2013/8

■

Benchmarking and regulation

Per J. Agrell and Peter Bogetoft

CORE

DISCUSSION PAPER

CORE DISCUSSION PAPER
2013/8

Benchmarking and regulation

Per J. AGRELL [1] and Peter BOGETOFT[2]

January 2013

Abstract

Benchmarking methods, and in particular Data Envelopment Analysis (DEA), have become well-established and informative tools for economic regulation. DEA is now routinely used by European regulators to set reasonable revenue caps for energy transmission and distribution system operators. The application of benchmarking in regulation, however, requires specific steps in terms of data validation, model specification and outlier detection that are not systematically documented in open publications, leading to discussions about regulatory stability and economic feasibility of these techniques. In this paper, we review the modern foundations for frontier-based regulation and we discuss its actual use in several jurisdictions.

Keywords: agency theory, regulation, energy networks.

JEL classification: Q40, L59, C51, C24

---

[1] Université catholique de Louvain, CORE and Louvain School of Management, B-1348 Louvain-la-Neuve, Belgium. E-mail: per.agrell@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.
[2] Department of Economics, Copenhagen Business School, DK-2000 Frederiksberg, Denmark; Yale School of Management, New Haven, CT 06511, USA. E-mail: pb.eco@cbs.dk

1. Introduction

One of the more prominent applications of state-of-the-art benchmarking is in the regulation of natural monopolies in general and electricity and gas networks, in particular. Benchmarking studies applied to inform such regulation has considerable economic impact on firms and consumers alike.

Large infrastructure industries like the networks to distribute electricity and gas, commonly referred to as Distribution System Operators DSOs, are characterized by considerable fixed cost and relatively low marginal costs. They therefore constitute natural monopolies and indeed network companies are generally given licenses to operate as legal monopolies.

Monopolies have limited incentives to reduce costs, and will tend to under-produce and overcharge the services provided since they are not subject to the disciplining force of the market. For electricity distribution, the monopoly characteristic is accentuated by the fact that there are no close substitutes for the offered services and that demand is relatively inelastic.

Most countries therefore empower regulators to act as a proxy purchaser of the services, imposing constraints on the prices and the modalities of the production. The regulator is usually affiliated with the national competition authority. One of the instruments used in the regulation is benchmarking, which is facilitated by the existence of different networks covering different areas that can be compared or, in some cases, by international comparisons of such firms.

Regulation economics was long considered a fairly uninteresting application of industrial organization. Early regulatory theory largely ignored incentive and information issues, drawing heavily on conventional wisdom and industry studies. This kind of institutional regulatory economics was challenged in the seventies with economists such as Friedman, Baumol, Demsetz and Williamson questioning the organization and succession of natural monopolies. However, the main breakthrough came in the late eighties with the introduction of information economics and agency theory. An authoritative reading in the area is Laffont and Tirole (1993a). Littlechild (1983) suggested a relatively simple yet high powered revenue or price-cap regime, while the idea of yardstick competition goes back to Lazear and Rosen (1981), Nalebuff and Stiglitz (1983) (1983) and Shleifer (1985) who show conditions for the implementation of first-best solutions for correlated states of nature. The results carry over, even for imperfectly correlated states of nature Tirole (1988), and as further analyzed using DEA in Bogetoft (1997) and related publications that we shall discuss in the Section 5. Hence, the comparators do not have to be identical, but the relative difference in the exogenous operating conditions has to be known or estimated, and benchmarking can obviously be helpful here. One way to think of modern regulation is as model based pseudo competition – the firms do not compete on the market but they compete via a benchmarking

model. An alternative to this is to introduce auction based competition for the market, i.e. competition as to which firm shall serve the market. Franchise auctions were discussed early by Demsetz (1968) and Laffont and Tirole (1993b), and again we shall discuss some DEA based extensions in Section 5. Key references to the practical combination of benchmarking and regulation are Agrell and Bogetoft (2001), Agrell and Bogetoft (2010) and Coelli et al. (2003).

In this paper, we first describe some classical regulatory packages and explain the role of benchmarking in these regimes. Next, we illustrate some of the models that have been developed in a selection of countries. We finally provides a summary of the theoretical literature linking DEA and incentives.

2. Classical regulatory packages

As explained above, modern economic theory views the regulatory problem as a game between a principal (the regulator) and a number of agents (the regulated firms). The regulation problem is basically one of controlling firms that have superior information about their technology and their cost reducing efforts as compared to the regulator. The availability and access to information is a key issue in the regulatory game and regulators can use benchmarking to undermine the informational asymmetry.

The regulatory toolbox contains numerous more or less ingenious solutions to the regulator's problem. To illustrate, we will distinguish four approaches

- Cost-recovery regimes (cost of service, cost-plus, rate of return),

- Fixed price (revenue) regimes (price-cap, revenue cap, RPI-X),

- Yardstick regimes, and

- Franchise auction regimes.

We will provide a brief introduction to these regimes below.

2.1. Cost-recovery regimes

Taking for granted the cost information supplied by the agents, the regulator may choose to fully reimburse the reported costs, often padded with some fixed mark-up factor. To illustrate, the reimbursement in a given period t for firm k may be determined as

$$R^k(t) = C^k_{OpEx}(t) + D^k(t) + (r + \delta)K^k(t)$$

where C^k_{OpEx} is the operating expenses , D^k is the depreciation reflecting capital usage, r is the interest rate reflecting the credit costs of investments with similar risks, δ is a mark-up, and K^k is the total investment, the capital or rate base.

3

Unless subject to costly information verification, a cost recovery approach results in poor performance. Firms have incentives to over-invest in capital and have no incentives to reduce operating expenditures since it just lowers revenue.

In reality, such schemes have therefore involved considerable regulatory administration in an attempt to avoid imprudent or unreasonable operating expenditures and investments to enter the compensation and rate base. As part of the regulatory effort, some benchmarking approaches have been used. However, even with large investments in information gathering, the information asymmetry and the burden of proof in this regime rest on the regulator, and there are reasons to doubt their ability to induce efficiency.

Cost recovery is often organized as negotiation and consultation based regimes. Whether rate reviews are initiated by complaints or are planned, reviews are often done as individual consultations. In contrast to the methods below, where a joint framework is used to evaluate all DSOs, the consultations are typically case-specific and they rely more on negotiations than on a comprehensive model estimation for the entire sector.

An idea is to combine negotiations with systematic investigations and benchmarking in such a way as to limit the negotiation space. In this way, the negotiations become more structured. Such restrained negotiations have been proposed in the Netherlands for the regulation of hospitals, cf. Agrell et al. (2007). The idea is that the regulator uses benchmarking to constrain the acceptable outcomes but leaves negotiations to industry partners, say hospitals and insurance companies.

2.2. Fixed price regimes (price-cap, revenue cap, CPI-X)

In response to the problems of the cost-recovery regime, several countries have moved to more high-powered regimes. These regimes typically allow the regulated firms to retain any realized efficiency gains.

In the price-cap regime, the regulator caps the allowable price or revenue for each firm for a pre-determined regulatory period, typically 4-5 years. The price or revenue cap model is usually quite simple, involving a predicted productivity development per year x plus, perhaps, individual requirements on DSOs, x^k, to reflect the level of historical costs and thereby the need to catch-up to best practice. The resulting allowed development in the revenue for DSO k is then

$$R^k(t) = C^k(0)(1 - x - x^k)^t, \qquad t = 1, \dots, T$$

where $R^k(t)$ is the allowed revenue in period t and $C^k(0)$ is the cost of DSO k in period 0. Note that x is used here not as input but as an efficiency requirement; this is in accordance with the standards in regulations where the above model is often referred to as CPI-x to reflect that there are adjustments for price developments and productivity requirements.

4

There are, of course, many modifications to this model. Thus, there will typically be adjustments for changes in the volume supplied and for general changes in the cost level due to inflation. We will show an example from Germany below.

The crucial feature of the fixed price regime is that there is a fixed, performance independent, payment. This means that, to maximize profit, the DSO will minimize costs. This is key to the incentive provision.

Another important feature is the fixation of payments during a regulatory period and the consequent regulatory lag in updating productivity development. The last feature is often emphasized by calling such schemes *ex ante regulation* as illustrated in Figure 1 below. Before a regulatory period starts, the regulator uses historical data from a review period to estimate x and x^k, and then commits to these values for the regulatory period of T years. At the end of the regulatory period, new estimations of x and x^k are made to set the revenue conditions for the next regulatory period.
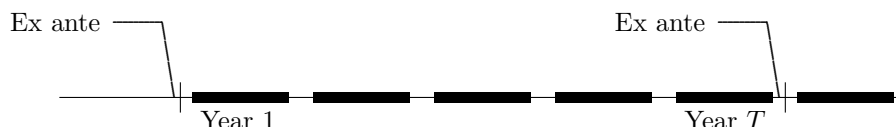

Figure 1: Ex ante regulation

The idea of price or revenue fixation is simple but in practice the cap is regularly reset, in hindsight, to reflect the realized profits in the previous period. This limits the efficiency incentives. Also, the initial caps have to strike a careful balance between informational rents, incentives for restructuring and the bankruptcy risks. Further, the price or revenue cap is usually linked to the consumer price index (CPI) or the retail price index (RPI) as a measure of inflation. Therefore, in spite of its conceptual simplicity, the challenges of fixing the initial caps, the periodicity of review and the determination of the X-factor make this regulation a non-trivial exercise for the regulator. In particular, since initial windfall profits are retained by the industry and dynamic risks are passed on to consumers, there is a potential risk of regulatory capture by consumer or industry organizations.

For now, however, the most important feature is that the price fixation regimes generally involve some systematic benchmarking exercise, often based on DEA and SFA, to guide the choice of individual requirements x^k and the general requirement x.

The general requirement x is often set by using a Malmquist-like analysis of productivity developments over the years prior to the regulatory period. Thus, if the analysis of past frontier shifts suggests that even the best are able to reduce

costs by 2 % per year, the regulator has a strong case to set x close to 2%.

Individual requirements x^k are typically linked to the individual efficiencies of the DSOs in the last period prior to the regulatory period. There are no general rules used by regulators to transform a Farrell efficiency E^k to an individual requirement x^k, except that the smaller E^k is, the larger x^k is. Some countries require the DSOs to catch-up very quickly. In the first Danish regulation of electricity networks, for example, the electricity producers were required to eliminate the inefficiency in just 1 year. Others, like the Netherlands, used one regulatory period of 3-5 years. Germany aims to have eliminated the individual efficiency differences in two periods, i.e., 10 years, while Norway, a pioneer in the use of incentive-based regulation, allowed for an even longer period of time in the initial implementation of a revenue cap system. It is clear that the analyses of historical catch-up values can guide this decision, but there is also a considerable element of negotiation in the rules that are applied. Moreover, it is difficult to compare these requirements across countries. A cautiousness principle would suggest that the requirements will depend on the quality of data and the benchmarking model. Also, a controllability principle would suggest that it should depend on the elements that are benchmarked. In particular, it is important if it is Opex (operating expenses) or Totex (= Opex+Capex) that are being benchmarked and that become subject to efficiency improvement requirements.

In Denmark, for example, the first model from 2000 had very rigorous requirements on Opex - but still allowed new capital evaluations (opening statements), which lead to increased Capex allowances. On average, the companies only used 80-85% of the revenue caps. This suggests that the regulation may not have been as demanding as it looked with immediate catch-up requirement in a linear model. Also, it seems that the importance of consumer preferences in the many cooperatively-owned distribution companies was not foreseen. Either way, this led to immense accumulated reserves by the end of 2003. In return, this meant that adjustments in the regulation could have only limited impact since the DSOs could always draw on past revenue cap reserves. The regulation was, therefore, abandoned at the end of 2003 and a new regulation was later established.

We will give some more detailed illustrations of some of the steps in regulatory benchmarking for revenue cap regulation in section 3 below, where we discuss the recently developed German benchmarking model.

2.3. Yardstick regimes

The idea behind yardstick regimes is to mimic the market as closely as possible by using real observations to estimate the real cost function in each period rather than relying on ex ante predicted cost developments.

Thus, for example, in its simplest form, the allowed revenue for DSO k in period t would be set ex post and determined by the costs in the same period of

other firms $h = 1, \ldots, k-1, k+1, \ldots, K$ operating under similar conditions

$$R^k(t) = \frac{1}{K-1} \sum_{h \neq k} C^h(t), \qquad t = 1, 2, \ldots$$

Observe that this is the revenue the firm could charge on average in a competitive environment.

Also, one can argue that the average is just one of many ways to aggregate the performance of the other firms. One alternative is to use best practice realized performance, i.e.,

$$R^k(t) = \min\{C^h(t) \mid h = 1, \ldots, k-1, k+1, \ldots, K\}, \qquad t = 1, 2, \ldots$$

Of course, if the DSOs are delivering different services under different contextual constraints, the above revenue cap formed as a simple average of the costs in the other firms, is not directly applicable. Instead, we use benchmarking to account for these differences.

The yardstick regime is attractive in the sense that the revenue of a given DSO is not determined by its own cost but by the performance of the other DSOs. This fixed price feature makes the firm a residual claimant, as in the price fixation regime, and this is the key incentive property.

Another advantage of yardstick competition is that the productivity development is observed rather than predicted. This provides insurance for the DSOs and at the same time it limits their information rents. This is accomplished by setting the revenue *ex-post*, i.e., after each period. This is illustrated in Figure 2. The allowed revenue in period t is only set after period t. Exogenous and dynamic risks will directly affect the costs in the industry, lifting the yardstick. Innovation and technical progress will tend to lower the yardstick. Thus, the regime endogenizes the ubiquitous x factor and caps the regulatory discretion at the same time.
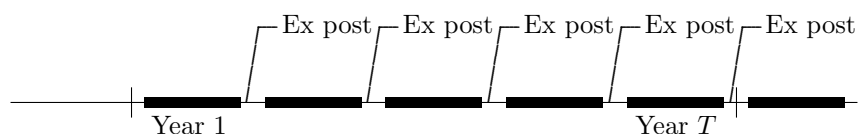

Figure 2: Ex post regulation

Despite its theoretical merits, the pure approach of only considering the observed cost in each period is linked to some risks in implementation. First, a set of comparators with correlated operating conditions must be established. Second, if the comparators are few and under similar regulation, there is risk of collusion.

Finally, a yardstick system that is not preceded by a transient period of asset revaluation or franchise bidding will face problems with sunk costs and possibly bankruptcy. A crucial question, in terms of yardsticks in electricity distribution, is, therefore, how to preserve the competitive properties while assuring universal and continuous service.

In section 5 below we will expand on the advantages of the yardstick idea and we will show how to cope with cases of imperfectly correlated costs and variations in output levels and mix by using DEA.

From the point of view of benchmarking, the yardstick regime requires the same model types as price fixation regimes, only now benchmarking has to take place more often, typically annually. A DEA-based yardstick scheme was introduced in Norway 2007 and will be discussed later. Also, the Dutch regulation of electricity DSOs has yardstick features.

2.4. Franchise auctions

A fourth approach to regulation is to substitute pseudo competition on the market with real competition for the market. The idea is to award delivery rights and obligations based on an auction among qualified bidders. Thus, for example, we could assign the distribution task to the bidder demanding the least. As an alternative, we could pay the winning bidder the lowest losing bid.

To formalize the latter, let each of K bidders for a project demand $B^h, h = 1, \ldots, K$. Agent k, therefore, is a winner if

$$B^k = \min\{B^1, B^2, \ldots, B^K\}$$

and we would compensate him

$$R^k = \min\{B^1, B^2, \ldots, B^{k-1}, B^{k+1}, \ldots, B^K\}$$

The bidding can be for a one-year contract, or more relevantly, it can be for a regulatory period of, for instance, three to five years.

It may seem surprising to pay the lowest losing bid rather than the required and lowest amount. The former is called the second-price principle, while the latter is called the first-price principle, and there are in fact good strategic reasons to choose the second-price variant of the procurement auction. It makes bidding much easier because it makes it a dominant strategy for all agents to bid their true costs. Moreover, if the payment depends on the actual bid of the winner, as in the first-price auction, the agents will submit bids with a mark-up because it would be the only way to make a margin. The resulting price to be paid will therefore often end up the same whether we use a first-price or a second-price mechanism.

It is clear that the second-price approach resembles a yardstick regime. We do, however, use bids rather than realized costs in the auction scenario. One can

Table 1: Some European regulation regimes and cost function methodologies for electricity DSOs

Code	Country	Regulation	Benchmark
AT	Austria	Revenue cap	DEA-SFA, best-of
BE	Belgium	Revenue cap	DEA
CH	Switzerland	Cost recovery	Ad hoc
DE	Germany	Revenue cap	DEA-SFA best-of
DK	Denmark	Revenue cap	COLS-MOLS
ES	Spain	Revenue cap	Engineering
FI	Finland	Revenue cap	StonED[1]
FR	France	Cost recovery	Ad hoc
GB	Great Britain	Revenue cap	COLS and Ad hoc
GR	Greece	Cost recovery	Ad hoc
HU	Hungary	Price cap	Ad hoc
IRL	Ireland	Price cap	Ad hoc
NL	Netherlands	Yardstick comp	DEA-OLS-MOLS
NO	Norway	Yardstick comp	DEA
SE	Sverige	Revenue cap	Ad hoc [2]

extend this scenario to situations with heterogenous bids by using, for example, DEA-based auctions to cope with differences in the services offered in a one-shot procurement setting. We shall discuss this below.

The second-price franchise auction regime conserves the simplicity of the fixed-price regimes but limits the informational rent. It also offers perfect adjustment to heterogeneity, as prices may vary across franchises. The problems for limited markets with high concentration are that bidding may be collusive, that excessive informational rents may be extracted and that competition may be hampered by asymmetric information among incumbents and entrants. Even under more favorable circumstances, the problems of bidding parity, asset transition and investment incentives must still be addressed, and the use of the franchising instrument in, for example, electricity distribution is likely to be scarce in the near future and to be available at first primarily for spatial and/or technical service extensions.

2.5. Applications

Table 1 gives a summary of the regulations used for electricity DSOs in 15 European countries.

Most countries rely on some revenue cap model and have derived general productivity and individual inefficiencies using benchmarking tools like DEA and SFA.

We see how some countries, like Sweden and Spain, have chosen to rely on technical engineering norms, sometimes referred to as ideal nets, in an attempt to

identify not only best practice but absolute technological possibilities.

Dynamically, the progression seems to be from a more heavy-handed cost recovery regime, over a model-based price fixation towards a high-powered market-based yardstick regime.

3. DSO regulation in Germany

In this section we will discuss the regulation of electrical DSOs in Germany. We will explain some of processes leading to the regulation and go through some highlights of the benchmarking models used.

Relevant references to the German regulation are Agrell and Bogetoft (2007), where we describe the pre-regulation analyses of a series of models to guide the final implementation plan from the regulator as described in Bundesnetzagentur (2007), which was largely transformed into an Ordinance, Government (2007). The 2008 analyses of a new dataset with the aim to serve in the first regulatory period is described in the white paper Agrell and Bogetoft (2008) and the results are summarized in Agrell et al. (2008).

3.1. Towards a modern benchmark based regulation

In 2005, it was decided to introduce a new regulation of German electricity and gas DSOs. Here, we will focus on the regulation of the electricity networks, but we note that the gas regulation and models are rather similar.

Previously, regulation occurred solely through competition law, and there was no regulator. With the new Electricity Act (EnWG), effective July 13, 2005, it was decided that "Regulation should be based on the *costs of an efficient and structurally comparable operator* and provide incentives based on efficiency targets that are *feasible and surpassable*".

The enactment of the Electricity Act marked the start of an intense and ambitious development process by the regulatory authority, the Federal Network Agency, Bundesnetzagentur (BNetzA). BNetzA performs tasks and executes power, which under the EnWG has not been assigned to the state regulatory authorities. The state regulatory authorities are responsible for regulating power supply companies with fewer than 100,000 customers connected to their electricity or gas networks and whose grids do not extend beyond state borders. In practice, the BNetzA approach has a significant impact also on the regulation of the DSOs under state regulation.

Through several development projects and a series of consultations with industry on the principles, BNetzA developed a specific proposal for how to implement the Electricity Act. As one of several consulting groups, we undertook a series of full-scale trial estimations of different model specifications. DEA and SFA models were developed based on more than 800 DSOs in each sector. This served several

purposes, some of which were to train the regulatory personnel in benchmarking methodology, to guide future data collection, to define a detailed implementation plan, and to facilitate an informed discussion with industry members.

The final proposal and detailed implementation plan by the regulator was largely transformed into the Ordinance that now provides specific guidelines for German regulation of electricity.

During 2008, we developed a new set of results to implement the Ordinance. Some highlights from this work are provided below. The new regulation became effective in 2009 for the 200 DSO under federal regulation. Smaller DSOs, with no more than 30,000 customers connected directly or indirectly to their electricity distribution system, could, instead of efficiency benchmarking to establish efficiency levels, take part in a simplified procedure. The efficiency level in the first regulatory period for participants in the simplified procedure is 87.5 percent. From the second regulatory period, the efficiency level for these DSOs is the weighted average of all efficiency levels established in nationwide efficiency benchmarking.

The regulation is currently in place and working, although there are still some aspects that are being tested in the court system by different operators.

From an international perspective, the German experience is remarkable because of the large number of DSOs, the abundance of data, as illustrated by the presence of about 250 variables for each DSO, and by the speed and efficiency with which a new regulation was established. Most other regulators have used a considerably longer period of time to undertake considerably less ambitious prototyping and full scale implementation.

3.2. Revenue cap formula

The German regulation is basically a revenue cap regulation. Each regulatory period is 5 years and the content of the first two regulatory periods have been detailed, giving the DSO more long-term forecasts on which to act.

The regulation is Totex based, i.e., both operating expenses (Opex) and capital cost expenses (Capex) are subject to regulation. Capital costs are based on either book values or standardized costs using replacement values and constant annuity calculations of yearly cost using life times of different asset groups.

The revenue cap of an individual DSO k in the German regulation in year t is determined by the formula

$$R^k(t) = C_{nc}^k(t) + (C_{tnc}^k(0) + (1 - V(t))C_c^k(0))(\frac{RPI(t)}{RPI(0)} - x(t))ExFa(t) + Q(t)$$

where C_{nc} is the cost share that cannot be controlled on a lasting basis (statutory approval and compensation obligations, concession fees, operating taxes etc.), C_{tnc} is the cost share that cannot be controlled on a temporary basis (essentially the

efficient cost level found as the total costs multiplied by the efficiency level), C_c are the controllable costs, $V(t)$ is a distribution factor for reducing inefficiencies (initially set to remove incumbent inefficiency after two regulatory periods, i.e., 10 years), RPI(t) is the retail price index in year t, $RPI(0)$ is the retail price index in year 0, and $x(t)$ is the general productivity development from year 0 to year t reflecting the cumulative change in the general sectoral productivity for year t of the particular regulatory period relative to the first year of the regulatory period. Also, ExFa is an expansion factor reflecting the increase in service provision in year t compared to year 0 and determined as

$$ExFa_j^k(t) = 1 + max\left(\frac{L_j^k(t) - L_j^k(0)}{L_j^k(0)}, 0\right)$$

where $L_j(t)$ is the volume of load at level j in year t of the particular regulatory period. The expansion factor for the entire network is the weighted average of all network levels. Lastly, Q(t) is the increase or decrease in the revenue cap from quality considerations. Revenue caps may have amounts added to or deducted from them if operators diverge from required system reliability or efficiency indicators (quality element). The quality element is left to the discretion of the regulator.

3.3. Benchmarking requirements

From a benchmarking perspective, the regulation is remarkable for being explicit with respect to a series of technical aspects such as cost drivers, estimation techniques, return to scale and outlier criteria.

The Ordinance is specific about a *minimal set of cost drivers*. Cost drivers such as connections, areas, circuit length, and peak load, are obligatory. Of course, this leaves a series of available alternatives even within these groups and it does not exclude cost drivers covering other aspects of the service provision.

The German incentive regulation is also explicit as to which *estimation techniques* to use in the benchmarking and how to combine the results of multiple models. According to Section 12 of the Ordinance, the efficiency level for a given DSO is determined as the maximum of four efficiency scores, $E_{DEA}(B)$, $E_{DEA}(S)$, $E_{SFA}(B)$, and $E_{SFA}(S)$, where E_{DEA} is the Farrell efficiency, calculated with a NDRS-DEA model, E_{SFA} is the Farrell input efficiency, calculated using a SFA model, and the arguments B and S denote book values and standardized capital costs, respectively. As such, the regulation takes a cautious approach and biases the decision in favor of the DSOs in case of estimation risk. Entities demonstrating particularly low efficiency are given the minimum level of 60 percent. In summary, the efficiency of DSO k is calculated using this equation

$$\max\{E_{DEA}^k(B), E_{DEA}^k(S), E_{SFA}^k(B), E_{SFA}^k(S), 0.6\}$$

It is worthwhile noting that the Ordinance does not prescribe any bias correction for the DEA scores, nor does it rely on confidence intervals for the scores, as they could be calculated in both the DEA models (via boot-strapping) or in the SFA models.

The Ordinance is also specific about *how to identify outliers*. Indeed, it prescribes two outlier criteria to be tested for each DSO, and if any of them is fulfilled, the DSO cannot be allowed to affect the efficiency of the other DSOs. The two criteria can be formalized in the following ways. Let $K^* = \{1, \ldots, K\}$ be the DSOs is the data set, and k be a potential outlier. Also, let, $E(h, K^*)$ be the efficiency of h when all DSO are used to estimate the technology and let $E(h, K^* \setminus k)$ be the efficiency when DSO k does not enter the estimation.

The *first outlier criterion* is that a single DSO should not have too large of an impact on the average efficiency. We can evaluate the impact on the average efficiency by considering

$$\frac{\sum_{h \in K^* \setminus k}(E(h, K^* \setminus k) - 1)^2}{\sum_{h \in K^* \setminus k}(E(h, K^*) - 1)^2}$$

The test compares the average efficiency of the other operators when k cannot affect the technology as compared to the average efficiency of the other DSOs when k is allowed to impact the evaluations. Since $E(h, K^* \setminus k) \geq E(h, K^*)$, this ratio is always less than or equal to 1, and the smaller the ratio is, the larger the impact of k, i.e., small values of the ratio will be an indication that k is an outlier. The asymptotic distribution of the ratio is $F(K - 1, K - 1)$ following Banker (1993).

The *second outlier criterion* is that no DSO k will be extremely super-efficient in the sense that

$$E(i, K^* \setminus k) > q(0.75) + 1.5(q(0.75) - q(0.25))$$

where $q(a)$ is the a quantile of the distribution of super-efficiencies, such that e.g., $q(0.75)$ is the super-efficiency value, below which exist 75% of DSOs. This criteria is inspired by Banker and Chang (2006).

In addition to these outlier rules, the ordinance prescribes the use of common econometric outlier detection methods like Cook's distance.

The Ordinance also prescribes the *return to scale assumption* to be used in the DEA models of the regulation, namely as a non-decreasing economy of scale, an IRS or NDRS technology.

The *high level of technical specifications in the German Ordinance* is remarkable and uncommon in an international context. There are several reasons for this. One is probably that it was considered a way to protect the industry against extreme outcomes. The cautious approach of specifying a minimal set of cost drivers and of

using the best-of-four approach with an added lower bound of 60% clearly provides some insurance ex-ante to the DSOs about the outcome of future benchmarking analyses. The extensive pre-Ordinance analyses and full scale testing of alternative models and techniques is, of course, also an important pre-requisite. Without such analyses it would not have been possible to design the regulation in such detail nor to engage in qualified discussion with the industry about alternative approaches. It is worthwhile to note that during the initial analyses leading to the Ordinance, no information was revealed about the efficiency of individual DSOs. Only the general level of efficiency and the distributions of efficiencies were public during this phase.

3.4. Model development process

The development of a regulatory benchmarking model is a considerable task due to the diversity of the DSOs involved and the economic consequences that the models may have. Some of the important steps in the German model development were:

Choice of variable standardizations: Choice of accounting standards, cost allocation rules, in/out of scope rules, assets definitions, operating standards etc. were necessary to ensure a good data set from DSOs with different internal practices.

Choice of variable aggregations: Choice of aggregation parameters, like interest and inflation rates, for the calculation of standardized capital costs, and the search for relevant combined cost drivers, using, for example, engineering models, were necessary to reduce the dimensionality of possibly relevant data.

Initial data cleaning: Data collection were an iterative process where definitions are likely to be adjusted and refined and where collected data were constantly monitored by comparing simple KPIs across DSOs and using more advance econometric outlier detection methods.

Average model specification: To complement expert and engineering model results, econometric model specification methods were used to investigate which cost drivers best explain cost and how many cost drivers were necessary.

Frontier model estimations: To determine the relevant DEA and SFA models, they must be estimated, evaluated and tested on full-scale data sets. The starting point were the cost drivers derived from the model specification stage, but the role and significance of these cost drivers were examined in the frontier models, and alternative specifications derived from using alternative substitutes for the cost drivers weree investigated, taking into account the outlier detecting mechanisms.

Model validation: Extensive second stage analyses were undertaken to see if any of the more than 200 non-included variables should be included. The second stage analyses were typically done using graphical inspection, non-parametric (Kruskal-Wallis) tests for ordinal differences, and truncated regression (Tobit regressions) for cardinal variables. Using the Kruskal-Wallis method, we tested, for example, whether there was an impact on 1) year of cost base, 2) the East-West location of the DSO, and the DSO's possible involvement in water, district heating, gas, or telecommunication activities. Using Tobit regressions, we tested a series of alternative variables related to cables, connections and meters, substations and transformers, towers, energies delivered, peak flows, decentralized generation, injection points, population changes, soil types, height differences, urbanization, areas etc.

It is worthwhile emphasizing, once again, that model development is not a linear process but rather an iterative one. During the frontier model estimation, for example, one may identify extreme observations that have resulted from data error not captured by the initial data cleaning or the econometric analyses and which may lead to renewed data collection and data corrections. This makes it necessary to redo most steps in an iterative manner.

The non-linear nature of model development constitutes a particular challenge in a regulatory setting where the soundness and details of the process must be documented to allow opposing parties to challenge the regulation in the courtroom.

Also, since corrections of previous steps typically have to repeatedly and since there is also typically a considerable time pressure in the regulatory setting, it is important to organize work appropriately. Scripts to support this can be developed using more advanced software and are very important and useful for such purposes since they allow massive recalculations in a short period of time and they document the calculation steps in great detail.

3.5. Model choice

The choice of a benchmarking model in a regulatory context is a multiple criteria problem. There are several objectives, which may conflict with one another. To emphasize this, note at least the following four groups of criteria.

Conceptual: It is important that the model makes conceptual sense both from a theoretical and a practical point of view. The interpretation must be easy and the properties of the model must be natural. This contributes to the acceptance of the model in the industry and provides a safeguard against spurious models developed through data mining and without much understanding of the industry. More precisely, this has to do with the choice of outputs that are natural cost drivers and with functional forms that, for example, have reasonable returns to scale and curvature properties.

Statistical: It is, of course, also important to discipline the search of a good model with classical statistical tests. We typically seek models that have significant parameters of the right signs and that do not leave a large unexplained variation.

Intuition and experience: Intuition and experience is a less stringent but important safeguard against false model specifications and the over- or underuse of data to draw false conclusions. It is important that the models produce results that are not that different from the results one would have found in other countries or related industries. Of course, in the usage of such criteria, one also the runs the risk of mistakes. We may screen away extraordinary but true results (Type 1 error) and we may go for a more common set of results based on false models (Type 2 error). The intuition and experience must therefore be used with caution.

One aspect of this is that one will tend to be more confident in a specification of inputs and outputs that leads to comparable results in alternative estimation approaches, e.g., in the DEA and SFA models. The experiential basis of this is that when we have a bad model specification, SFA will identify a lot of unexplained variation and therefore attribute the deviations from the frontier to noise rather than inefficiency. Efficiencies in the SFA model will therefore be high. DEA, on the other hand, does not distinguish noise and inefficiency, so in a DEA estimation, the companies will look very inefficient. Therefore, results that deviate too drastically in the DEA and SFA estimations may be a sign that the model is not well specified. It should be emphasized that the aim is not to generate the same results using a DEA and a SFA estimation. The aim is to find the right model. Still, intuition and experience suggest that a high correlation between the DEA and SFA results is an indication that the model specification is reasonable. Therefore it also becomes an indirect success criterion.

Regulatory and pragmatic: The regulatory and pragmatic criteria calls for conceptually sound, generally acceptable models as discussed above. Also, the model will ideally be stable in the sense that it does not generate too much fluctuation in the parameters or efficiency evaluations from one year to the next. Otherwise, the regulator will lose credibility and the companies will regard the benchmarking exercise with skepticism. Of course, one will not choose a model simply to make the regulator's life easy, so it is important to remember that similar results are also a sign of a good model specification, cf. the intuitive criteria above. The regulatory perspective also comes into the application of the model. If the model were not good, a high powered incentive scheme, for example, would not be attractive since it would allocate too much risk to the firms. Lastly, let us mention the trivial but very important requirement to comply with the specific conditions laid out in the regulatory directives like the German Ordinance.

Since some of these objectives may conflict we need to make some trade-offs.

16

As an example, it may be that the Ordinance prescribes a cost driver group that in some models is not significant. In that case, there will be a conflict between the statistical logic and the law, and we have to make a trade-off in favor of the latter.

The multiple criteria nature of model choice is a particular challenge in regulation. When we have multiple criteria, they may conflict as mentioned above, and this means that there is no optimal model that dominates all other models. We have to make trade-offs between different concerns to find a compromise model, to use the language of multiple criteria decision making, and such trade-offs can be challenged by the regulated parties.

3.6. Final model

The final German electricity DSO model used the input and outputs shown in Table 2.

Table 2: German model of electricity DSOs

Input	Outputs (cost drivers)
Total costs:	Connections.hs.ms.ns
Totex	Cables.circuit.hs.share.cor
or	Lines.circuit.hs.share.cor
Totex.standard	Cables.circuit.ms
	Lines.circuit.ms
	Net.length.ns
	Peakload.HSMS.unoccupied.cor
	Peakload.MSNS.unoccupied.cor
	Area.supplied.ns
	Substations.tot
	Decentral.prod.cap.tot

From an international perspective, this model specification is comparable in terms of the cost driver coverage included. Regulatory models of electricity DSOs generally have cost drivers related to transport work, capacity provision and service provision. We do not have any transport work cost drivers, but this lack is in accordance with engineering expectations and is confirmed by both model specification tests and second-stage testing. The number of cost drivers is at the high end of what we have used elsewhere.

The DEA models were IRS (NDRS) models, as prescribed in the Ordinance, and the outliers were excluded using the two DEA outlier criteria above. In practice, only the last outlier criterion was really effective.

In the SFA models, we used a normed linear specification where the norming constant was Connections.hs.ms.ns. The reason for norming (deflating) the data was to cope with heteroscedasticity; the absolute excess costs, i.e. the inefficiency terms in a SFA model, will increase with the size of the company even if the percentage of extra costs is fixed. Likewise, the noise term is expected to have variance that increases with the size of the DSO. We could, of course, have handled the heteroscedasticity problem using a log-linear specification, but we did not do so to avoid the specification's curvature problem — the output-isoquants in a log-linear specification curve the opposite way than normal output-isoquants do. This difference is not surprising, as the log-linear model corresponds to a Cobb-Douglass model, which is really a production and not a cost function. Furthermore, the normed linear model is conceptually easy to interpret.

To supplement the analyses, we performed sensitivity evaluations of the impact of using a normed linear or a log-linear SFA specification and investigated the impact of using a linear with constant terms which would be more similar to a VRS model. The end results were insensitive to these model variations.

A summary of the resulting efficiency levels is provided in the Table 3 below.

Table 3: Final efficiencies in German electricity model, Agrell and Bogetoft (2007).

Model	Mean	Std.Dev.	Min	$\#E < 0.6$	$\#E = 1$
BestOfTwoTotex	0.898	0.074	0.729	0	40
BestOfTwoTotex.stand.	0.920	0.058	0.795	0	43
BestOfFour	0.922	0.059	0.795	0	49

We see that the resulting efficiency evaluations are high and that with 10 years to catch up, the yearly requirements are modest. Of course, the catch-up requirements will also be evaluated in terms of the cost elements involved, but there are considerable non-benchmarked cost elements, and a relatively large share of the total costs is Opex.

Although the resulting requirements may seem modest, this situation is not necessarily a bad outcome for the regulator. First, it may reflect the fact that the German DSOs are relatively efficient, and second, it may facilitate the institutionalization of model-based regulation. Also, despite the modest estimated average inefficiency of 7.8%, the economic stakes are still considerable at a national level.

Of course, for most companies, the stakes are relatively modest, and for individual consumers, the stakes are very modest indeed. This limited effect actually provides a rationale for central regulation; the individual economic gains are small, making it unlikely that individuals will spend many resources challenging the DSO charges.

4. DSO regulation in Norway

In 2007 the Norwegian regulator for electricity DSOs, the Norwegian Water Resources and Energy Directorate (NVE), moved from an ex ante revenue cap regulation to a DEA-based yardstick competition regime along the lines of Bogetoft (1997). The benchmarking model used in the Norwegian yardstick regulation was first developed in Agrell and Bogetoft (2004). The 2010 version of the regulation is summarized in Langset (2009). A comparison of regulation in the Nordic countries is provided in Agrell et al. (2005a).

The Norwegian revenue cap is determined as

$$R^k(t) = 0.4C^k(t) + 0.6C_{DEA}^k(t-2) + IA^k(t)$$

where R^k is the revenue cap, C_{DEA}^k is the DEA-based cost norm for companies based on data from year $t-2$ and $IA^k(t)$ is the investment addition to take into account the new investments from year t. The actual costs $C^k(t)$ are calculated as

$$C^k(t) = (Opex^k(t-2) + QC^k(t))\frac{CPI(t)}{CPI(t-2)} + pNL^k(t) + DE^k(t-2) + rCap^k(t-2)$$

where QC is quality compensation by firm k to consumers as a consequence of lost load, CPI is the consumer price index, NL is the net-loss, p is the price of power, DE is depreciation, Cap is the capital basis and r is the interest rate on capital set by the regulator.

The cost norm C_{DEA}^k is calculated in two steps. The main calculation is a DEA CRS model with 8 cost drivers covering lines, net stations, delivered energy, numbers of ordinary and vacation users, forests, snow and coastal climate conditions. The second step is a regression-based second stage correction based on border conditions, decentralized power generation and number of coastal islands in the concession area.

NVE has internationally been a pioneer in the design of model-based regulation of electricity DSOs. In 1991, they introduced Rate of Return Regulation (ROR) and in 1997 they moved to a DEA-based revenue cap regulation that was in place until the introduction of the yardstick regime in 2007. The movement to a yardstick-based regime can be seen as a natural next step in the attempt to mimic a competitive situation in a natural monopoly industry. Still, the transition from a well-established revenue cap system required careful planning.

One challenge was to convince the industry that a yardstick regime is less risky than an ex ante revenue cap system. The latter enables the companies to predict the future income several years in advance. At first this may seem to be a big advantage but, since it does not include the cost side except for the use of a more or less arbitrary inflation adjustment, it actually does not protect the profit, which

should be the main concern for the companies. The yardstick regime offers more insurance because technological progress and costs are estimated directly using the newest possible data.

Another challenge was to calibrate the transition to avoid dramatic changes for any individual firms moving from one benchmarking practice to another.

A third challenge was to enable the firms to close their financial accounts in due time. This is a general challenge of the yardstick competition. A firm's allowed income for period t can only be calculated after data from all firms have been collected regarding year t. Assuming that the firms are able to deliver this information sometime in the middle of year $t + 1$, the regulator needs at least half a year to validate data and make his calculations. This means that the allowed income for year t will only be known in year $t + 2$. Therefore, in practice, such regulation often works with a time-lag such that the cost norm for period t is based on data from period $t - 2$. This also means that the difference between an ex ante revenue cap and a yardstick-based regime is reduced; the latter becomes similar to a revenue cap with annual updating of the cost norms.

The structural properties of an industry, i.e. the firms' scale, scope, ownership etc., may be just as important as the cost reduction efforts of the individual firms. 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.

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, though they might lead to more outputs being produced with fewer inputs. As discussed in Bogetoft and Otto (2011) and Bogetoft (2012), the Norwegian Water Resources and Energy Directorate handles this, by calculating the so-called harmony effect from Bogetoft and Wang (2005) and by compensating a merged firm for the extra requirements corresponding to this effect. At the same time, mergers will tend to affect the performance evaluation basis and may lead to more rents to the firms because the cost norm becomes less demanding. The regulator, who considers allowing a merger, must therefore trade-off the gains from improved costs to the firms with the losses from a shrinking information basis. The latter is the regulatory equivalent of the negative market effects of a merger in a non-regulated sector.

5. DEA based incentive schemes

In this section, we provide an introduction to and summary of the theoretical foundation for DEA based incentives and regulation.

The connection between DEA and the formal literature on games was first suggested by Banker (1980) and Banker et al. (1989). Linkage with the formal performance evaluation and motivation literature, most notably the agency theory and related regulation and mechanism design literature, has subsequently been the subject of a series of papers including Agrell and Bogetoft (2001); Agrell et al. (2005b), Bogetoft (1994a,b, 1995, 1997, 2000) Bogetoft and Hougaard (2003), Bogetoft and Nielsen (2008),Bowlin (1997), Dalen (1996); Dalen and Gomez-Lobo (1997, 2001), Førsund and Kittelsen (1998), Resende (2002), Sheriff (2001), Thanassoulis (2000) and Wunsch (1995).

In the following we will highlight some of these results.

5.1. Framework

The basic problem addressed in this line of research is the following. Assume that a principal (regulator) has access to data on the multiple inputs used x^k and and multiple outputs produced y^k in each of K agents (firms, DSOs)

$$(x^k, y^k), \qquad k = 1, \ldots, K$$

Based hereon, what can the principal reasonably ask the agents to do in the future and how should he motivate and compensate them to do so?

The answer to these questions depends on the *organizational context* and in particular on the technological, informational and preferential assumptions of the parties.

The relevance of DEA is in general related to situations, where the principal faces considerable *uncertainty about the technology*. In a single input multiple output cost setting, the principal may know that the cost function is increasing and convex, but otherwise have no *a priori* information about the cost structure. In pure moral hazard models, we also assume that the agents face a similar uncertainty.

The general case also empowers agents to take *private actions*, which the principal cannot observe. The action could be to reduce costs or increase the quality of the work done. This leads to a usual *moral hazard problem* since the principal and the agents may conflict as to which actions the agents shall take. The traditional setting depicts the agents as work averse, tempted to rely on their good luck and to explain possibly bad performances with unfavorable circumstances.

In some models, we also consider the possibility that the agents have *superior information* about the working conditions before contracting with the principal. A hospital manager may have good information about the primary cost drivers at his hospital while the Ministry of Health may have little information about what causes the total bill to increase. This leads to the classical *adverse selection problem* where an agent will try to extract information rents by claiming to be working under less favorable conditions.

5.2. Interests and decisions

It is common to assume that the principal is risk neutral and that the agents are either risk averse or risk neutral. The principal's aim is to minimize the costs of inducing the agents to take the desired (hidden) actions in the relevant (hidden) circumstances. An agent's aim is usually to maximize the utility from payment minus the dis-utility form private effort.

In the combined moral hazard and adverse selection models, we usually make a simplifying assumption about the structure of the agent's trade-offs between effort and payment. We assume that the aim of any agentis to maximize a weighted sum of profit and slack:

$$U(y^k, B^k) = u(B^k) - v(y)$$

or more specifically,

$$U(y^k, B^k) = (B^k - x^k) - \rho(x^k - c(y^k))$$

where y^k is the outputs produced, B^k is the payment received, slack $(x^k - c(y^k))$ is a measure of the extent to which input utilization x^k exceeds the minimal possible $c(y^k)$, and $\rho^k \in (0, 1)$ is the relative value of slack.

We will rely on these assumptions below, but we realize that, although widely used in the literature, they constitute a stylized caricature of intra-organizational decision making and conflict resolution. This is not satisfactory and is in sharp contrast to the nuanced production description that state-of-the-art performance evaluation techniques like DEA enables. Moreover, recent applications have demonstrated that to derive regulation and incentive schemes with a more sound theoretical basis, we need to know more about what goes on inside the black box of the firm. Only thus can we study, in more detail, the combined use of incentive regulation and regulation by rights and obligations that are used in practice and only in this way can we make valid statements about the speed and path of improvements that a new performance-based scheme may foster. The recent idea of *rational inefficiency* discussed in Bogetoft and Hougaard (2003), Bogetoft and Andersen (2009), Asmild et al. (2009), Bogetoft et al. (2009), and Asmild et al. (2013), is an attempt to provide a more nuanced view of the preferences involved in the selection of multiple dimensional production plans and slack elements. A discussion of this, however, is beyond the scope of the discussion in this paper.

5.3. Super efficiency in incentive schemes

One of the first lessons, from the incentive perspective, is that the traditional Farrell score is not useful. The Farrell output efficiency F gives all units on the

relative efficient frontier a score of 1. This severely limits the ability to give high-powered incentives based on Farrell measures. The Farrell measures can give incentives to match others but not to surpass the norm and push out the frontier. Combining this with the multiple dimensional characteristics of the typical DEA model and thereby with the ability to be special in different ways, *the Nash Equilibria (NE)* that can be implemented using the Farrell measure will often involve minimal effort and maximal slack.

Figure 3 illustrates this. Here, we assume that the cost to the agents is proportional to the length of the production vectors and that payment is decreasing in the Farrell output efficiency score F,

$$F^k > F^{*k} \Rightarrow B^k(F^k) \leq B^k(F^{*k})$$

such that maximal payment is received when a firm is efficient with a score of $F = 1$. If Firm 1 planned to produce at A and moves from A to C, it would get the same payment but use less effort. A is therefore not a best response. Next, Firm 2 could move from the planned B to an easier life in D, again reducing private costs of effort without affecting its payment. This procedure can continue until they both use minimal effort and receive maximal payment.
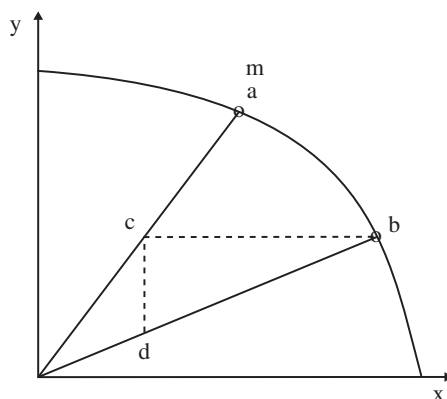

Figure 3: Nash equilibria under Farrell incentives

This somewhat discouraging outcome can easily be remedied by making the payment decreasing in the super-efficiency rather than in the usual output efficiency. In Figure 3, the output-based super efficiency for Firm 1 in A is approximately 0.6, but if the payment is sufficiently decreasing in F^{SUP}, it would not pay to reduce the effort. It does not pay to reduce the effort if the marginal reduction in payment exceeds the marginal decease in the cost of effort.

More generally, using super-efficiency, one can support the implementation of most plans, even in so-called un-dominated Nash-equilibria as first demonstrated

in Bogetoft (1995)

5.4. Incentives with individual noise

Another fundamental result concerns a pure moral hazard context with ex post evaluations of the performance of the firms when there is

- Considerable technological uncertainty a priori,

- Risk averse firms and

- Individual uncertainty (noise) in the firms' performances.

Technological uncertainty is represented by a large class of *a priori* possible technologies, e.g. the set of production functions that are increasing and concave or the set of functions that are increasing. One can now ask when the DEA frontier is sufficient to write an optimal contract, i.e., when

$$B^{*k} = B^k(x^k, y^k, C^{DEA}(\cdot \mid x^{-k}, y^{-k}))$$

This is the case where optimal relative performance evaluations can be made by comparing the performance of a given firm against the DEA best practice frontier, estimated from the performance of the other firms.

It turns out that i) DEA frontiers support optimal contracts when the distributions of the individual noise terms are exponential or truncated, and that ii) DEA frontiers, based on large samples, support optimal contracts when noise is monotonic, in the sense that small noise terms are more likely than large noise terms, cf. Bogetoft (1994a). Hence, even when we have individual noise elements and not just the structural uncertainty, which intuitively seems to favor DEA, DEA- based contracts will be optimal for special distributional assumptions and for general assumptions, if the sample is sufficiently large.

5.5. Incentives with adverse selection

Another set of results concern combined adverse selection and moral hazard problems with

- Considerable asymmetric information about the technology

- Risk neutral firms,

- Firms seeking to maximize Profit $+ \rho$ Slack utility.

The firms are supposed to have superior technological information. In the extreme case, they know the underlying true cost function with certainty, while the regulator only knows the general nature of the cost function. Thus, the regulator may know that there are fixed unit costs of the different outputs but not the exact unit cost because it is the firm's private information. Alternative assumptions may be made about the information available to the regulator. We may assume, for example, that the regulator only knows that the cost function is increasing and convex.

The optimal solution in this case depends on whether the actual costs, i.e., the minimal possible cost plus the slack introduced by the firm, can or cannot be verified and therefore contracted upon.

If the actual costs x cannot be contracted upon, the optimal solution is to use

$$B^{*k}(y^k) = b^k + C^{DEA}(y^k \mid x^{-k}, y^{-k})$$

i.e. the optimal compensation equals a lump sum payment plus the DEA estimated cost norm ex ante. The size of the lump sum payment b^k depends on the firm's alternatives, i.e., its reservation profit, which in turn depends on profit potentials in other markets or the surplus from contracting with other regulators, for example, private insurance companies. One consequence of this result is that a best way to downsize an organization when there is considerable uncertainty about the cost drivers may be via a lawn-mowing approach where all product types are downsized by the same amount as shown in Bogetoft (2000). This situation corresponds to a situation where the only ex ante data is the historical production of the firm in question.

If, instead, we assume that the actual costs of the firm can be contracted upon, the optimal reimbursement scheme becomes

$$B^{*k}(x^k, y^k) = b^k + x^k + \rho^k(C^{DEA}(y^k; x^{-k}, y^{-k}) - x^k)$$

as demonstrated in Bogetoft (1997) and Bogetoft (2000). This shows that the optimal compensation equals a lump sum payment plus actual costs plus a fraction ρ^k of the DEA estimated cost savings. The structure of this payment scheme can be interpreted as a *DEA based yardstick model*: Using the realized performances of the other firms, the regulator creates a cost yardstick against which the regulated firm is evaluated. The regulated firm is allowed to keep a fraction ρ of its saving compared to the yardstick costs as his effective compensation. Figure 4 illustrates this reimbursement scheme.

These results provide an incentive rationale for using DEA-based revenue cap and yardstick competition systems in contexts where the regulator faces considerable uncertainty about the underlying cost structure. Note that the performance of the other firms can, in both cases, be interpreted as either historical data, as it
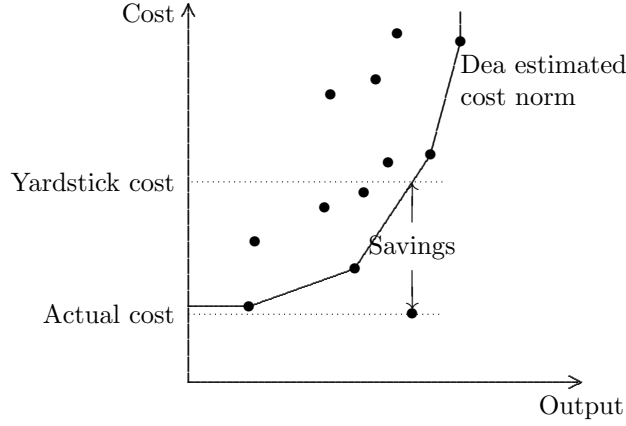
Figure 4: DEA based yardstick competition

is generally used in the revenue cap regulation, or as actual data, as is the idea in the ex post yardstick regulation regime.

5.6. Dynamic incentives

In the previous section, we considered incentives for a single period based on historical or current information. Dynamic cases with multiple periods are more complicated since they give rise to new issues like the

- Possibility to accumulate and use new information from one or more firms,

- Need to avoid the ratchet effect, i.e., deliberate sub-performance in early periods to avoid facing too tough standards in the future, and

- Possibility of technical progress (or regress).

The structure of the optimal dynamic scheme is similar to the ones developed above as demonstrated in Agrell et al. (2002) and Agrell et al. (2005b). Thus, the optimal revenue cap for a firm is determined by a DEA-based yardstick norm. Assuming verifiable actual costs, and taking into account the generation of new information, the ratchet effect and the possible technical progress, the optimal scheme becomes

$$B_t^{*k}(x_t^k, y_t^k) = b_t^k + x_t^k + \rho^k(C^{DEA}(y_t^k \mid x_{1-t}^{-k}, y_{1-t}^{-k}) - x_t^k)$$

where $C^{DEA}(y_t^k \mid x_{1-t}^{-k}, y_{1-t}^{-k}) - x_t^k)$ is the DEA-based cost norm that uses all the information from the other firms generated in periods 1 through t. By relying only on information from the other firms in setting the norm, we avoid the ratchet effect, and by relying on all previous performances, we presume that there is no technical regress.

26

Of course, the dynamic case can be further extended, e.g., by including incentives to innovate and to share innovative practices. Also, it could be extended to situations where the catch-up capacity is somewhat constrained such that immediate catch-up, as it is assumed here, is avoided.

5.7. Bidding incentives

The results summarized above all concern incentives and coordination of activities in view of realized production plans. The realized production plans may be generated ex ante or they may be part of a future multiple agent production context. An interesting extension of these ideas concerns the possibility of using DEA and related benchmarking techniques to select the winner of a procurement auction and the compensation to provide to the winner. The results above can be extended in this way, although the exercise is non-trivial.

The DEA-based auction discussed in Bogetoft and Nielsen (2008) extends the idea of a second-price auction auction to a multiple output case where the services (outputs) offered by the different agents are not the same and where the DEA serves to interpolate a reasonable second price, even in cases where no other bidder is offering the same output profile. Alternatively, one can look at the DEA baed auction as an extension of the so-called second score auction for multiple dimensional procurement settings suggested by Che (1993).

Consider a situation where a principal is interested in procuring a multiple dimensional service bundle y. The associated required payment from a bidder by x. The value to the buyer if he gets y and pays x is $V(y) - x$, and the value to the provider is $x - c(y)$ where $c(y)$ is the minimal costs of producing y. The questions addressed in the procurement auction model are how to evaluate the bids, how to choose a provider and how to compensate the provider.

The DEA based auction mechanism and runs according to the following protocol

1. The bidders submit price-output bids, $(x^k, y^k), k = 1, \ldots, K$

2. Each bid is assigned a score, $S(x^k, y^k) = V(y^k) - x^k, k = 1, \ldots, K$

3. The bid with the highest score wins

4. The winner is compensated with the smaller of the second-score price and the DEA cost, and losers are not compensated, i.e. if agent i wins he is compensated

$$B^i(x, y) = min\{C^{DEA-i}(y^i; k), V(y^i) - S^{(2)}\}$$

where S^2 is the second highest score.

It is relatively easy to see, that it is an optimal strategy for any agent to simply demand his minimal production costs, i.e. to truthfully reveal his costs. The truth-telling property follows from the fact that the price bid only affects the chance of winning, not the terms in case of winning. It follows now that the outcome is socially (allocatively) efficient in the sense that the net benefit is maximized, i.e. the winning agent is the one maximizing the social welfare as given by the value created minus the underlying minimal cost of providing the services. Lastly, we note that as long as the underlying costs function comply with the mild regularity of the DEA models, the DEA based auction is individually rational and leads to lower expected procurement costs than the normal second score auction. The payment is the minimum of the second-score payment and the DEA yardstick cost, calculated based solely on the bids from other bidders.

Fig. 5 illustrates how the DEA based auction in the case of VRS technology. The winner is Agent 1, while Agent 2 is the second runner up. Agent 1 is reimbursed by a convex combination of the requests forwarded by Agent 2 and Agent 3, as this gives a lower payment than the second-score payment $V(y^1) - S^{(2)}$.
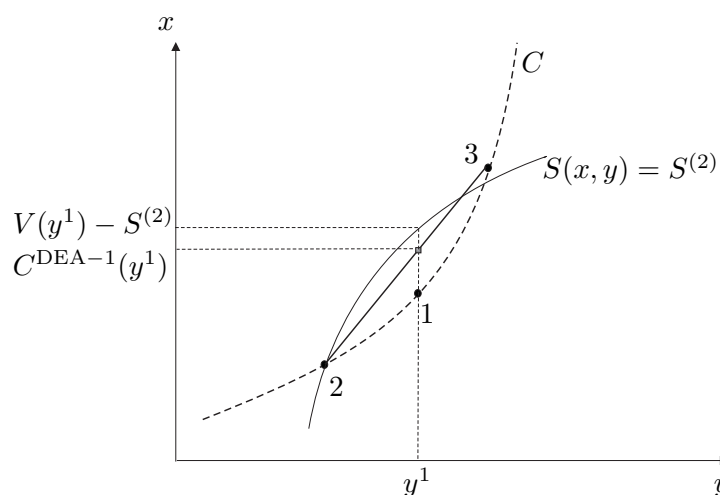

Figure 5: The DEA-based auction

6. Summary

Benchmarking can be used to facilitate motivation and contracting. One of the areas where modern benchmarking techniques like DEA and SFA are widely used for motivation purposes is in the regulation of natural monopolies like local or regional electricity and gas distribution systems. In regulatory contexts, the firms generally have superior information about the cost structures, and benchmarking

28

helps the regulator to undermine the firms' superior information and, thereby, their ability to extract information rents.

In this paper, we discussed how different regulations need benchmarking. We saw that price fixation schemes, like a revenue cap system, need benchmarking at least once before every regulatory period, i.e., at least once every 3-5 years, to evaluate the general productivity developments as well as individual incumbent inefficiencies that will determine how much cost reduction the regulator can reasonably request. We also saw that a more advanced regulation like yardstick competition will need yearly benchmarks to evaluate ex post the reasonable costs of the previous year. Lastly, we saw that franchise auctions can make use of benchmarking of the bids to compare different offers across service levels. We also surveyed the systems used in 15 European countries.

As a more specific example, we covered the regulation of German electricity distribution systems operators. We saw how the German approach is cautious. It evaluates every DSO using four different models and relies on the most positive evaluation in setting the allowed income. We also saw how outlier detection based on super-efficiency was part of the regulatory set-up, and we covered the many different steps in a regulatory benchmarking model from the choice of variable standardizations and aggregations, over data cleaning to average model specification, frontier estimations and extensive second stage analyses with the aim of developing a model that is conceptually sound, adheres to general statistical principles, complies with intuition and experience, as well as with regulatory requirements while also taking into account also what is feasible and not just desirable. The economic stakes in a regulatory context may be considerable. Taking the cautious German approach, we estimated a potential savings of about 0.4 billion Euros.

Having covered some practical applications, we turned to part of the theoretical basis of DEA-based contracting. We showed that DEA-based contracts may be optimal in some settings, particularly when there is considerable uncertainty about the underlying cost functions. With risk neutral firms, a DEA-based yardstick regime may be the optimal regulation. A specific implementation of this is the new DSO regulation introduced in Norway since 2007.

References

Agrell, P. J., Bogetoft, P., 2001. Should health regulators use dea? In: Fidalgo, E. e. a. (Ed.), Coordinacion e Incentivos en Sanidad. Asociasion de Economia de la Salud, pp. 133–154.

Agrell, P. J., Bogetoft, P., 2004. Nve network cost efficiency model,. Tech. rep., Norwegian Energy Directorate NVE.

Agrell, P. J., Bogetoft, P., 2007. Development of benchmarking models for german electricity and gas distribution. Consultation report, Bundesnetzagentur, Bonn, Germany.

Agrell, P. J., Bogetoft, P., 2008. Electricity and gas dso benchmarking whitepaper. Consulattion report, Bundesnetzagentur.

Agrell, P. J., Bogetoft, P., 2010. A primer on regulation and benchmarking with examples from network industries. Tech. Rep. version 05, SUMICSID AB.

Agrell, P. J., Bogetoft, P., Bjørndalen, J., Vanhanen, J., Syrjänen, M., 2005a. Nemesys subproject a: System analysis. Consultation report, Nordenergi.

Agrell, P. J., Bogetoft, P., Cullmann, A., von Hirschhausen, C., Neumann, A., Walter, M., 2008. Ergebnisdokumentation: Bestimmung der effizienzwerte verteilernetzbetreiber strom. Consultation report, Bundesnetzagentur.

Agrell, P. J., Bogetoft, P., R.Halbersma, M.C.Mikkers, 2007. Yardstick competition for multi-product hospitals. NZa Research Paper 2007/1, NZa, Netherlands.

Agrell, P. J., Bogetoft, P., Tind, J., 2002. Incentive plans for productive efficiency, innovation and learning. International Journal of Production Economics 78, 1–11.

Agrell, P. J., Bogetoft, P., Tind, J., 2005b. Dea and dynamic yardstick competition in scandinavian electricity distribution. Journal of Productivity Analysis 23, 173–201.

Asmild, M., Bogetoft, P., Hougaard, J. L., 2009. Rationalising dea estimated inefficiencies. Journal of Business Economics (Zeitschrift für Betriebswirtschaft ZfB), 81–86.

Asmild, M., Bogetoft, P., Hougaard, J. L., 2013. Rationalizing inefficiency: Staff utilization in branches of a large canadian bank. Omega 41, 80–87.

Banker, R., Chang, H., 2006. The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research 175 (2), 1311–1320.

Banker, R. D., 1980. A game theoretic approach to measuring efficiency. European Journal of Operational Research 5, 262–268.

Banker, R. D., 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 39 (10), 1265–1273.

Banker, R. D., Charnes, A., Cooper, W. W., Clarke, R., 1989. Constrained game formulations and interpretations for data envelopment analysis. European Journal of Operational Research 40, 299–308.

Bogetoft, P., 1994a. Incentive Efficient Production Frontiers: An Agency Perspective on DEA. Management Science 40, 959–968.

Bogetoft, P., 1994b. Non-Cooperative Planning Theory. Springer-Verlag.

Bogetoft, P., 1995. Incentives and Productivity Measurements. International Journal of Production Economics 39, 67–81.

Bogetoft, P., 1997. DEA-based yardstick competition: The optimality of best practice regulation. Annals of Operations Research 73, 277–298.

Bogetoft, P., 2000. DEA and activity planning under asymmetric information. Journal of Productivity Analysis 13, 7–48.

Bogetoft, P., 2012. Performance Benchmarking - Measuring and Managing Performance. Springer, New York.

Bogetoft, P., Andersen, J., 2009. Rational inefficiency in fishery. Journal of Business Economics (Zeitschrift für Betriebswirtschaft ZfB), 63–79.

Bogetoft, P., Fare, R., Grosskopf, S., 2009. Rationalizing (in)efficiency through price coordination. Journal of Business Economics (Zeitschrift für Betriebswirtschaft ZfB, 81–86.

Bogetoft, P., Hougaard, J. L., 2003. Rational inefficiencies. Journal of Productivity Analysis 20, 243–271.

Bogetoft, P., Nielsen, K., 2008. DEA based auctions. European Journal of Operational Research 184, 685–700.

Bogetoft, P., Otto, L., 2011. Benchmarking with DEA, SFA, and R. Springer, New York.

Bogetoft, P., Wang, D., 2005. Estimating the Potential Gains from Mergers. Journal of Productivity Analysis 23, 145–171.

Bowlin, W., 1997. A proposal for designing employment contracts for government managers. Socio-Economic Planning Sciences 31, 205–216.

Bundesnetzagentur, 2007. Bericht der bundesnetzagentur nach § 112a enwg zur einführung der anreizregulierung nach § 21a enwg. Report, Bundesnetzagentur.

Che, Y.-K., 1993. Design competition through multidimensional auctions. RAND Journal of Economics 24 (4), 668–680.

Coelli, T., Estache, A., Perelman, S., Trujillo, L., 2003. A primer on efficiency measurement for utilities and transport regulators. Tech. Rep. 129, World Bank Publications.

Dalen, D. M., 1996. Strategic responses to relative evaluation of bureaus: Implication for bureaucratic slack. Journal of Productivity Analysis 7, 29–39.

Dalen, D. M., Gomez-Lobo, A., 1997. Estimating cost functions in regulated industries under asymmetric information. European Economic Review 31, 935–942.

Dalen, D. M., Gomez-Lobo, A., 2001. Yardstick on the road: Regulatory contracts and cost efficiency in the norwegian bus industry. Tech. rep., Norwegian School of Management.

Demsetz, H., 1968. Why regulate utilities? Journal of Law and Economics 11.

Førsund, F., Kittelsen, S., 1998. Productivity development of norwegian electricity distribution utilities. Resource and Energy Economics 20, 207–224.

Government, T. F., 2007. Verordnung zum erlass und zur änderung von rechtsvorschriften auf dem gebiet der energieregulierung. Germany Teil I Nr. 55, Bundesgesetzblatt.

Laffont, J.-J., Tirole, J., 1993a. A Theory of Incentives in Procurement and Regulation, 1st Edition. The MIT Press.

Laffont, J.-J., Tirole, J., 1993b. A Theory of Incentives in Procurement and Regulation. MIT Press.

Langset, T., 2009. Rundskriv eø 4/2009 om beregning av inntektsrammer og kostnadsnorm for 2010. (In Norwegian) NVE 2009 04925-4, The Norwegian Water Resources and Energy Directorate (NVE).

Lazear, E., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89, 841–864.

Littlechild, S., 1983. Regulation of british telecommunications' profitability: report to the secretary of state. Tech. rep., Department of Industry, London.

Nalebuff, B. J., Stiglitz, J. E., 1983. Prizes and incentives: Towards a general theory of compensation and competition. Bell Journal of Economics 14, 21–43.

Resende, M., 2002. Relative efficiency measurement and prospects for yardstick competition in brazilian electricity distribution. Energy Policy 30 (8), 637–647.

Sheriff, G., 2001. Using data envelopment analysis to design contracts under asymmetric information. Tech. rep., University of Maryland.

Shleifer, A., 1985. A theory of yardstick competition. Rand Journal of Economics 16, 319–327.

Thanassoulis, E., 2000. DEA and its use in the regulation of water companies. European Journal of Operational Research 127, 1–13.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press.

Wunsch, P., 1995. Peer comparison and regulation: An application to urban mass transit firms in europe. Ph.D. thesis, Department of Economics, Universite Catholique de Louvain, 182pp.

Recent titles

CORE Discussion Papers

2012/33   Jean-François MERTENS and Anna RUBINCHIK. Pareto optimality of the golden rule equilibrium in an overlapping generations model with production and transfers.

2012/34   Hélène LATZER and Florian MAYNERIS. Income distribution and vertical comparative advantage. Theory and evidence.

2012/35   Pascal MOSSAY and Takatoshi TABUCHI. Preferential trade agreements harm third countries.

2012/36   Aitor CALO-BLANCO. The compensation problem with fresh starts.

2012/37   Jean-François CARPANTIER and Arnaud DUFAYS. Commodities volatility and the theory of storage.

2012/38   Jean-François CARPANTIER and Christelle SAPATA. Unfair inequalities in France: A regional comparison.

2012/39   Shui-Ki WAN, Shin-Huei WANG and Chi-Keung WOO. Total tourist arrival forecast: aggregation vs. disaggregation.

2012/40   Thierry BRECHET, Yann MENIERE and Pierre M. PICARD. The clean development mechanism in a global carbon market.

2012/41   Nikolaos GEORGANTZIS, Rafael MONER-COLONQUES, Vicente ORTS and José J. SEMPERE-MONERRIS. Theoretical and experimental insights on firms' internationalization decisions under uncertainty.

2012/42   Berardino CESI and Dimitri PAOLINI. Peer group and distance: when widening university participation is better.

2012/43   Arnaud DUFAYS. Infinite-state Markov-switching for dynamic volatility and correlation models.

2012/44   Miguel JARA, Dimitri PAOLINI and J.D. TENA. Management efficiency in football: An empirical analysis of two extreme cases.

2012/45   Guillaume ROELS, Philippe CHEVALIER and Ying WEI. United we stand? Coordinating capacity investment and allocation in joint ventures.

2012/46   Mikel BEDAYO, Ana MAULEON and Vincent VANNETELBOSCH. Bargaining and delay in trading networks.

2012/47   Pierre M. PICARD and Ridwan D. RUSLI. State owned firms: Private debt, cost revelation and welfare.

2012/48   Shin-Huei WANG, Luc BAUWENS and Cheng HSIAO. Forecasting long memory processes subject to structural breaks.

2012/49   Thierry BRECHET, Carmen CAMACHO and Vladimir M. VELIOV. Adaptive model-predictive climate policies in a multi-country setting.

2012/50   Vladyslav NORA and Hiroshi UNO. Saddle functions and robust sets of equilibria.

2012/51   Thomas BAUDIN, David DE LA CROIX and Paula GOBBI. DINKs, DEWKs & Co. Marriage, fertility and childlessness in the United States.

2012/52   David DE LA CROIX and Omar LICANDRO. The longevity of famous people from Hammurabi to Einstein.

2012/53   Marijn VERSCHELDE, Jean HINDRIKS, Glenn RAYP and Koen SCHOORS. School staff autonomy and educational performance: within school type evidence.

2012/54   Thierry BRECHET and Susana PERALTA. Markets for tradable emission permits with fiscal competition.

2012/55   Sudipto BHATTACHARYA, Claude D'ASPREMONT, Sergei GURIEV, Debapriya SEN and Yair TAUMAN. Cooperation in R&D: patenting, licensing and contracting.

2012/56   Guillaume WUNSCH, Michel MOUCHART and Federica RUSSO. Functions and mechanisms in structural-modelling explanations.

2012/57   Simone MORICONI, Pierre M. PICARD and Skerdilajda ZANAJ. Commodity taxation and regulatory competition.

2012/58   Yurii NESTEROV and Arkadi NEMIROVSKI. Finding the stationary states of Markov chains by iterative methods.

Recent titles

CORE Discussion Papers - continued

2012/59   Tanguy ISAAC and Paolo PIACQUADIO. Equity and efficiency in an overlapping generation model.
2012/60   Luc BAUWENS, Giuseppe STORTI and Francesco VIOLANTE. Dynamic conditional correlation models for realized covariance matrices.
2012/61   Mikhail ESKAKOV and Alexey ISKAKOV. Equilibrium in secure strategies.
2012/62   Francis BLOCH and Axel GAUTIER. Strategic bypass deterrence.
2012/63   Olivier DURAND-LASSERVE, Axel PIERRU and Yves SMEERS. Sensitivity of policy simulation to benchmark scenarios in CGE models: illustration with carbon leakage.
2013/1   Pierre PESTIEAU and Maria RACIONERO. Harsh occupations, health status and social security.
2013/2   Thierry BRECHET and Henry TULKENS. Climate policies: a burden or a gain?
2013/3   Per J. AGRELL, Mehdi FARSI, Massimo FILIPPINI and Martin KOLLER. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems.
2013/4   Adel HATAMI-MARBINI, Per J. AGRELL and Nazila AGHAYI. Imprecise data envelopment analysis for the two-stage process.
2013/5   Farhad HOSSEINZADEH LOTFI, Adel HATAMI-MARBINI, Per J. AGRELL, Kobra GHOLAMI and Zahra GHELEJ BEIGI. Centralized resource reduction and target setting under DEA control.
2013/6   Per J. AGRELL and Peter BOGETOFT. A three-stage supply chain investment model under asymmetric information.
2013/7   Per J. AGRELL and Pooria NIKNAZAR. Robustness, outliers and Mavericks in network regulation.
2013/8   Per J. AGRELL and Peter BOGETOFT. Benchmarking and regulation.

Books

G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en heritage*. Itinera Institute.
M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.
V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.