other hand recognizes that some of the variation will be noise and only shift the line – in case of a linear mean structure – part of the way towards the COLS line. DEA estimates the technology using the so-called minimal extrapolation principle. It finds the sample production set (i.e. the set over the cost curve) containing data and satisfying a minimum of production economic regularities. Assuming free disposability and convexity, we get the DEA model illustrated in figure 4.1. Like COLS, it is located below all cost-output points, but the functional form is more flexible and the model therefore adapts closer to the data. Finally, SDEA combines the flexible structure with a realization, that some of the variations may be noisy and only requires most of the points to be enveloped.

Figure 4.1 Different models in simple example



4.12 In figure 4.1 we have included a fifth frontier, termed engineering. The idea is to base the modeling on data from engineers about best possible performance, perhaps in idealized settings. We will discuss engineering approaches later in this chapter.

Pros and cons

4.13 We will now focus on the pros and cons of these methods in general, and in particular their relative merits in a regulatory context. Again, it goes beyond the scope of this pre-project to explain and prove the pros and cons in any details. It is our experience however that good scholars will tend to share our few below. In the cases where there may be different opinions among the specialists in these methods, we will give a few more explanations.

4.14 Some of the strengths of non-parametric methods like DEA include

- Requires no or little preference, price or priority information
- Requires no or little technological information
- Makes weak *a priori* assumptions
- Handles multiple inputs and multiple outputs
- Provides reel peers

- Identifies best practice
- Cautious or conservative evaluations (minimal extrapolation)
- Supports learning and in some cases planning and motivation
- Game theoretical foundation of the industry-regulator relation

4.15 Some of the strengths of parametric methods like SFA are

- Strong theory of significance testing (sensitivity, re-sampling, bootstrapping, asymptotic theory)
- Separates noise and efficiency
- Smoothes out some dynamic differences
- May leave lower rents when functional form known
- Creates anonymous peers, may be relevant in regulation

Basic trade-offs

4.16 As indicated, the different approaches have numerous pros and cons. From a regulators point of view, the relative importance of these merits depends on the overall perspective, i.e. the answers to the why, what and whom perspective.

4.17 In our view, however, a basic difference from a general methodological point of view and from regulatory point of view is whether one wants flexibility in the mean structure or precision in the noise separation. The inevitable tradeoff is illustrated in Figure 4.2 below.

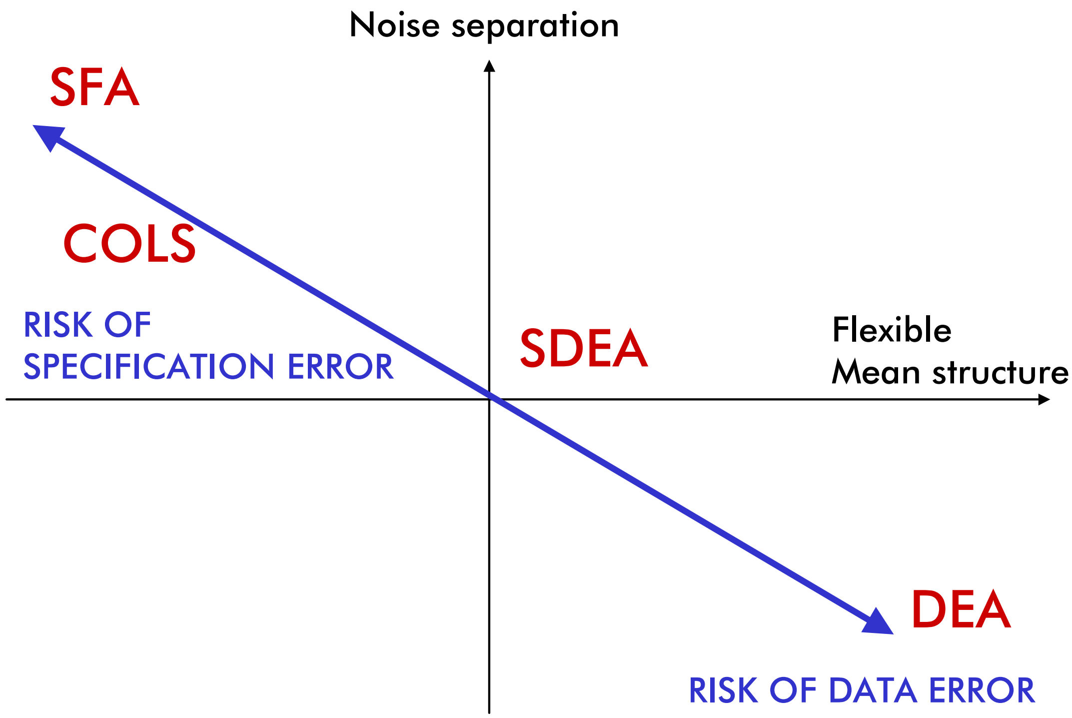


Figure 4.2 Tradeoffs in technology and noise specification, respectively.

4.18 An important property of a benchmarking approach is its ability to reflect and respect the characteristics of the industry. This requires that it is a *flexible* model in the wide sense that its shape (or its mean structure to use statistical terms) is able to adapt to data instead of relying excessively on arbitrary text book assumptions. This is particular important in attempts to support learning, individual motivation and coordination. It is probably less important models aimed at evaluating system wide shifts, e.g. in non-individualized motivation and incentive provision. The non-parametric models are by nature superior in terms of flexibility.

4.19 Another important property of a benchmarking approach is its ability to cope with noisy data. A *robust* estimation method give results that are not too sensitive to random variations in data. This is particularly important in individual benchmarking and perhaps learning – and probably less important in industry wide motivation and coordination studies. The stochastic models are particularly useful in this respect.

4.20 Ideally, then, we would like to use flexible models that are robust to random noise. The problem however is that all of this comes at a cost. The estimation task become bigger, the data need larger and still we cannot avoid a series of strong assumptions about the distributions of the noise terms. Coping with uncertainty requires us to dispense somewhat with flexibility and vice versa.

4.21 We furthermore argue that a lack of stochasticity can be partly compensated by a flexible mean structure – and a restricted mean structure can be somewhat compensated by allowing for random elements. This means that DEA and SFA may be very useful methods and that we do not necessarily need to move to SDEA.

4.22 We will continue our discussion of the pros and cons of parametric versus non-parametric and between stochastic and non-stochastic models below.

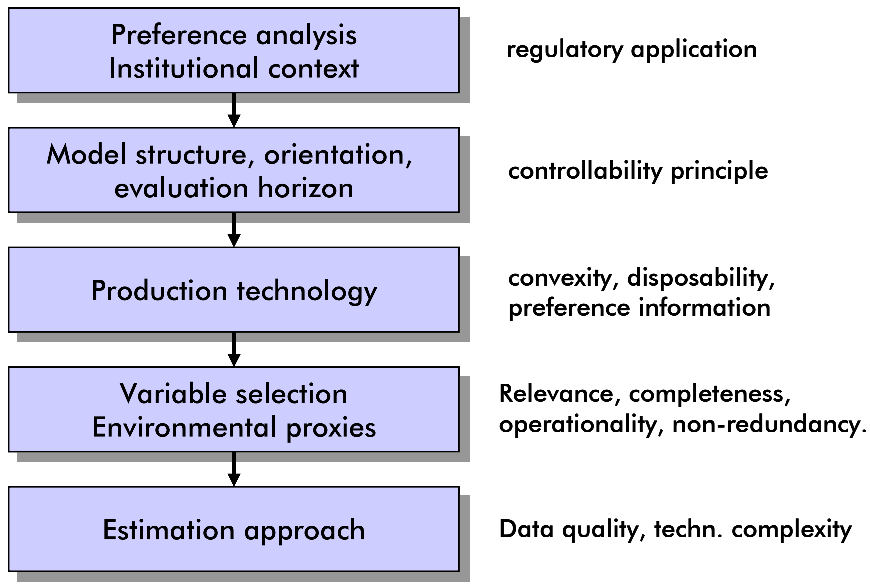
STEPS IN A BENCHMARKING STUDY

4.23 The value of benchmarking tools – as most tools – depend on how skillfully they are used. With the forthcoming of professional computer codes, the ease of efficiency analyses has increased – and hereby also the risk of un-reflected misuses of the frontier approaches. A particular problem in the business of frontier modeling is the lack of simple warning indicators and model specification tests. The risk increases when the modelers do not have string methodological training. Text books seldom contains detailed guidelines for proper uses of the tools they describe. A safeguard against misuses is sound application procedures. Based on our own experience, we now outline a series of relevant steps in such a procedure.

4.24 The model development can include the following steps: 1) Analysis of regulatory interface with benchmarking (preference structure and application), 2) Choice of model structure, orientation and evaluation horizon, 3) Choice of production technology (returns to scale and disposability), 4) Choice of variables and environmental proxies, 5) Choice of estimation approach (parametric or non-parametric)

4.25 The steps are illustrated in figure 4.3 below. We now comment on the individual steps.

Figure 4.3 Model development steps



Regulatory interaction

4.26 The regulatory approach and the benchmarking model are closely interdependent. This is the general theme of this report. Here we simple remind that the scope, frequency and scale of the regulation regime shall ideally guide the choice of optimal benchmarking method. In repeated moderately incentivized settings with audited data collection, deterministic non-parametric methods, such as data envelopment analysis DEA, are often selected as primary benchmarking tools. In one-shot assessments of incumbent inefficiency and settings with high-powered regimes and potentially noisy data, parametric approaches, such as stochastic frontier analysis SFA or multi-output econometrics, are appropriate.

4.27 It is also important to adapt the benchmarking approach to current legislation, as well as to the long-term vision of the regulator. The methodology for this sort of *dynamic regulatory trajectory* has been subject to study in Agrell and Bogetoft (2001) and Estache and Martimort (1998).

Model structure

4.28 The modeling proceeds to investigate the activity under the *controllability principle*. The idea is to tailor the evaluation horizon with the degree of controllability over the activity, if necessary splitting the comprehensive model in a long-run and a short-run model. In distribution regulation, this corresponds to the need to incentivize efficient infrastructure investments as well as efficient grid operation (Sweden, Finland and Norway). However, regulation may also start by assessing stranded cost due to inefficient investments and then operate a comprehensive long-run model with the adjusted capital input (Holland, UK and Denmark).

4.29 The *orientation* is normally given by the controllability principle as well. That it, the discretionary and non-discretionary variables are identified and discretionary inputs (or outputs) are reduced (or expanded). In transmission benchmarking, the focus is usually on cost minimization in a unbundled cost structure, but more advanced solutions may be relevant for integrated utilities. The recent development of directional distance functions offers a flexible approach that can take into account both the *controllability* of different resources and the preferences towards alternative directions.

4.30 The *preferences* for alternative improvement directions may reflect the regulator's trade-offs, say between economic and environmental concerns. Alternatively, the preferences may reflect intra-firm improvement strategies, e.g. as they are settled depending on the relative power of different employee groups.

Production technology

4.31 Non-parametric as well as parametric models usually convexity assumptions, disposability assumptions and return to scale assumptions.

4.32 Most models use a global *convexity* assumption. That is they assume that any weighted average of any pair of feasible productions plans are feasible as well. Although it is widely used and can be motivated in some cases, it is fair to say that it is traditionally assumed for technical convenience to simplify the duality between the production and cost space. Also, in efficiency studies it is done to increase the discriminatory power by extending the production possibility set. On the other hand, there is by now a series of models invoking less convexity assumptions, e.g. Agrell and Tind(2001), Bogetoft(1996), Bogetoft, Tama, Tind(2000), Borger and Kerstens(1996), Deprins, Simar and Tulkens(1984), Petersen (1990), Tulkens(1993). These models are theoretically appealing as they rely less on a priori assumptions and they are in general easier for the industry to accept as they rely less on the idea of mixed organizations – and of course tend to put everyone in a better light.

4.33 In terms of *disposability, i.e.* whether or not the production space is characterized by congestion constraints, rather strong assumptions are usually imposed, say strong free disposability where more inputs can always produce less outputs.

4.34 In terms of *return to scale*, the traditional models either make no assumptions or presume a – possibly local – version of the constant return to scale hypothesis. There are several common motivations to use a constant return to scale assumption like in Norway, i.e. to assume that if we adjust inputs up-ward or down-wards with a given factor, we can do the same on the output side and vice versa. One is that one can always use multiples of smaller units. This prohibits decreasing return to scale were more inputs generates small and smaller increases in the output. A second is – as with convexity - to retain sufficient discriminatory power. A third is that we would like companies to work on the constant return to scale parts of a technology to ensure that they have the right scale. We shall return to this structural argument – which is actually not valid – below.

4.35 The model structure also depends on the possible use of *partial price (preference) information*. The idea is that we may know something about the range of substitution possibilities – e.g. that the resources for one heart surgery exceed that of 3 knee operations and fall short of that of 20 knee operations. The use of such information can reduce informational rents, but it also violates the endogeneity of the input-output weighting in the method.

4.36 The basic assumptions of an efficiency analysis model should ideally be tested. Validation with *statistical tools* allows the analyst to settle on the right model with arguments that withstand industry challenge. There is a growing literature on statistical test but an early approach by Banker(1996) can be used to test all of the above models, i.e. the validity of a constant return to scale assumption, a free disposability assumption and a convexity assumption.

Variables and environmental proxies

4.37 The choice of variables for a given model structure involves looking for a set that is *relevant, complete, operational* and *non-redundant.*

4.38 *Relevance* means that the set of variables should reflect the industry's and the authority's comprehension of the system. The variables should be defined such that decision makers and legislators can relate to and refer to them in the regulation. In the modeling, a compromise is found in the interval between the industry's process-oriented desire to capture the details of the process and the authority's tendency to aggregate to increase comparability.

4.39 *Completeness* means that the set of variables fully capture the objectives (or regulated costs/revenues) of the decision making units. Non-modeled activities are to be explicitly acknowledged to avoid opportunistic action.

4.40 *Operationality* makes it preferable to use variables that are unambiguously defined and measurable. Qualitative indexes, subjective assessments of utility or service value are inadequate in this sense.

4.41 *Non-redundancy* is another word for Occam's razor, prescribing the least complicated means that achieves the end. Overlapping and partially redundant variables may interfere and introduce avoidable noise in the analysis.

4.42 The *model's degree of freedom* is a technical concept that relates the number of observations to the dimensionality of the model. The lower the dimensionality of the model, the higher it's discretionary ability. In the parametric, statistical model, the concept is related to the power of subsequent hypothesis tests. In the non-parametric models, heuristic upper limits on the number of variables have been proposed as well. They require that the number of observations must exceed 3*(no of inputs + no outputs) or (no inputs)(no outputs). With a fair number of distribution companies, this allows for rather flexible non-parametric models.

4.43 Regulatory benchmarking is the art of ensuring a fair treatment of all firms without leaving excessive rents. The proper use of *environmental variables* in the benchmarking models assures these two conflicting objectives. Categorical variables are related to climate, topology, density or other imposed regional heterogeneity in operating conditions. In particular mountainous regions such as Austria, Sweden and Norway are subject to such conditions, whereas models for the fairly homogenous countries in Western Europe have ignored this aspect. E.g., the final regulatory models for Sweden 2000 in Agrell and Bogetoft (2002), included four control variables (climate zone, transforming capacity/interconnection station, subscribed capacity in MW, minimal spanning net-length). However, we recommend that the final choice of environmental variables be made after exhaustive pilot-runs with alternative configurations and statistical tests like above. In this manner, the regulator has access to convincing evidence to various objections to the benchmarking by the regulated firms.

4.44 The choice of variables for the model need not be unique. It can in many case we useful to have an *arsenal of complementary models*. First of all, it gives more credibility to the results if they are verified in a series of models. Secondly, to the extent that the different specifications lead to contradicting results, one can let the benefit of the doubt protect the evaluated – like it is done in a Norwegian context with respect to alternative capital measures, cf. chapter 2 above. The idea of picking the best result fits particularly nicely with the DEA idea of putting everyone in their best possible light. In fact, DEA results can be interpreted as the best results one can obtain using linear (or convex) cost functions, cf. Bogetoft(2000). Thirdly, using a spectrum of specification can be useful to understand the nature of the inefficiency and to decompose the differences among them. Again this has a nice theoretical basis as several types of inefficiency, e.g. technical, scale and allocative inefficiencies are defined precisely from the effects of using one or another model assumption. The use of models with different variables is probably less common than the use of different model assumptions like return to scale assumptions. Still, it is routinely done in second stage analysis of the results. Also, in a Swedish context we have found it very useful to work with cost models that has either direct operating expenditure or direct consumer charges as inputs, cf. chapter 6. By comparing the outcome of the two, one can identify if more efficient companies simply generates more profit to the owners and one can better identify possible strategic behavior

Estimation approach

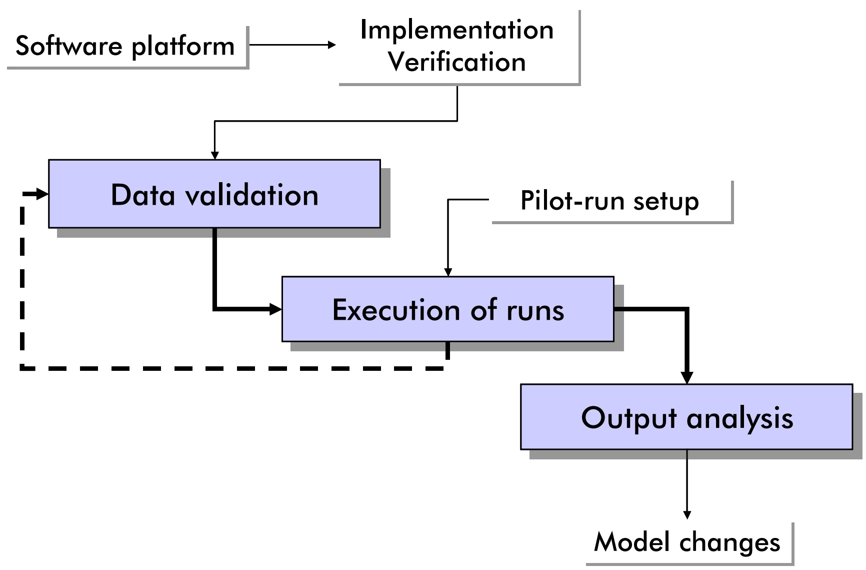
4.45 In the principal choice of an estimation method, a number of issues can be used to evaluate the appropriateness of a particular method. We discussed four classes of estimation methods above, and summarized some of their strength and weaknesses. Here we simple remind that one of important concerns from the point of view of designing incentives based on benchmarking models is the (partial) choice between flexibility and robustness to noise.

STEPS IN MODEL IMPLEMENTATION

4.46 In applied research, it is of course not enough to have good theoretical models and good scientific procedures for how to chose models etc. It is also important to have a good implementation of whatever model is decided on. We discuss a few issues that we find particularly important in the implementation of benchmarking models in a regulatory context.

4.47 The model implementation phase includes steps like 1) Choice of software platform, 2) Implementation of the chosen model using the software, 3) Data validation and review, 4) Experimental design for pilot-runs, 5) Execution of runs, 6) Output analysis. The steps are illustrated in figure 4.4 and commented on below.

Figure 4.4 Stages in model implementation



Software platform

4.48 There are a number of commercial software packages on the market, including OnFront, WDEA, Frontier Analyst, DEA Excell Solver, and EMS for DEA, and DEAP and LIMDEP for SFA and COLS. In terms of speed and reliability, OnFront is a good DEA software although somewhat conservative and restricted in the models one can implement. All the software can probably be useful for running minor, more explorative analysis For large studies undertaken periodically, however it may be more appropriate to use general programming languages and optimization software like SAS or GAMS. This allows the analyst to combine efficient optimization with the flexibility of specifying a individual models. One of the advantageous of the non-parametric model is that it is based very directly on production economics and activity analysis. This makes the formulation of new models, e.g. to estimate the gains from reallocating obligations relatively easy.

Implementation and verification

4.49 In order to minimize implementation errors, it can be useful to cross-validated preliminary models with other software. For many of the dedicated soft wares various types of numerical problems have been demonstrated.

Data validation

4.50 Data for distributors should be careful validate before being used to model estimation. Validation can use partial measures, international comparisons and outlier analysis.

4.51 *Partial measures* can be developed in collaboration with industry specialists. For regulatory DEA models in Sweden, we developed a battery of seven partial measures that was used to filter the reported data. In international studies, we used a simpler databank of three key ratios. Certainly, *international data* material can also be useful to screen and validate the delivered data – especially when the number of comparable units are small.

4.52 *Outlier analysis* consists of screening extreme observations in the model against average performance. Depending on the approach chosen (DEA, SFA), outliers may have different impact. In DEA, particular emphasis is put on the quality of observations that define best practice. The outlier analysis in DEA can use statistical methods as well as the dual formulation, where marginal substitution ratios can reveal whether an observation is likely to contain errors.

Experimental design

4.53 To assure that the chosen model fulfils the model criteria stated above, in particular with respect to environmental and categorical variables, it is useful to make a set of runs to determine the sensitivity of the model with respect to these parameters. The theory for experimental design and inclusion results for non-parametric methods effectively reduce the necessary runs.

Output analysis

4.54 The output (scores, peers, weights and dual prices) from the pilot runs is processed in two steps, where the first step assesses the raw sensitivity of model results to selected parameters. The second step involves regression tests and similar against excluded explanatory factors, which helps evaluating the credibility of the model results and to understand what drives the results.

4.55 We note that inefficiency indicated by benchmarking may be due to one or more explanations:

- Technical inefficiency
- Scale inefficiency
- Allocative inefficiency
- Industry efficiency development
- Excluded variables
- Low capacity utilization following investments
- Non-accounted quality differences in outputs/inputs
- Environment: climate, regulation or local market
- Uncertainty or measurement error

Model changes

4.56 It is important to observe that models are not chosen and implemented in a linear manner. Several rounds of modification will be needed to get a relative robust structure. This is particularly important in motivational applications where the firms' rewards may depend directly on the model results. Frequent shifts in the model and unexpectedly large changes in efficiency and productivity over time will tend to undermine the credibility and commitment capability of the regulator and hereby weaken motivation in the future. It is therefore very important that a robust model structure is selected to begin with.

Status and Further Advances

4.57 Based on the above and combined with the information in chapter 2, we suggest that *the Norwegian model seems well-executed although surprisingly simple and without too much discriminatory power*. We note however that we have not directly analyzed the available data and tested alternative model specifications, i.e. we have not systematically applied the procedural guidelines above to replicate the resulting Norwegian model

4.58 We close this chapter by identifying several routes for further progress, viz. the introduction of more detailed models with sufficient discriminatory power by combining with partial weight and (industrial) norms, the support of learning via interactive benchmarking, the support of motivation via more detailed analysis of the courses and values of inefficiency and the support of more advanced coordination via extended structural analysis.

Detailed modeling using international comparisons, partial weights and engineering extensions

4.59 In view of the relative simple DEA model underlying the Norwegian regulation and the surprisingly large number of efficient units, it is clear that many obvious improvements – including the extension mentioned below – require a more detailed underlying model. The inclusion of more variables, however, may further reduce the discriminatory power, and a need for additional data is therefore obvious. There are at least three ways to extend the data basis.

4.60 One is to use *international data*. This is not a trivial tasks but we believe that the use of in particular detailed Swedish and Danish data would be relatively simple.

4.61 Another is to supplement the actual observations with *engineering results*. The DEA model is well suited for such extensions, and in other businesses this has a long tradition. In agriculture, for example activity based farm models usually rely on a mixture of actual activity information and experimental activity information. Recently, regulators in Chile, Spain and Sweden has experimented with the combination of engineering norm and DEA models. We shall return to this below.

4.62 In terms of the degrees of freedom, there seems to be some room for extending the Norwegian models with more variables given the number of units. In fact, if the so-called primal information about inputs and outputs was supplemented with just *partial so-called dual information* about the relative importance of different types of inputs and outputs, the number of inputs and outputs could be expanded considerably. In a recent study, Petersen and Olesen(2001) estimated a hospital model with more than 700 outputs using less than 100 decision making units by introducing bounds on the relative costs on of different outputs.

Learning and internet based benchmarking

4.63 We have argued repeatedly that it is a potentially important regulatory task to provide performance information to the individual firms. This may inspire individual learning, be a motivation device on its own, in particularly in forms that are not entirely managed by economists, and support structural adjustments by offering information about other firms.

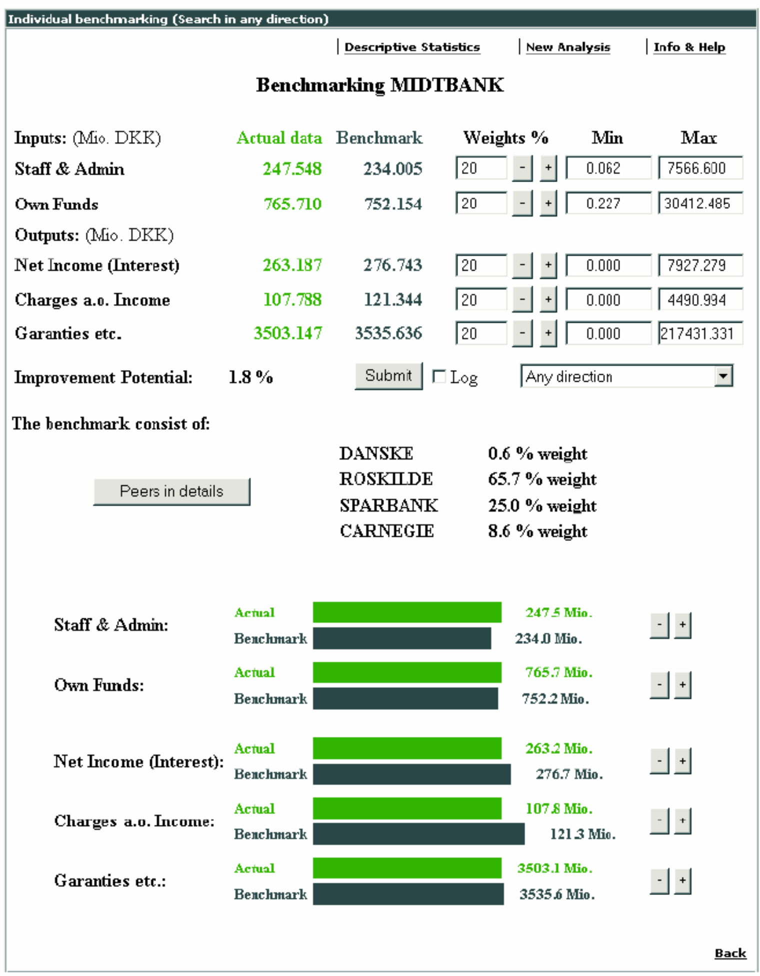
4.64 The learning aspects is also the primary theme of much of the modern benchmarking literature. Without turning to very detailed process benchmarking, many recent contributions combine DEA and Multiple Criteria Decision Making, for example.

4.65 From a learning perspective, the most advanced proposal is probably to let the users, e.g. the firms, undertake user-defined advanced analyses over the internet. This idea is to combine the power of DEA analysis with the flexibility of linear programming and the ease of the internet. The user in such a system can define *whom* to be compared to, e.g. units of approximately the same size or from the same region or with a similar ownership structure, *what* to emphasize, e.g. the relative importance of different outputs, and *how* to improve, e.g. whether or not no focus on reducing certain labor inputs or certain capital inputs.

4.66 The approach has been implemented using simple models of 50 different Danish industries, as well as using advance models of the banking and farming industries, cf. e.g. Bogetoft and Nielsen(2001). Also, in energy regulation, the Swedish regulator has emphasized the internal use of such information by educating key staff to make detailed analysis of individual firms using the DEA basis.

4.67 A screen shot from the Danish banking application is given in Figure 4.5 below. The actual inputs and outputs of the evaluated unit and its benchmark is provided in two columns – or at the lower part of the screen shot as bars. By pressing the + and – bottoms, or by giving numerical weights or constraints, the user can determine which direction in the input-output space he wants to explore. Also, by changing the frontier, he can explore what to do depending on his aspiration in terms of general industrial ranking, say to be in the group of 10% most efficient firms.

Figure 4.5 Screen shot from Danish banks web-based benchmarking system



Motivation and the nature of inefficiency and menu of contracts

4.68 The benchmarking literature has – by and large – considered inefficiency as waste that one would ideally try to eliminate. Also, the literature has been ignorant about the relative distribution of inefficiency among different inputs, e.g. different types of labor.

4.69 A recent contribution of Bogetoft and Hougaard(2002) provides an alternative perspective which is both in better accordance with much of economic and organizational theory. It recognizes the possible benefits of apparent inefficiencies by modeling the preferences of the firms as one of maximizing both profits and slacks. Also, it suggests specific ways to model the trade-off among different types of slacks (inefficiencies).

4.70 One of the applications of this approach is that it allows *better incentive schemes* by capturing more fully the gains from inefficiencies. That is, the stylized assumption in the regulation literature of costs of effort etc can now be substituted with more specific models reflecting the internal power structure among different labor groups or between labor and capital interests, for example. The internal control and incentive structure most likely will vary according to ownership. The approach therefore suggest a way to take this into account and this could potentially form the basis of new regulations using menus of contracts etc.

4.71 Another application of a more detailed inefficiency generation model is that it becomes easier to *predict organizational responses* to changes in the regulatory regime. The impact of regulatory imposed direct constraints (e.g. minimal quality requirements) or changes in relative prices of the inputs and outputs (e.g. via price-caps, taxes etc), for example, will lead to different expansion and contraction paths depending on the organizations value of different types of slack.

Coordination and advanced structural analysis

4.72 As we have argued in our FP5 pre-project, the structural properties of the energy industry (firm scale, scope, ownership etc) may be more important than the details of the regulatory reimbursement schemes. At the same time, the incumbent regulatory regime may have an impact on the structural adjustment, both very directly if the regulators refuse to approve changes in the structure, and indirectly if the payment plans make socially attractive changes non-profitable for the individual firms.

4.73 A good example of these problems is the question of how to treat mergers. When payments are correlated with efficiency, the payment plans will tend to discourage mergers in convex models even though they might lead to more outputs being produced with less inputs. (There are of course ad hoc rules to cope with this, e.g. somewhat arbitrary and sub-optimally to allow the firms in a CPI-X scheme to keep all the gains in the present regulatory period). At the same time, mergers will tend to affect the performance evaluation basis and may lead to more rents to the firms by the cost norm becoming less demanding. The regulator considering to allow a merger must therefore trade-off the gains from improved costs to the firms with the losses from a shrinking information basis.

4.74 Recent advances in structural analysis based on DEA models may be very useful here. Models have been developed that allows the analyst to decompose overall industrial inefficiency into *technical inefficiency* that be reduced via increasing *learning*, *harmony or scope inefficiency* that can be reduced by exchanging inputs and outputs among the firms, e.g. by *outsourcing* certain task to neighboring distribution companies, and *size inefficiency* that deals specifically with the economies of scale and requires genuine *mergers* to be realized.

4.75 It is worthwhile to note that there may very well be gains from increased cooperation even though all companies are optimally scaled in the traditional sense. The reason is that optimal scale size is very direction specific and in fact many of the traditional applications of scale efficiency make limited sense in a world with producers operation in different geographic areas.

4.76 For more on these ideas, see e.g. Bogetoft(2002), Bogetoft, and Wang(1999), Bogetoft, Strange and Thorsen(2001), Brännlund, Färe and Grosskopf(1995) and Brännlund, Chung, Färe and Grosskopf(1998).

5. The Economics of Benchmarking Information

5.01 This Chapter is devoted to the economics of the information revealed by benchmarking, but not merely for use in yardstick regimes and to leverage the classical asymmetric information problem.

Reduce asymmetric information

5.02 Distribution and transmission of electricity are examples of economic activities with considerable difficulty to observe the effectiveness of operations and investments. Apart from the technical difficulties involved in the measurement of the services, time, demand, supply and exogenous heterogeneous service conditions add to the complexity. Finally, widely varying economic and technical asset-lives further complicate the direct observation of performance. Clearly, the operator of such services is better informed than the clients or the regulator. In economic theory, this *asymmetric information* problem is the primary motivation for benchmarking, i.e. comparative studies of production and/or services. Benchmarking can be obtained by direct information acquisition, such as tender offers in contestable markets, by observation of external operations that are correlated to the regulated service, or by construction of ideal observations using experimental or prototype services.

Yardstick models

5.03 Schleifer (1985) showed using a simple average price yardstick that benchmarking information could be effectively used under asymmetric information. In a dynamic single-product scenario with a high number of firms, the firms have strong incentives to reveal their true costs if they are reimbursed with the average cost norm.

5.04 Laffont and Tirole (1986) consolidated the theory by providing results also for imperfect observations, correlated to the cost, and by providing links to the principal-agency literature on optimal effort provision, moral hazard and incentive powers in contracts. The seminal paper, later extended for the dynamic case, by-pass options and various information assumptions in Laffont and Tirole (1993), put a clear foundation for the use of partial cost information to extract rents and parameterize regulation mechanisms.

DEA-based yardstick models

5.05 However promising in theory, the early benchmarking results were not immediately applicable for the multi-product and multi-input case, where econometrics still was a required, yet informational heavy, investment. Seminal work by Bogetoft (1994b, 1995) offered a solution to the moral hazard problem in the utilization of DEA-based norms as a basis for yardstick contracts. The results showed that DEA norms, under deterministic and certain stochastic settings, provide optimal incentives to reveal and maintain technical and allocative efficiency. Further developments by Bogetoft (2000) and Agrell and Bogetoft (2000b) for activity planning and dynamic regulation, respectively, demonstrate promising and practical applications of benchmarking techniques to improve regulation.

Sector measures, TFP and panel data

5.06 Full use of benchmarking information for non-contestable markets (e.g. natural monopolies) has been applied to individual reimbursement only in selected settings, e.g. public transportation and health care (independent practitioners), where sufficient consensus has been reached upon the specification of the benchmarking model. However, use of benchmarking to determine collective productivity targets (e.g. the X-factor) or performance levels (output and quality levels) is widespread among regulators around the world (cf. Chapter 6). Using simple statistical measures to find average performance, econometric techniques to correct for exogenous conditions or non-parametric methods to determine best practice performance, regulators have also informed regulatory rulings and rate reviews, without necessarily formalizing the role of benchmarking or the choice of methodology. Finally, regulators in multiple sectors have provided benchmarking information for learning purposes as a compensation for, or in anticipation of, incentive-based regulation regimes.

Avoid capture

5.07 The publication of benchmarking can be seen theoretically as an act of accountability on behalf of the regulator, that lowers the risk of capture, the regulatory risk and increases social welfare. Consumers may observe the effectiveness of the regulator's actions as well as firms' behavior. Attempts by firms to capture the regulator could be more easily detected and thus less profitable (tempting). Likewise, attempts by pressure groups to capture the regulator to extract inadequate rents could be challenged by the firms using the benchmark.

Cost efficient information processing

5.08 Benchmarks do not have to be used for motivation to increase welfare. The regulator may be the least cost processor of private industry information that firms could use for quicker catch-up under an arbitrary regulation regime. The benchmark format rather than accounting data lowers the search cost of information and provides non-monetary motivation to managers in the firms.

Reporting discipline

5.09 Benchmarks with high explanatory power, more ambitious models, may increase welfare by restraining the firms' scope for opportunistic action, e.g. allocation of costs or claim for exogenous costs. Repeated benchmarking using multiple methods provides valuable information that increases the costs of lying "consistently" with previous reports.

Specific investments

5.10 Boyer, Jacques and Moreaux (1998) study the impact on social welfare when technological choices are made observable on commitment and flexibility. It is concluded that observability lowers the incentive to some specific investments unless the rents can be appropriated quickly enough. On the other hand, firms may also prefer to be observed to discourage competitors to undertake substitute investments in research, that may threaten future rents. In our regulated context, this means that too detailed benchmarking, e.g. coordination information, could lower technological progress. Regimes must also take into account the ratchet-effect on technology catch-up when it comes to innovation, cf. Agrell, Bogetoft, Tind (2002) or FP2.

5.11 Lewis and Poitevin (1995) look at the regulatory review process from an information disclosure viewpoint. Using a simple setup they show that welfare is maximized when the firm provides evidence for its deviation from a general assumption, rather than the opposite. It also shows that mandatory, rather than voluntary, disclosure of additional information does not increase efficiency. Benchmarking methods can thus provide regulators with a standardized costefficient information gathering instrument that could focus the proceedings on special conditions. Note that the authors are economists and not lawyers.

Summary

5.12 Economic theory provides other rationale than the improved cost information to conduct performance assessment studies. Relative performance provides non-monetary motivation to managers, transmits high-quality information to facilitate faster productivity catch-up and signals regulatory priorities and independence. However, regulation instruments based on incomplete task description also promote myopic and opportunistic behavior by firms, limiting the social welfare gains from overly high-powered incentive schemes.

5.13 The main result in the current context is that even a non-complete model, be it a model primarily for learning or motivation, may still incite positive effects on market development.

6. International Benchmarking Practice

6.01 In this part, we will describe the benchmarking models used in several other countries. The countries include Sweden, Finland, Denmark, Holland, England, Spain, New Zealand and Australia. This part draws on private and published information on the regulatory use of frontier models, not only limited to electricity distribution.

Diversity or Chaos?

6.02 The majority of Western electricity regulators use, or plan to use, benchmarking instruments for distribution utilities, cf. Table 6.1. Early implementations of primitive econometric COLS benchmarks in 1980s have been followed with a wave of non-parametric DEA applications and some technical benchmarks in the very late 1990s. A listing of the inputs, outputs and environmental variables of these models is given in Table 6.3 below. Immediately, two findings may strike the reader. First, DEA enjoys a spectacular popularity among regulators and utilities. Second, the structure, scope and reference sets of the models are widely varying. Jamasb and Pollitt (2001) report on the frequency of variable classifications in 20 benchmarking studies for electricity distribution (cf. Table 6.2) and find a number of factors that are used both as input and output in the sample. Indeed, how come the non-parametric techniques are so popular in regulation if there is no consensus on their specifications?

6.03 Returning to the framework in Chapter 3, we find that the conception of the models largely explain their structure. Let us distinguish the models with the triple i/o/e for inputs/outputs/environmental factors, the sum of which gives an indication of the dimensionality of the model. Studies that pursue a learning orientation have generally more environmental variables and mixed reference sets. Consider e.g. the 1/1/6 study of UK distribution made by Weyman-Jones (1985). Pure regulatory applications to operationalize the sector X-factor are more aggregated and involve often international comparators (extreme case OFGEM 1/1/0, but also the 1/5/1 model of Dte), whereas individualized regimes are usually more detailed and use national data, such as the Swedish model 1/5/3. Finally, coordination models use very specific data and few environmental variables, e.g. the 5/4/1 model of the Swedish utility Sydkraft.

Table 6.1 Regulation regimes and benchmarking methods.

Country	Reg.App.	Eval.Meth.	Development / In use
Australia	Ex ante	CPI-DEA/SFA/Stat	U
Austria	Ex post	DEA/SFA	D
Denmark	Ex ante	CPI-COLS	U – with renegotiation
England	Ex ante	CPI-DEA/COLS	U
Finland	Ex post	DEA	U
Netherlands	Ex ante	CPI-DEA	U
New Zealand	Ex ante	CPI-DEA	U
Norway	Ex ante	CPI-DEA	U
Spain	Ex ante	Ideal-Net	D
Sweden	Ex post	DEA/Ideal-net	U/D

6.04 The seemingly contradictory classification of variables can also be approached from the same viewpoint: the needs of the analysis and the character of the variables depend on the horizon of study. Whereas a long-run model, such as the Swedish LR model, employs the an exogenous proxy for the network size, short-run models in the Scandinavian annual ex-post regimes use the actual network length as an environmental variable. Similar arguments can be made for network losses, that are byproducts of the distribution services in the long run, but whose additional cost can be controllable in the short run (as in the NVE model). In general terms, the models have the relationship that is depicted in Figure 6.1 below. In the long run, there are no fixed inputs, only exogenous conditions. For shorter evaluation horizons, a certain share of the inputs must be considered as fixed, if we are evaluating the management of the activity. For general coordination purposes, these allowances are not necessary. We may thus imagine the various models classified, not only according to their positioning for the three questions Why? What? Whom?, but also according to their evaluation horizon.

6.05 When considering the diversity of models, the sector structure should also be taken into account. The Scandinavian situation with many independent observations of fairly high quality is an international exception. Countries such as UK, the Netherlands, Chile and Spain have demonstrated great willingness to develop yardstick schemes on benchmarking schemes, but most have been forced to resort to parametric (UK) or technical methods (Chile, Spain) due to limited or low-quality national datasets. In situations with limited data sets, access to international data may limit the informational rents.

However, this demands the compilation of high-quality data of comparable processes. Jamasb and Pollitt (2000, 2001) and Edvardsen and Førsund (2001) emphasize the need for international standardization of data and environmental modeling as a prerequisite for international benchmarking. The NorDEA (or Nordic Open) initiative to benchmark larger distribution utilities (Edvardsen and Førsund, 2001) is a positive experience in this context of higher usefulness than other projects with larger membership. Nevertheless, we stress that careful analysis must be made as to the current and future access to data before committing to a particular method. We will return to this issue in the next Chapter and the Future projects.

Table 6.2 Frequency of variable classification in 20 benchmarks (Jamasb-Pollitt, 2001).

Variable	Input	Output
Energy delivered	2	12
No of customers	1	11
Split in HC/LC		5
Network size	11	4
Split in HC/LC	2	
Transformer capacity	11	1
No of transformers		1
Service area	2	6
Maximum demand	1	4
Net losses	4	-
Labor	15	-
OpEx	7	-
Administrative cost	2	
Maintenance cost	1	
CAPEX	1	
Capital	5	
Customer density	2	

Single vs. Multiple models

6.06 As discussed in the FP2 report on regulatory formalism and discretion, a regulator may choose to constrain his discretion by using a formal model, or to support his discretion by using a battery of models. Let us illustrate with two examples.

6.07 The British regulator OFGEM has officially settled on a cost benchmarking process for the 12 national distribution zones. However, the regulator has also commissioned other studies and supports a series of technical decisions for the benchmarking with various arguments from the supplementary methods. The decision making process has been accused of lack of transparency and regulatory opportunism. Here, the translation of scores to revenues through the general X-factor becomes the focus of the firms' interest. To run the strictly motivational benchmark, that carries very limited firm-level information, with the limited national data set, the choice of methodology is necessarily very constrained.

6.08 In the case of the Swedish regulator STEM, the revenue cap for the 240 firms is set by a technical model, corresponding to the monetary motivation objective, and the performance assessment in the long- and short run is made through a family of interconnected non-parametric models, corresponding to the learning objective. Industry independently develops and refines the public data to run coordination models (cf. Sydkraft) using the same and other methodologies. In this scenario, the learning oriented benchmarks have to be focused more at explanatory power than differentiation among units for the approach to be consistent. However, the separation of revenue generation format from performance assessment limits the firms' incentive to manipulate data for short-term monetary benefits. Instead, managers, competitors and customers take interest in the performance assessment to gauge the effectiveness and the viability of the operations strategy of the firm.

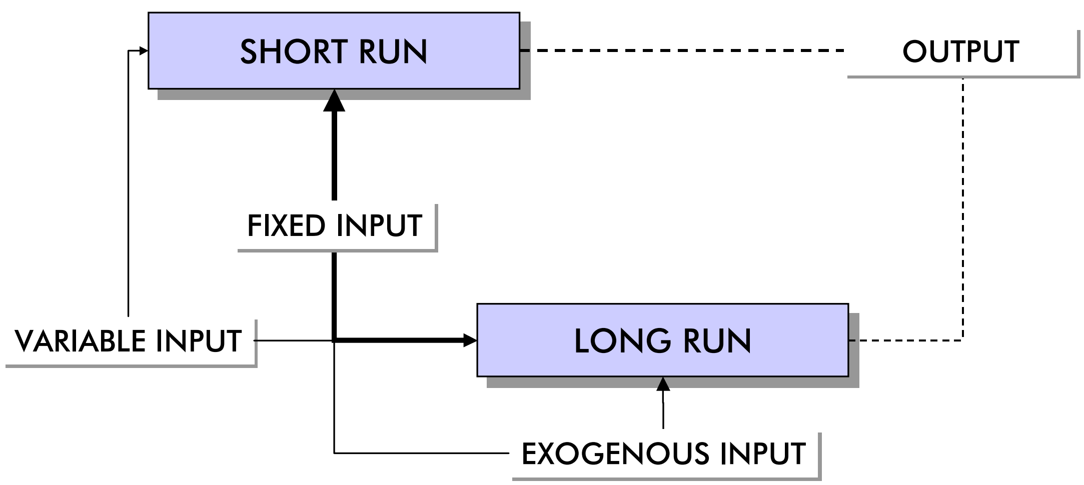


Figure 6.1 What to benchmark depends on why we are benchmarking.

A Dutch Treat to the Firms

6.09 The Dutch regulator Dte launched unilaterally a transitory CPI-DEA regime 2000 based on a non-parametric model with only 20 firms for the first year. The model (Dte in Table 6.3) uses constant returns to scale and is primarily motivational. However, due to the very limited data set and the high aggregation level of the model, the regulator and the firms soon disagreed on the interpretation of the scores. A series of legal appeals and industry-regulator negotiation rounds followed and the final compromise was usually a fairly reasonable individualized catch-up requirement, X = {-2, ..., 8%}. Although the actual benchmarking model and process in this case were refined during the first regulatory period, the aggressive strategy of an under-specified model led to a much faster catch-up of the informational asymmetry – by the regulator! In return for some limited discretion through the model, the regulator gained valuable information with respect to cost drivers and environmental conditions when the burden of proof was reversed. Such results would likely not have been attained had the regulator tried to used pure discretion under asymmetric information, nor with an over-specified learning oriented model. In the future, as the number of firms is likely to continue to decrease, the regulator may further revise the use of the model as the transition period is succeeded by the permanent yardstick regime. The example illustrates how the regulator may use an (imperfect) model to elicit information from the firms at low cost – a true Dutch treat.

Experiences with Technical Benchmarking Models

6.10 At least three countries have developed engineering models to benchmark distribution utilities: Chile (23 firms), Spain (350 firms) and Sweden (240 concessions). The actual technical solutions are different due to the sector differences. The engineering cost models in Chile and Spain are fairly aggregated and the operation is dominated by third-party consultants. In Grifell-Tatje and Lovell (2001) it is shown that the ideal cost provided by the benchmark is not a lower bound to performance, the error being largely standard costing for line constructions. The Swedish model is a mathematical green-field simulator using actual GIS-data for the connection points to clients and the transmission grid. The resulting grid cannot be used for actual operations, but is claimed to provide a lower and upper bound to feasible total costs. The initial investments in technical models are heavy, both for the regulator and the reporting firms (Sweden) and the methodology implicitly draws on a predefined technology. The attractiveness lies in the potentially tight bounds that are completely exogenous for the firms, i.e., the investment problem may be effectively dealt with using quality standards and/or separate quality benchmarks. The temptation of micro-management has to be addressed, either by third-party reliance or algorithmic simplicity. So far, the effectiveness of the Spanish and Chilean systems have been unclear due to poor data, regulatory capture and lack of necessary skills to optimize the method (Ferrier Swier, 2002). The Swedish model has not yet been fully implemented and the light-handed regime relies until 2004 on the DEA models. A principal scheme of the ideal net model and its interface towards the non-parametric learning-oriented models is given in Figure 6.2. Notice how the information contents of the DEA benchmarking models is enhanced by the provision of an exogenous estimate of network density from the ideal net model. Analogously, the normative results from the ideal net model are successively validated against efficiency results, guiding regulatory monitoring prioritization.

Hybrid Methods and The Way Forward

6.11 Austria has recently initiated a regulatory benchmark with a double methodology in which a non-parametric model is supported by an engineering model for environmental parameters. By using engineering estimates for environmental conditions, the Austrian regulator hopes to maintain the motivational strength of an aggregated DEA model under the explanatory strength of an engineering-based benchmark. The project points to a new level of integration between models and approaches in regulatory context, not unlike what has been used in other sectors. It also opens an alternative for the many situations where lack of data previously would urge the benchmark to be made by rough-cut methods.

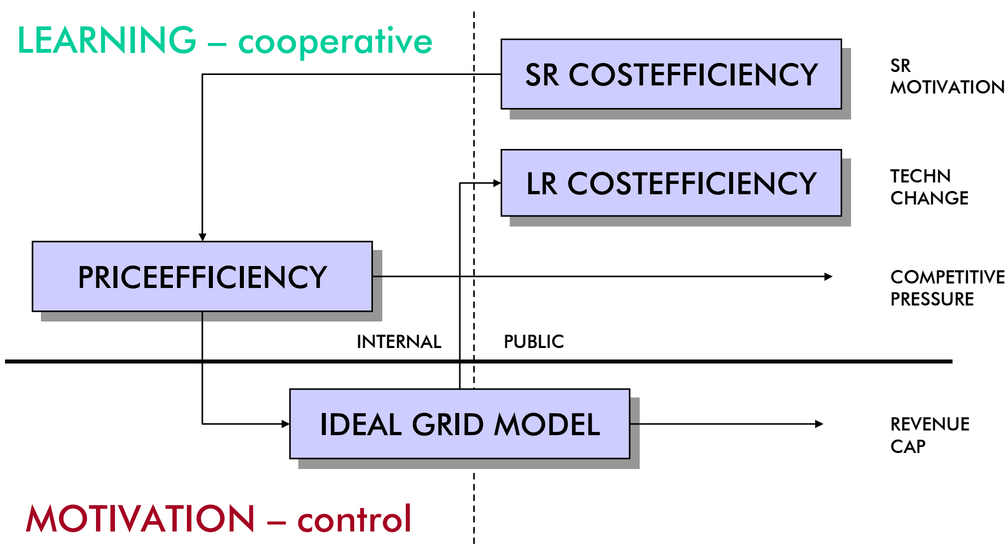


Figure 6.2 The STEM model battery and the benchmarking objectives.

Summary

6.12 The recent surge in the use of benchmarking methods for natural monopolies in general, and electricity networks in particular, is a natural response to the deregulation and the need to bridge the information asymmetry. As argued in Agrell and Bogetoft (2001), the popular non-parametric methods offer many interesting features for regulators, such as the flexible functional form and the conservative efficiency estimates. The development has continued to a greater understanding of the diversity of model specifications and the strengths and limitations of each benchmarking methodology. Rather than looking for the "optimal" benchmark that would satisfy all parties, regulators are acknowledging the trade-off between discriminatory power, explanatory power, data requirement and potential usefulness for learning, motivation and coordination in the regulation. The framework is illustrated in Figure 6.3 below and will be used in the conclusions. To aim for the coordination models upper right is infeasible in most regimes due to data access and division of tasks in the market. The regulator can then hope to either get an optimal motivation model in the upper left with sufficient relevance to be accepted, or to go for a learning-oriented model with more allowance for individual conditions at the expense of informational rents. If data or administrative resources are scarce, the solution in the lower left may be the only feasible, simple aggregate cost benchmarks that neither are robust to environmental noise, nor to accounting maneuvers in integrated firms. Without considering the context and the entire regulatory regime, as discussed in FP2, the impact on social welfare by the choice cannot be determined.

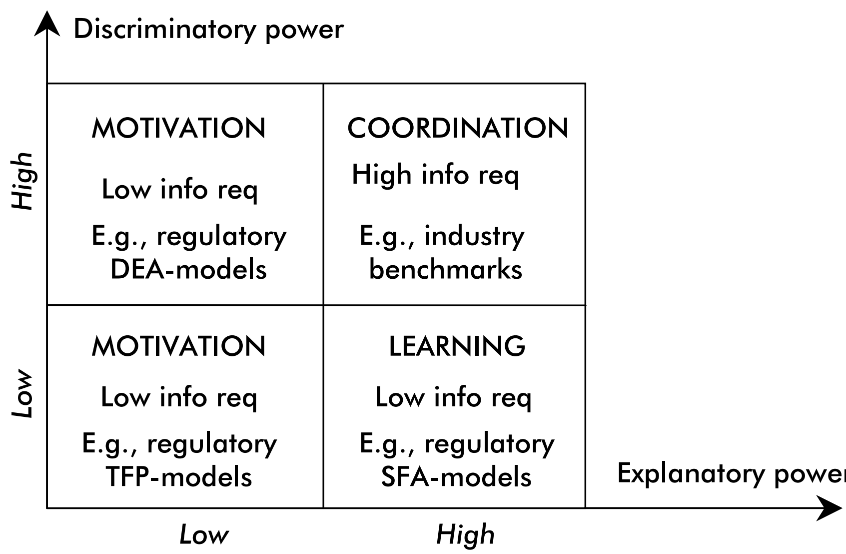


Figure 6.3 Classification of benchmarking models with respect to discriminatory and explanatory power.

Table 6.3 Variables for selected benchmarking models.

Model	Inputs	Outputs	Environmental factors
Dte 1[1] DTe (2000)	OpEx[2]	Delivered energy No of connections[3]	
Dte2 DTe (2000)	OpEx	Delivered energy No of connections	
Dte3 DTe (2000)	OpEx	Delivered energy No of connections HC No of connections LC Network size[4],[5] No of transformers[6]	
Dte4 DTe (2000)	OpEx	Delivered energy No of connections HC No of connections LC No of transformers Network length/customer	
Dte DTe (2000)	Total Operating Cost (TOC)[7]	Delivered energy No of connections Small No of connections Large Maximum demand LC Maximum demand HC	Network length
NordDEA model Edvardsen – Førsund (2001)	Operating cost Labor (kh) Capital (physical) Distribution losses (GWh)	Delivered energy No of connections Network size	
STEM SR-Agrell Bogetoft(2000)[8]	OpEx[9]	Delivered energy HC Delivered energy LC No of connections HC No of connections LC Maximum demand (MW)	Network length Transformers/installed capacity (MW) [Installed capacity MW[10]] Climate index[11]
STEM LR-Agrell Bogetoft(2000)[12]	OpEx Net losses (MWh) Capital (kkr)	Delivered energy HC Delivered energy LC No of connections HC No of connections LC Maximum demand (MW)	Network length[13] (GIS) Climate index[14]
NUTEK 1993[15]	Labor (kh) Network length HC Network length LC Transformer capacity (MVA)	Delivered energy HC Delivered energy LC No of connections HC No of connections LC Utilization time for assets Peak demand (MW)	
Hjalmarsson and	Labor (kh)	Delivered energy HC	

[1] Netherlands Electricity Regulatory Service
[2] OpEx (Operating Expenditure) = Variable costs = Materials + services+ staff + other costs
[3] Connections or invoiced customers, depending on available information.
[4] Proxy for customer density
[5] Interpretation for inputefficiency with given networks, not maximization of network size. The most important outputs are delivery quality, delivered energy and installation of capacity to meet peak demand.
[6] Proxy for network complexity
[7] Total operating cost = OpEx + depreciation on material assets, using a recalculated RAB.
[8] Short run benchmarking model for managerial costefficiency.
[9] Net of activated labor, depreciation and transmission fees.
[10] Not used in actual runs 1997, 2000 and 2001.
[11] Suppressed in 2001.
[12] Long run benchmarking model for technical efficiency.
[13] Defined as GIS proxy from Net Utility model, actual length used in runs 1997, 2000 and 2001.
[14] Suppressed in 2001.
[15] Variabler som hade en signifikant inverkan på produktionsresultatet

Veiderpass (1992)	Network length HC Network length LC Transformer capacity (MVA)	Delivered energy LC No of connections HC No of connections LC	
Hougaard (1994) (Four models)	Labor (kh) Operating cost (ex labor) Total Operating cost Net losses Capital	Delivered energy No of connections Network size	
Kittelsen (1994)	Labor (kh) Net losses (MWh) Transformers Lines (kkr) Material and services (kkr)	Delivered energy No of connections	Network size
Kittelsen – alt. modeller	Labor (kh) Net losses (MWh) Transformers Lines (kkr) Material and services (kkr)	Maximum demand(kW) Industrial demand Commercial demand Residential demand	Distance indicator[16] Corrosion index[17] Climate index
STEM 2, Ek (1998)	OpEx or TOC[18]	Delivered energy HC Delivered energy LC No of connections HC No of connections LC	Network HC (km) Network LC (km) Installed transformers (MVA)
NVE	Labor (fte) Net losses (MWh) Capital (kkr) (net assets) Material and services (kkr)V	Delivered energy No of connections	Network size
Sydkraft	Network HC (km) Network LC (km) Installed transformers (MVA) Net losses (MWh) OAMC[19]	Delivered energy HC Delivered energy LC No of connections HC No of connections LC	Lines LC per customer LC
Roos and Färe (Etapp 1)	Network HC (km) Network LC (km) Installed transformers (MVA) Net losses (MWh) OAMMC[20]	Delivered energy HC Delivered energy LC	Population density No of connections
ELTA (Workshop 22.5.2000)	Labor (fte) Network LC (km) Installed transformers (kVA)	Delivered energy No of connections	Road length
HKKK (Helsinki School of Economics and Business Administration) proposal for potential I and O	OpEx Investments	Delivered energy (weighted) Average interruption length	Average snow depth Share of forest No customers (weighted) Customer density Change in delivered energy
Net Utility Model. (STEM)	TOC	Netlength LC[21] Netlength HC Installed power	
Weyman-Johnes (1985)	Labor (fte)	No of connections	Network size Installed transformers (MVA) Del energy Max demand Customer density Share industrial customers

[16] Traveling time in minutes to regional center.
[17] scale 1,0 to 4,0
[18] Real annuity of replacement value. OpEX from annual reports and the capital costs are calculated as an annual annuity from the replacement value given a normalized depreciation period (30 years, real interest 4%). Replacement values calculated using the EBR component catalogue and the firms' asset registers.
[19] Costs for administration, operation and measurement excl fixed costs.
[20] Costs for administration, operation, measurement and asset maintenance.
[21] Calculated from GIS data using a minimal spanning tree algorithm.

7. New Regulatory Challenges

7.01 The Norwegian model is, from an international and methodological viewpoint, a fairly aggregate model with a clear focus on incentive provision in the current regulatory scheme. Its lack of exogenous variables and output differentiation limits its value for industry interpretation.

7.02 However, the distinction of regional versus local grids and the application of variable returns to scale imply limited data sets and a surprisingly low differentiation among firms. It has been suggested that this effect is indeed a sign of high data quality in a mature business, which should make the exception evidence of outliers. Whatever its reasons, we notice that already in an early catch-up phase, the motivational model has reached the limits of its potential. As we will outline in this Chapter, new challenges in the industry and the regulation will inevitably prompt for revisions of the current model, in one or the other direction.

7.03 This part is a benchmarking oriented connection to pre-project 5 on non-grid investments and incentive systems. The interaction with alternative energies and means to adjust energy demand raises a series of questions. How can a benchmarking model be conceived and updated to handle uncertainty in technology and output definition? How can a benchmarking model support trade-offs among multiple energies and multiple ways to save energy? How can consistency between multi-utility regulation be maintained? And so on. These are some of the issues we comment on here.

7.04 The general messages of pre-project 5 (FP5) are that 1) structural aspects are potentially much more important than the details of the reimbursement scheme and that 2) in a broader perspective, the focus should be more on social value optimization rather than partial efficiency improvement in a single energy line.

Structural elements

7.05 In the previous chapter, we have already indicated how the structural efficiency of the energy sectors can be analyses using recent advances in productivity analysis, cf. e.g. Bogetoft(2002), Bogetoft, and Wang(1999), Bogetoft, Strange and Thorsen(2001), Brännlund, Färe and Grosskopf(1995) and Brännlund, Chung, Färe and Grosskopf(1998).

7.06 From the point of view of future benchmarking exercises, such structural analyses require a basic production model that is reasonable rich in the input and output descriptions. Only in this way can the full potential from cooperation be identified.

Output and value focus

7.07 Despite of the variety of international benchmarking models, cf. the discussion in the previous chapter, it is fair to say that in the broader perspective, they are all very similar. Traditional performance evaluation models have several variables that essentially are process oriented and concern the inputs or means rather than ends. They use for example km lines, number of transformers etc. in addition to some output-based measures like costumer characteristics or peak capacity offered. The Norwegian approach is no exception in this regard.

7.08 Another approach which is to focus on the true outputs generated, namely the values to end-users, rather than the ways resources are spend. This means that the regulation should stress product attributes and final values, say heating, easy use of household equipment, flexible powering of industrial machinery etc. and neither processes nor ex ante boundaries among the energies. A true output or value based approach is theoretically more appropriate although in practice also much more challenging.

7.09 The philosophy of the Swedish net-utility model discussed in the previous chapter is precisely that companies should be rewarded for the values that they generate, not the ways by which they generate them.

7.10 By stressing the end-values rather than the means of providing them, the benchmarking can also be used to compare companies supplying different types of energies. We may compare for example residual production and distribution of electricity for heating with central heating systems distributing via hot water.

7.11 The resulting challenge in future benchmarking studies is of course to find appropriate proxies for end-values and to included these in a possible multi-energy benchmarking exercise.

Better trade-offs and more comparisons

7.12 In general, a value model will improve the incentives of the companies to make proper *intra-firm trade-offs* among their means of offering energy. Inefficiency is a multi-utility framework, for example, may now reflect inadequate use of a given technology as well as inappropriate trade-offs among alternative technologies.

7.13 Also a value approach will support *inter-firms trade-offs*. The regulator will have a more appropriate – and larger, cf. below, basis for prioritizing among different energies if the values they generated are included in the same model. The same reasoning applies to quality regulation, where industry specific valuations (such as in the KILE approach) may inform also new technologies, such as distributed generation and demand-side management projects.

7.14 A final advantage of the end-value approach and hereby the possibility to make cross energy comparisons is also that it may increase the number of units to compare. This is advantageous to limit rents – and it is particularly useful when the exiting companies tend to become more and more similar so as to reduce the information that can be extracted.

Disciplining self-reporting

7.15 In FP5 – and FP2 – we have stressed also the possible use of a *Charter of Accountability* approach where the single or multiple-utility firm must account for its system wide trade-offs, cf. e.g. Agrell and Bogetoft(2002a). In such an approach, a company designing its net to reduce the risk of outages, for example, should documents this to the regulator.

7.16 In general, self-reporting is a useful way to make use of the fact that agents have relevant private information. There is a large literature on agency models with communication that make this point, cf. for example the survey in Bogetoft(1994a). However, self-reporting without any means of disciplining the sender usually degenerates to so called *cheap talk* – or at least the reports must be heavily underutilized. In such a context, benchmarking may get an new role. Its purpose becomes one of disciplining self-reporting, rather than to make final evaluations.

7.17 From the point of view of future benchmarking, this perspective would require an examination of alternative ways to discipline self-reporting. To illustrate, a benchmarking approach that with just a small probability generated perfect information will extremely useful (cf. 5.04). Also, it is well-known for example that information that is correlated with the underlying private information will be useful, even if the correlation is imperfect. One can image therefore that the disciplining benchmarking may be very different from the present ones typically used in regulation. The benchmarking may not even model energy-production as such – but just some aspect correlated with energy-production.

Motivation for whom?

7.18 The implicit tradeoff between explicative and discriminatory power in the model design reveals that the motivation objective has been emphasized in the current regime. However, in a dynamic perspective, two issues remain to be solved to achieve full effect of the regulation.

7.19 First, the unexpectedly rapid convergence of the efficiency results to a high proportion of efficient firms is a sign a rapid productivity catch-up, which is good. On the other hand, it also means that the individual motivation effect of the benchmarking is likely to decrease once full efficiency has been achieved, especially under an irregular ex ante regime with relatively slow feedback.

7.20 Second, the majority of the firms are still publicly owned, although some concentration into private ownership has been noticed lately. Theoretically as well as empirically, there is no reason to believe that the firms' managers react differently to profit incentives. For private firms with effective decentralization of internal incentive mechanisms, the mere application of a high-powered regime would suffice to induce efficiency. For firms with passive owners and strong internal management culture, motivation in terms of a more informative and easily accessible benchmarking may provide stronger incentives. The managers-engineers can communicate their performance to their customers and also acquire private rents in the employment market if the model is perceived as a credible indication of *managerial* performance.

8. Further Work

8.01 In this Chapter, we sketch some further projects that may provide NVE and OED with necessary information to make the strategic decisions discussed. Overall, the suggested projects follow the conclusions in the previous chapter, i.e., regulatory continuity and rigorous instrumental and mechanism analysis.

8.02 The current situation for the NVE benchmarking is characterized by a well-working data collection process, fairly low-powered integration into the regime, high aggregation and surprisingly low discrimination. Compare some other "peer" models in Figure 8.1 below. Our suggestions focus around studies that could improve either the explanatory value of the model (going right) or the discriminatory power of the model (going up). As discussed in the preceding Chapter, the position on this question should be coordinated with strategic studies on the regime itself (FP2) and the new sector (FP5).

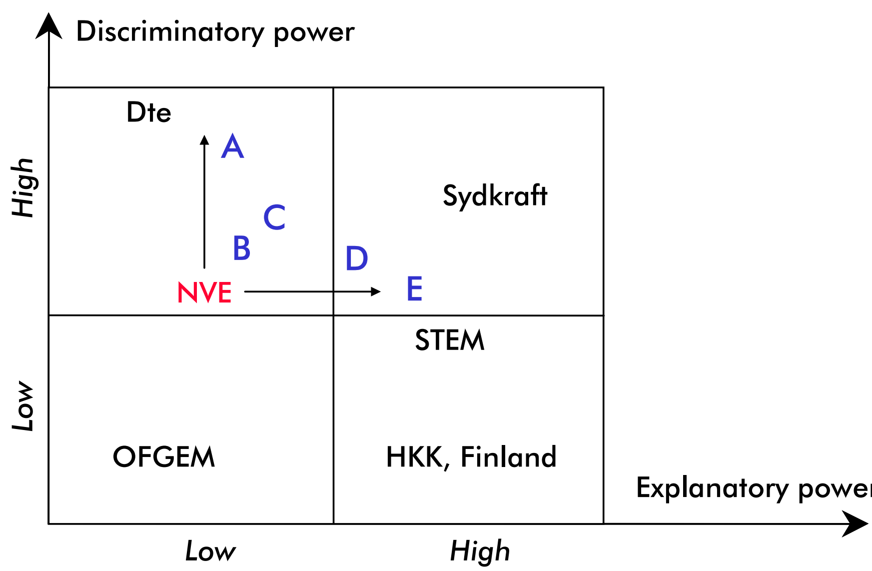


Figure 8.1 NVE benchmarking, quo vadis?

8.03 In the following five proposals, we outline three projects that primarily support the *motivation-incentive* property of the current benchmarking instrument. The projects A, B, C correspond to critical issues related to capital valuation, model specification, discriminatory ability, model scope and implementation feasibility. We also give two suggestions for the *informative-learning* property of the regime. Projects D and E aim at providing the regulator and the industry with stable and viable analyses on the structure, sources and scope of the alleged sector inefficiency.

A. Hybrid Approach with Technical Models

8.04 Projects FP5 and FP2 have highlighted the need to coordinate regulation among substituting goods and to leave the field to innovation in the industry. As technology becomes more volatile and the need for pre-commitment on technology less, the benchmarking models need to follow the evolution towards more aggregate and customer-close measures. One innovative way to assure this in the future would be to use hybrid measures with engineering models, where e.g. the final energy delivered to the consumer is the only output and the total resources are measured at final level. We propose a pre-study, drawing on our integration work between the net utility model in Sweden and the DEA models, to investigate the potential advantages and risks with such approach. The outcome of the study is a recommendation for further development and a suggested timeplan for possible implementation.

B. Dynamic ex-post adaptation

8.05 As pointed out in FP2 (subproject A) and above, the NVE regulation and performance assessment model is well adapted for a transition towards an ex-post settlement. Such a change may limit the regulatory involvement in e.g. investment review and output management, while leaving decisions to the best informed party. However, it is primordial to assure a smooth transition in order to signal regulatory continuity and long-term interests. In this project, we propose to guide, date and inform the transition process by providing detailed analysis on crucial policy issues and strategic alternatives for successful implementation. This project involves a methodological part (that might be amended by NVE as a white paper), a managerial brief on transition mile stones and an applied part providing evidence of similar projects.

C. Quality regulation and benchmarking

8.06 Service quality, development and innovation are key concepts for the energy sector in the coming years. When reviewing the regulation regime, it is important to consolidate incentive plans towards these dimensions, rather than risking to distort service valuations. This is particularly relevant in the future scenario of multi-utilities (cf. FP5), where inter-firm tradeoffs should be facing the optimal signals to guide decision making. The current inclusion of the KILE cost in the NVE benchmarking model is one approach to deal with service quality. In this project, we would like to explore this and other models from the perspectives of short-run and long-run incentives under mono- and multi-utility assumptions. The proposed project, that may be combined with the regulatory reform project FP2:E, is providing both constructive advice on incremental improvements and sharp analysis of areas of crucial future interest.

D. Inefficiency, ownership and data gathering

8.07 The Norwegian distributors are among the smallest in Europe and the ownership structure is very homogenous, whereas the operating conditions are very heterogeneous. Together this poses a major challenge to NVE in the future, knowing if and how to affect industrial structure and the effectiveness of the learning-motivation orientation. Are public firms more or less prone to motivation or to public benchmarking? How to assure sufficient data material in the future?

8.08 Using detailed Swedish and Danish data, we propose an in-depth study of the structural sources of inefficiency, especially ownership, aiming at strategic advice for the revision of the benchmarking method and the regulation regime.

E. Optimal scale and industrial structure

8.09 The concept of scale efficiency in DEA is flawed and not very useful. Yet the Norwegian distribution sector faces a challenge in industrial restructuring in the near future. NVE may preempt problems in the industry's merger policy by undertaking a data and methodological study on optimal scale and size in Norway, using primarily Swedish data, and an estimation of the potentials for efficiency catch-up gains, size effects and harmony effects. Using results in Bogetoft and Wang (1999), we propose a study to clarify this important dimension for the Norwegian electricity sector and the future revision of the model.

References

Agrell, P. and P. Bogetoft, (2001) *On Dynamic Regulation of Natural Monopolies Using Frontier Analysis*, invited talk at CREPP, Université de Liège, 2001

Agrell, P. and P. Bogetoft, (2002) *Ekonomisk Nätbesiktning 2000, Report* ER 7:2002, Swedish Energy Agency (in Swedish).

Agrell, P. J and P. Bogetoft (2000), Ekonomisk Nätbesiktning. Final Report commissioned by STEM. SUMICSID AB. (In Swedish)

Agrell, P. J, P. Bogetoft, and J. Tind (2000b), Multi-period DEA Incentive Regulation in Electricity Distribution. Submitted to Journal of Productivity Analysis.

Agrell, P. and J. Tind, (2001) A Dual Approach to Nonconvex Frontier Models *Journal of Productivity Analysis*, vol. 16, pp. 129-147.

Alexander, I. and Estache A. (2000) Infrastructure Restructuring and Regulation: Building a Base for Sustainable Growth Working Paper No.: 2415 World bank

Averch, H. and L. L. Johnson (1962), Behavior of the Firm under Regulatory Constraint. American Economic Review 52, 1052-1069.

Banker, R.D., *Hypothesis Test Using Data Envelopment Analysis*, Journal of Productivity Analysis, 7, pp. 139-159, 1996.

Bernstein JI, Sappington DEM (1999), Setting the X factor in price-cap regulation plans, J REGUL ECON 16 (1): 5-25

Bernstein JI, Sappington DEM (2000), How to determine the X in RPI-X regulation: a user's guide TELECOMMUN POLICY, 24 (1): 63-68

Bogeoft, P. and K. Nielsen(2002), *Internet Based Benchmarking*, Working Paper, KVL.

Bogetoft, P, Quota Trading among Inefficient Firms, WP, KVL, 2002.

Bogetoft, P. (1990), Strategic Responses to DEA Control, Working Paper, CBS.

Bogetoft, P. (1994a), Non-Cooperative Planning Theory, Springer-Verlag.

Bogetoft, P. (1994b), Incentive Efficient Production Frontiers: An Agency Perspective on DEA, Management Science, 40, pp.959-968.

Bogetoft, P. (1995), Incentives and Productivity Measurements, International Journal of Production Economics, 39, pp. 67-81.

Bogetoft, P. (2000), DEA and Activity Planning under Asymmetric Information, *Journal of Productivity Analysis,* 13, pp. 7-48.

Bogetoft, P. (2000), DEA and Activity Planning under Asymmetric Information, Journal of Productivity Analysis, 13, pp. 7-48.

Bogetoft, P. and D. Wang (1999), Estimating the Potential Gains from Mergers, Working Paper, KVL.

Bogetoft, P. and D. Wang, *Estimating the Potential Gains from Merger,* Working Paper, Royal Agricultural University, pp. 1-24, 1999

Bogetoft, P. and J.L.Hougaard (2002), Rational Inefficiency, *Journal of Productivity Analysis*, 2002.

Bogetoft, P. J. Tama, J.Tind, *Convex Input and Output Projections of Nonconvex Production Possibility Sets*, Management Science, vol. 46(6), pp. 858-869, 2000.

Bogetoft, P., (1997), DEA-Based Yardstick Competition: The Optimality of Best Practice Regulation, Annals of Operations Research, 73, pp. 277-298.

Bogetoft, P., *DEA on Relaxed Convexity Assumptions*, Management Science, 42, 1996, pp. 457-465.

Bogetoft, P., N. Strange and B.J.Thorsen, *Efficiency and Merger Gains in The Danish Forestry Extension Service,* Forestry Science to appear,2001.

Borger, B. and K. Kerstens, *Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches*, Regional Science and Urban Economics, 26, 1996, pp. 145-170.

Boyer, M., A. Jacques, and M. Moreaux (1998) *Observability, Commitment and Flexibility*. Working Paper 98-36, CIRANO, Montreal.

Brännlund, R., R. Färe and S. Grosskopf, *Environmental Regulation and Profitability: An Application to Swedish Pulp and Paper Mills,* Environmental and Resource Economics, 6, pp. 23-36, 1995.

Brännlund, R., Y. Chung, R. Färe and S. Grosskopf, *Emission Trading And Profitability: The Swedish Pulp and Paper Industry*, Environmental and Resource Economics, 12, pp. 345-356, 1998.

CEER Council of European Energy Regulators (2001), Quality of Electricity Supply: Initial Benchmarking on Actual Levels, Standards and Regulatory Strategies, Autorità per l'energia elttrica e il gas.

Charnes, A., W. Cooper, A.Y. Lewin and L.M. Seiford (1994), Data Envelopment Analysis: Theory, Methodology and Application, Kluwer Academic Publishers.

Charnes, A., W.W. Cooper and E. Rhodes (1978) Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 2, pp. 429-444.

Charnes, A., W.W. Cooper and E. Rhodes (1979), Short Communication: Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 3, pp. 339.

Deprins, D., D. Simar, and H. Tulkens, *Measuring Labor Efficiency in Post Offices*, pp. 243-267 in M. Marchand, P. Pestieau, and H. Tulkens, "The Performance of Public Enterprises: Concepts and Measurements", North Holland, 1984.

ECON (1998) Statnett efficiency ranking - rapport nr. 55/98

ECON (2002) Nettregulering 2007: Inntektskontroll. Notat 71-02.

Edin, K.-A. and H. Svahn (1998) Reglering av tariffer för elnät -- Sporre eller hinder för rationaliseringar ? Politik & Samhälle, Stockholm. (In Swedish)

Edvardsen, D. F., and F. R. Førsund (2001), *International Benchmarking of Distribution Utilities*. Working Paper, Frisch Centre, Oslo University.

Electricity Act (1997) Rixlex (Swedish). Unofficial translation at www.stem.se.

Estache and Martimort (1998) *Transaction Costs, Politics, Regulatory Institutions, and Regulatory Outcomes,* EDI Regulatory Reform Discussion Paper, World Bank.

European Commission (1996) Directive 96/92/EC.

Farrell, M.J. (1957), The measurement of productive efficiency, Journal of the Royal Statistical Society, Series A, III, pp. 253-290.

Farrier Swier (2002) *Comparison of Building Blocks and Index-Based Approaches.* Utility Regulators Forum, Australia.

Førsund, F. and D. Edvardsen (2001) NorDEA – Efficiency Analysis of Nordic Electricity Distributors. Frisch Institute, Research Report. Forthcoming.

Grasto, K. (1997) Incentive-based Regulation of Electricity Monopolies in Norway - Background, Principles and Directives, Implementation and Control System. Publication 23/1997, Norweigian Water Resources and Energy Administration, POB 5091, 0301 Oslo, Norway.

Jamasb, T. and Pollitt, M. (2000) Benchmarking and Regulation of Electric Distribution and Transmission Utilities: Lessons from International Experience. Workshop on International Benchmarking of Electric Utilities, Robinson College, Cambridge, December.

Jamasb, T. and Pollitt, M. (2001) International Benchmarking and Yardstick Regulation: An Application to European Electricity Utilities. DAE Working Paper 0115, Dept of Applied Economics, University of Cambridge.

Joro, T., P. Korhonen, and J. Wallenius (1998), Structural Comparison of Data Envelopment Analysis and Multiple Objective Linear Programming, Management Science, 44, pp. 962-970.

Kittelsen, S. A. C. (1993) *Stepwise DEA; Choosing Variables for Measuring Technical Efficiency in Norwegian Electricity Distribution*, SNF-arbeidsnotat nr. A 55/93, SNF, Oslo.

Kittelsen, S. A. C. (1994) *Effektivitet og Regulering i Norsk Elektrisitetsdistribusjon*, SNF-rapport 3/94, SNF, Oslo.

Kittelsen, S. A. C. (1996) *DEA for NVE - Et Måleverktøy for Effektivitet i Elforsyningen*, SNF-rapport 85/96, SNF, Oslo.

Kittelsen, S. A. C. and Torgersen, A. M. (1993) *Teknisk Effektivitet i Norske Elektrisitetsfordelningsverk*, SNF-arbeidsnotat nr. A 27/93, SNF, Oslo.

Laffont, Jean-Jacques and Meleu, M (1997) Reciprocal Supervision, Collusion and Organizational Design. Scandinavian Journal of Economics, 99, 519-540.

Laffont, Jean-Jacques and Tirole, Jean (1986) Using Cost Observations to Regulate Firms. Journal of Political Economy, 94, 614-641.

Laffont, Jean-Jacques and Tirole, Jean (1987) Auctioning Incentive Contracts. Journal of Political Economy, 95(5), 921-937.

Laffont, Jean-Jacques and Tirole, Jean (1988) Repeated Auctions of Incentive Contracts, Investment, and Bidding Parity with an Application to Takeovers. RAND Journal of Economics, 19(4), 516-537.

Laffont, Jean-Jacques and Tirole, Jean (1990) Optimal Bypass and Cream Skimming. American Economic Review, 80, 1042-1061.

Laffont, Jean-Jacques and Tirole, Jean (1991) The Politics of Government Decision Making: A Theory of Regulatory Capture. Quarterly Journal of Economics, 107, 1089-1127.

Laffont, Jean-Jacques and Tirole, Jean (1993), A Theory of Incentives in Procurement and Regulation. MIT Press.

Larsson, M.B-O (1999) *Nätnyttomodellen*, Konsultrapport, Energimyndigheten (STEM).

Lewis, T. and M. Poitevin (1995) *Disclosure of Information in Regulatory Proceedings*. Working Paper 95-01, CIRANO, Montreal.

NVE (1997a) Retningslinjer for Inntektsrammen for Overføringstariffene. Report NVE, Norweigian Water Resources and Energy Administration, POB 5091, 0301 Oslo, Norway.

NVE (1997b) Benchmark. Publication 27/1997, Norwegian Water Resources and Energy Administration, POB 5091, 0301 Oslo, Norway.

NVE (2001) Inngangsverdier til analysene av kostnadseffektivitet 1996 – 1999. Note EMØ 04/2001, Norwegian Water Resources and Energy Administration, POB 5091, 0301 Oslo, Norway.

Olesen, O. B. and N. C. Petersen (1995). Chance constrained efficiency evaluation. Management Science 41(3): 442-457.

Petersen, N.C., *Data Envelopment Analysis on a Relaxed Set of Assumptions*, Management Science, 36, 1990, pp. 305-214.

Pollitt, M. G. (1995) Ownership and Performance in Electric Utilities: The International Evidence on Privatization and Efficiency. Oxford University Press, Oxford.

Resende M (2002), Relative efficiency measurement and prospects for yardstick competition in Brazilian electricity distribution, Energy Policy 30 (8): 637-647

Riechmann C (2000), Strategic pricing of grid access under partial price-caps - electricity distribution in England and Wales, Energy Economics 22 (2): 187-207.

Schleifer, A. (1985). A Theory of Yardstick Competition. Rand Journal of Economics, 16: 319-327.

Swedish Electricity Act (1992). Rixlex, Stockholm.

Swedish Electricity Act (1997) Amendments until SFS 1999:770. Rixlex, Stockholm.

Tulkens, H., *On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit*, The Journal of Productivity Analysis, 4, 1993, pp. 183-210.

Vogelsang I (2002), Incentive regulation and competition in public utility markets: A 20-year perspective, J REGUL ECON 22 (1): 5-27

von der Fehr, N.H., K.P.Hagen and E.Hope (2002), Nettregulering, SNF-rapport 1/2002, Stiftelsen for samfunns-og næringslivsforskning.

Weyman-Jones, T. G. (1995) Problems of Yardstick Regulation in Electricity Distribution. In Bishop, M., Kay, J., Mayer, C. (Eds) The Regulatory Challenge. Oxford University Press, Oxford.

SUMICSID

SUMICSID AB
Tunbyn 502 Tel: +46 60 56 51 41
S-855 90 Sundsvall, SWEDEN Fax: +46 60 56 51 41

sweden @ sumicsid.com

SUMICSID (BELGIUM)
Avenue des Tourterelles 37 Tel: +32 10 47 83 65
B-1150 Bruxelles, BELGIUM Fax: +32 10 47 83 24

belgium @ sumicsid.com

SUMICSID (DENMARK)
Fru Ingesvej 19 Tel +45 57 83 15 18
DK-4180 Sorø, DENMARK Fax: +45 57 83 15 18

denmark @ sumicsid.com

http://www.sumicsid.com