Regulatory Benchmarking
Models, Analyses and Applications

Agrell, Per J.; Bogetoft, Peter

Link to publication in CBS Research Portal

Regulatory Benchmarking: Models, Analyses and Applications

Per J. Agrell and Peter Bogetoft

Regulatory benchmarking: Models, analyses and applications

Per J. Agrell[a], Peter Bogetoft[b,*]

[a]*Louvain School of Management and CORE, Universite catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.*
[b]*Copenhagen Business School, Porcelaenshaven 16 A, DK-2000 Frederiksberg, Denmark, and Yale School of Management, 35 Prospect Street New Haven, CT 06511, United States*

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:
DEA, agency theory, regulation, energy networks.

---

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.

---

*Corresponding author. Phone: +45 2332 6495 or +1 203 393 5915
Email address: `pb.eco@cbs.dk` (Peter Bogetoft)

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 (1993). 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) 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), where we show the optimality of DEA based yardstick schemes. 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 (1993).

Key references to the practical combination of benchmarking and regulation are Agrell and Bogetoft (2001), Agrell and Bogetoft (2010) and Coelli et al. (2003).

There a small but important literature on the combination of incentives and benchmarking. This constitutes the theoretical foundation of DEA based 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 per-

formance 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). A survey of the main insight is provided in Bogetoft and Otto (2010) and Bogetoft (2012).

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.

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 private information is a key issue in the regulatory game and regulators can use benchmarking to undermine the informational asymmetry. The firms have superior information about their own activities and abilities, and they may therefore have incentives to supply insufficient effort or hide relevant cost information. By comparing the firms using benchmarking the regulator can alleviate these problems. In some situations the regulator may even learn more about the general available technological possibilities than the firms since the regulator collects such information from multiple firms.

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.

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

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

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.

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.
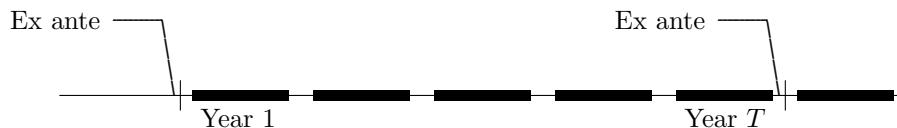
Figure 1: Ex ante regulation

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 4 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.
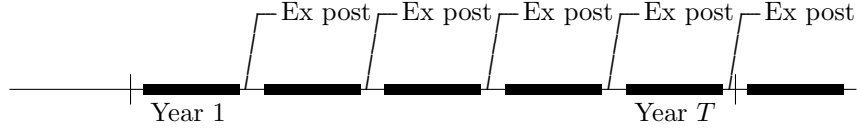
Figure 2: Ex post regulation

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.

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

8

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 extend this scenario to situations with heterogenous bids, i.e. as when the bidders offer different service profiles, 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.

3. Use of regulatory benchmarking

Table 1 below gives a summary of the benchmarking methodologies used for electricity DSOs and TSOs in 22 European countries. It is based on Haney and Pollitt (2009) with our updates for the period after 2008.

In terms of the regulatory regime, the use of a revenue cap is clearly the dominant approach. Dynamically, the progression seems to be from more heavy-handed cost recovery regimes, passing through a period of model-based price fixation towards a high-powered market-based yardstick regime. No country has so far relied heavily on concession auctions when it comes to the network activities. A main reason is that the transfer of network ownership from an incumbent operator to a new operator is likely very problematic due to asymmetric information. Bidding is however used heavily in other parts of the European energy sector, e.g. when licenses for building new wind farms are handed out.

It is important to understand also that most regulations have elements of different regimes. The dominant regime may be a revenue cap, but in all cases, it is only part of the cost base that is actually regulated. Another part of the costs are,

within reasonable bounds, simply accepted as so-called pass-through costs. This happens for example in many cases when it comes to the cost of new activities that the political parties agree on, e.g. regarding energy saving initiatives. To identify the exact share of cost pass-trough requires detailed knowledge of the regulatory specificities. In the cases we have detailed knowledge of, it is not uncommon that only 60-70 percent of the actual costs are being incentivized. In some cases, where only Opex and not Capex is regulated in the revenue cap, it is of course less.

In terms of the method used, it is typically DEA when it comes to the TSO benchmarking and a combination of DEA, SFA, MOLS and a series of ad hoc approaches that are used in the DSO regulation. The newest addition to the list of techniques applied is the StoNED approach that have recently been adopted by the Finish regulator EMV, cf. Kuosmanen (2012). We see that a few countries, like Spain and previously Sweden, rely on technical engineering norms, sometimes referred to as ideal nets or reference network model, in an attempt to identify not only best practise but also in some sense the absolute technological possibilities. Other countries, like Germany, have supplemented the development of for example DEA and SFA models with partial engineering models to gain insight into the relative importance of different cost drivers and net characteristics.

4. Application 1: 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).

4.1. 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*".

Table 1: Some European regulation regimes and cost function methodologies for electricity distributors (DSO) and transmission operators (TSO). Participation in benchmarking at a nat[ional] or int[ernational] level without direct implementation in regulation is denoted by *.

Country	Regime	Method DSO	Method TSO
Austria	Revenue cap	DEA-MOLS(nat)	DEA(int)*
Belgium[1]	Revenue cap	DEA(nat)*	DEA(int)
Denmark	Revenue cap	MOLS(nat)	DEA(int)
Estonia	Revenue cap	MOLS(nat)	DEA(int)*
Finland	Revenue cap	StoNED(nat)	DEA(int)
France	Cost recovery	Ad hoc	DEA(int)*
Germany	Revenue cap	DEA-SFA(nat) best-of	DEA(int)
Great Britain	Revenue cap	MOLS(nat)	DEA(int)*
Greece	Cost recovery	-	DEA(int)*
Hungary	Price cap	Ad hoc	Ad hoc
Iceland	Revenue cap	Ad hoc DEA(int)*	DEA(int)
Ireland	Price cap	Ad hoc	Ad hoc
Italy	Revenue cap(opex)	Ad hoc	DEA(int)*
Lithuania	Price cap	Ad hoc	DEA(int)*
Luxemburg	Cost recovery	Ad hoc	DEA(int)*
Netherlands	Yardstick	MOLS(nat)	DEA(int)
Norway	Yardstick	DEA(nat)	DEA(int)
Portugal	Revenue cap	SFA(nat)	DEA(int)
Spain	Revenue cap	Engineering	DEA(int)*
Slovenia	Price cap	DEA(nat)	-
Sweden	Rate-of-return	Ad hoc DEA(nat)*	DEA(int)*
Switzerland	Cost recovery[2]	Ad hoc DEA(nat)*	-

11

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.

4.2. 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

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

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. It is interesting that German regulation as regulation is many countries use special add on procedures to deal with expansions that take place within a regulation period. From an academic point of view, it seems that the established models can be directly used to measure the efficient cost impact of changing the output profile, and that the extra costs allowances could therefore be allocated based hereon in a way that is more consistent with the rest of the regulation.

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.

13

4.3. 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

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

The use of such a cautious approach comes, of course, at the cost of customers who pay a price for the bias. It may not be immediately obvious why the customers should bear the risk of estimation error. An alternative could be to spread the risk by specifying an unbiased estimator that would not favor either the DSOs or the customers. From a theoretical perspective, we shall later see that it may not be optimal to use an unbiased estimator. Truly, DEA is biased but still one can make optimal contracts on DEA in some cases, cf. Bogetoft and Otto (2010) and Bogetoft (2012). More importantly, however, it is not clear if the risk of the consumers are best minimized by using an unbiased estimator or a cautious approach like the German. The cost to consumers from interruptions of electricity supply that may result from a too harsh regulation may be more important than the cost savings possible.

The use of the cautious approach can also be challenged as an inconsistency. In Finland, the regulator previously applied the average of DEA and SFA efficiency scores, but in the recent review their decided in favor of the StoNED estimator. Kuosmanen (2012) refers to the previous practice as 'naive model averaging', and they criticize the parallel use of estimators that are based on conflicting assumptions or specifications. We do not share this criticism. As long as benchmarking scholars cannot clearly rank one method as being superior to another, we see no

reason the regulator should make that call. It is also not just "an easy way out" of methodological discussion to apply multiple methods. In fact, one can argue that it makes life easier for the regulator and the model developers to only have to relate to one set of assumptions and one set of results, and that the simultaneous application of multiple methods puts additional discipline on the model development approach.

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. To the best of our knowledge, this is generally the case in European regulations. Although many regulators have done back-office calculations of the measures and they may well have guided the decision making in many cases, the regulations tend not to use such measures in a direct way.

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).

15

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.

4.4. 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.

4.5. 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. 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.

4.6. Final model

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

Table 2: German DEA-SFA model for electricity DSOs. Agrell and Bogetoft (2007)
.

Input	Outputs (cost drivers)
Total costs:	Connections.hs.ms.ns
Totex	Lines.circuit.ms
or	Lines.circuit.hs.share.cor
Totex.standard	Cables.circuit.hs.share.cor
	Cables.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

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.

5. Application 2: 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

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

where R^k is the revenue cap, C^k_{DEA} 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^k_{DEA} 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, owner-ship 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.

6. Application 3: International benchmarking of electricity transmission

Transmission services, the backbone of the energy infrastructure, is an extreme example of a natural monopoly at national level. The regulators here face the classical problem of information asymmetry, but without the possibility of bridging the gap using national observations as in the previous applications for distribution system operators. For these purposes, the national regulatory authorities (NRA) in Europe under the premises of the Council of European Energy Regulators (CEER) have regularly organized international benchmarking studies of transmission system operators (TSO) that are put in actual practice by at least part of the participating NRAs, see Agrell and Bogetoft (2014) for a more indepth account.

Already the international organization of regulatory benchmarking is rare, besides Europe we notice occasional projects in Latin America (Estache et al. (2004)) for electricity distribution. The initial work by the European Commission was focused at relatively simple OLS-studies of final tariffs, such as Perez-Arriaga et al. (2002) or specific asset cost comparisons s (cf. ICF Consulting (2002) for a study oin 300 kV transmission lines). A desktop study using a two-output single-input DEA model was applied to seven TSO for the period 1999-2005 von Geymueller (2009).

The initial benchmarking studies in 2003, 2005 and 2009 (cf. Agrell and Bogetoft. (2009)) developed from used partial unit cost measures (e.g cost per normalized grid unit) towards the application of frontier analysis tools such as DEA. The findings of the studies were proven reliable for application in several appeals of regulatory rulings. Notwithstanding, most NRA use and intend to use international benchmarking as complementary information to regulatory supervision Haney and Pollitt (2013). However, as some countries as Germany, the Netherlands, Portugal and Denmark use international benchmarking actively for incentive regulation, some critique has been raised regarding the stability and reliability of such assessments Brunekreeft (2013). Here, we will examine some specificities and key aspects of the international regulatory models for electricity transmission.

6.1. Specificities

The objective of a regulatory assessment in transmission is to obtain robust estimates for the efficient cost of structurally comparable operators. Contrary to distribution system studies, the transmission studies immediately face all dimensions of heterogeneity: investment tenure and activation, non-standardized task scope, environmental and topological factors, climate, financial structure, current and past technical regulation, et c. In addition, even with a strong adherence in the latter studies (19 of 28 countries in 2012), the reference set is too limited to address heterogeneity exclusively by econometric means. Hence, a substantial effort in international benchmarking is devoted to *ex ante* activity analysis and cost allocation using activity-based accounting systems. Seven core activities are distinguished in the studies following Agrell and Bogetoft. (2009):

- X *Market facilitation* (management cost for and/or interventions on electricity exchanges)

- S *System operations* (maintenance of the real-time energy balance, congestion management, and ancillary services such as disturbance reserves and voltage support)

- P *Grid planning* (planning and drafting of grid expansion and network installations involving the internal and /or external human and technical resources, including access to technical consultants, legal advice, communication advisors and possible interaction with governmental agencies for pre-approval granting.)

- C *Grid construction* (tendering for construction and procurement of material, interactions, monitoring and coordination of contractors or own staff performing ground preparation, disassembly of potential incumbent installations, and recovery of land and material.)

- M *Grid maintenance* (preventive and reactive service of assets, the staffing of facilities and the incremental replacement of degraded or faulty equipment)

- A *Administration* (administrative support and associated costs include the non-activated salaries, goods and services paid for, central and decentralized administration of human resources, finance, legal services, public relations, communication, organizational development, strategy, auditing, IT and general management.)

- F *Grid financing* (long-term financing of the assets through equity and debt)

Frontier analysis is a relevant approach only to activities that are relatively homogeneous in process and controllable by the evaluated entities. The activities X, S, P and F above are all associated with problems linked to the horizon of observability (planning incurs cost today for outputs in the future), to exogenous market regulation (market facilitation costs X depend on the national rules for e.g. primary and secondary reserves, non-standardized in Europe by 2015), and to non-controllable factors (governance and national credit ranking have major influence on the financial costs F). Consequently, the transmission DMU is essentially a 'wire-company' only operating assets for the transport of electricity using high voltage equipment.

A second specificity relates to the orientation of the model. A transmission network basically performs three types of services: transport of energy (delivered energy), provision of transmission capacity (transformers) and customer service (information, billing, connection etc). Intuitively the grid (the capital) is then an input to achieve some of the tasks. Since the retail sales of the commodity (energy) is not among the competencies of the TSO, the final demand is exogenous. Thus, an input-orientation formulation is promoting a utilization metric, where the smallest or least costly grid asset transmitting the highest volume will be considered benchmark peer. Given an an asset-intensive activity like electricity transmission with high societal costs of interruption and unverifiable quality *ex ante*[3], such partial production possibility cannot be used for regulation. Since the TSO may to a certain extent mitigate capacity problems by system operations, the exogenous character of the capacity provision service could be questioned. Since the elasticity of the cost function is low with respect to pure volume changes (basically only energy losses in the network, about five per cent of the overall operating expenditure), such model becomes disproportionally dependent on a correlated and lagged output factor. Moreover, a utilization metric would explicitly contradict overarching system objectives for the decarbonization of the European energy system through the creation of 'electricity highways' to promote higher penetration of renewable energy resources and the use of electric vehicles.

In consequence, the capacity provision is normally modelled through the grid in itself or proxies thereof. On the one hand this naturally improves the fit of the cost function, since a proxy of the capital is included as an output. On the other hand it limits the model detecting non-grid solutions to congestion problems

---

[3]Quality, measured as the System Average Interruption Duration Index (SAIDI), the System Average Interruption Frequency Index (SAIFI) or the [volume] of Energy Not Supplied (ENS), are all *ex post* measures of service quality, most of which are very low or zero at the transmission level for European continental transmission grids. The quality indicators are largely influenced by stochastic events (weather, seismic), the cost causality is severely lagged in time and their use in a static model is not advisable.

25

(which may be relevant to overall effectiveness rather than efficiency) and gold-plating investments not called for. The latter problem is highly hypothetical since the return on capital even in the case of efficient firms is claimed to be low and discouraging (cf. Glachant et al. (2013)).

6.2. Model

Raw accounting data from the TSOs cannot be used, since depreciation periods differ, rules for activations vary and the use of nominal capital values would induce an age bias in the analysis. Thus, all investments are transformed to real annuities using standardized asset life times in nine groups, T_j. Investment data is aggregated by year, a weighted investment life time $\tilde{T}_{it}$ is calculated for each TSO and year.

$$\tilde{T}_{it} = \sum_j \left\{ \frac{T_j w_j \left(A_{ijt} - A_{ij,t-1} \right)}{\sum_j w_j \left(A_{ijt} - A_{ij,t-1} \right)} \right\} \tag{1}$$

The annuity factor α transforms lumpy investment patterns into smooth capital expenditure in real terms.

$$\alpha_{it} = r \left(1 - (1+r)^{-\tilde{T}_{it}} \right)^{-1} \tag{2}$$

The operating expenditure (opex) for each TSO is calculated as the sum of non-staff and labor-cost adjusted staff costs, all in real terms and in EUR.

$$VC_{it} = \left(V_{it}^N + V_{it}^S L_{it} \right) PI_{it} \xi_{it} \tag{3}$$

The capital expenditure (capex) is defined as the sum of real investment annuities.

$$FC_{it} = \xi_{it} \sum_{k=1}^{t} I_{ik} PI_{ik} \alpha_{ik} \tag{4}$$

The main output variable is a proxy for the grid provision task, the normalized grid. In Frontier Economics et al. (2013), power system engineers estimated a system of 2169 relative weights for opex v and capex w spanning over assets organized in nine groups (lines, cables, transformers, etc) and four categories (voltage, power, cross-section, short-circuit current). The grid output in terms of opex NGV is defined from a weighted set of assets in use, where R_v calibrates the average normalized opex to the average opex for a reference year.

$$NGV_{it} = R_v \sum_j v_j a_{ijk} \tag{5}$$

The normalized grid capex NGF is defined as a weighted sum of capex-active (younger than T_j) assets using real annuities, calibrated using a constant R_f to average capex in the sample.

$$NGF_{it} = R_f \sum_{k=1}^{t} \sum_{j} w_j A_{ijk} \alpha_{ik} \tag{6}$$

Finally, the normalized grid totex is the sum of opex and capex, analogously to the costs.

$$NG_{it} = NGF_{it} + NGV_{it} \tag{7}$$

6.3. Model specification

The common parameters for the study are defined in Table 4 below. CPI was chosen as inflation adjustor due to data availability, but tested in the sensitivity analysis below. A large set of environmental parameters, Y-parameters, was collected, whereof finally was retained $DenseArea$ as the total area (in km^2) of highest population density in EUROSTAT as a proxy for density-induced operating cost increases, and $AngleLineSum$, the totex-weighted linelength of angular towers, as a proxy for the extra costs induced by routing complexity (infrastructure, topology, urban sprawl). Some descriptive statistics are presented in Table 5.

Table 4: Parameters, e^3GRID 2012, Frontier Economics et al. (2013)

Parameter	UoM	Value	Source
r	%	4.36	
PI	index	CPI	EUROSTAT
R_v	constant	1.0808	Org data
R_f	constant	0.8801	Org data
L	index	EU salary index	EUROSTAT
ξ	EUR	(average 2011)	EUROSTAT

Table 5: Descriptive statistics 2007-2011, $n = 102$

Variable	Pctl(25)	Mean	Pctl(75)
TC	89229250.6	315085413.5	265250347.2
NG	104990.6	335631.4	302842.7
$DenseArea$	414.6	5251.7	5253.2
$AngleLineSum$	11372.9	37758.1	35628.6

After extensive model specification tests using various cost function approaches, a final cost function in log-linear functional form was found and documented in Table 6 below. As seen, the coefficients with expected signs are strongly significant both in normal OLS and in robust OLS. The model has a high explanatory value, is not subject to heteroscedasticity (p-value for Breusch-Pagan test 0.097) and the multicollinearity is low (maximum variance influence factor, VIF $8.3 \leq 10$).

Table 6: Results for Totex cost model, OLS and robust OLS

	OLS	Robust OLS
$\log(NG)$	0.554***	0.475***
	(0.096)	(0.052)
$\log(DenseArea)$	0.117***	0.137***
	(0.011)	(0.009)
$\log(AngleLineSum)$	0.217**	0.284***
	(0.083)	(0.040)
Constant	9.233***	9.477***
	(0.516)	(0.338)
Observations	102	
Adjusted R^2	0.912	
Residual Std. Error	0.351 (df = 98)	
F Statistic	349.668*** (df = 3; 98)	
Breusch-Page test (p-value)	0.097*	
Multi collinearity test (max VIF)	8.3	
Note:	*p<0.1; **p<0.05; ***p<0.01	

As the sum of the coefficients in the loglin cost function is less than unity, the cost function exhibits non-decreasing returns to scale, which is used in the DEA model below.

6.4. Robustness measures

Robustness with respect to the reference set and the possible influence of outliers is especially important for transmission system benchmarking where the heterogeneity is high, the sample set is small and the stakes are high. In the model

specification phase, before the environmental factors are determined, the robustness is ensured using robust regression[4]. An interesting example here is the selection of the final model compared to an alternative model with the system peak load (in MW) as a third output parameter. The results of the estimation of this model for a loglinear case are given in 7 below. The statistical properties are similar to those of the final model in 6, although the alternative model fails in the Breusch-Pagan test for heteroscedasticity. However, the model is rejected also on conceptual grounds: the negative sign (both in normal and robust OLS) indicates that the peakload parameter in this model does not form a natural and monotonous cost function. Indeed, the peak load is correlated to part of the asset base (transformers) which are already covered in the normalized grid output, creating a collinearity between the two parameters. In contrast, the finally retained parameter $AngleLineSum$, is reflecting the the topology through the angular towers, which is not covered by the normalized grid parameter. Following the model specification phase, the actual estimations are performed using the outlier detection techniques from the German Ordinance, also discussed in Agrell and Bogetoft (2014) to detect what was revealed as three outlier TSOs in terms of performance. The final results were also subject to extensive sensitivity analysis with respect to various parameters that potentially are influential in the model, such as the inflation adjustment factors (here: CPI). The impact in this particular case is limited, the average efficiency decreases from 86 % to 84 % when a producer-price index (PPI) is used rather than the CPI.

6.5. Weight restrictions in DEA

As seen from the parametric results, the main cost driver is the normalized grid, complemented by the two environmental proxies. To avoid nonsensical nonparametric results in DEA, a restriction for the dual weight is defined as cones around the coefficients above. This procedure assures that the marginal costs are comparable to those of the average firm. The approach also corresponds to regulatory practice in the sense that it endogenizes the parameters for the weight restrictions, otherwise left at the discretion of analysts or regulators. The robustness of the weight restriction approach is validated by complementary analyses using robust regression for the coefficients, as well as validation with larger data sets.

6.6. Dynamic analyses

The international benchmarking of TSOs is based on a time-series analysis, where a Malmquist-approach for the model above is used to derive estimates for

---

[4]The robust regression was performed using Iterative Reweighted Least-Squares regression (IRLS) by Holland and Welsch (1977)

Table 7: Results for alternative Totex cost model, OLS and robust OLS

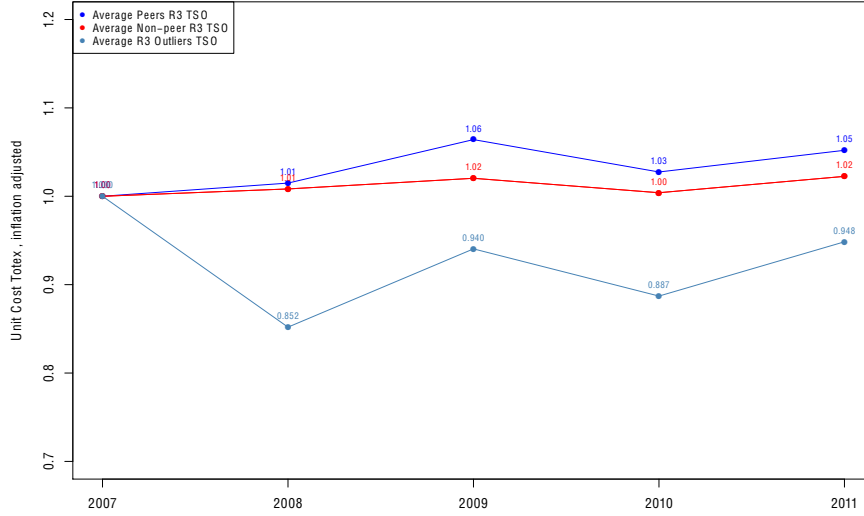|	OLS	Robust OLS
$\log(NG)$	0.9414***	1.006***
	(0.050)	(0.076)
$\log(DenseArea)$	0.1488***	0.1354***
	(0.008)	(0.012)
$\log(PeakLoad)$	-0.1936***	-0.2567***
	(0.051)	(0.074)
Constant	8.344***	8.170***
	(0.270)	(0.466)
Observations	102	
Adjusted R^2	0.916	
Residual Std. Error	0.343 (df = 98)	
F Statistic	368.6*** (df = 3; 98)	
Breusch-Page test (p-value)	0.101	
Multi collinearity test (max VIF)	6.4	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 3: Unit cost development, peers, non-peers and outliers, 2007-2011.

efficiency change, technical change (frontier shift) and productivity improvements in the sector. The obtained results for the technical change (TC) are applicable in the regulation as the so-called general X-factor applied to all firms. Interestingly for the period 2007-2011, TC was determined to be negative, whereas the efficiency change (catch-up) was significant and positive. More detailed analysis, as in Fig 3 shows that contrary to the period 2003-2006, the peer units on average increase their total expenditure and only the outliers (detected as previously described) show significant cost decreases. These and other findings prompted informative exchanges among the TSOs and NRAs on the reasons for the productivity slow-down among the efficient TSOs.

7. 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 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 22 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.

In a similar way, we covered in some details the regulatory benchmarking of European TSOs. We here emphasized the challenges of having to do international comparisons of a small sample of firms that do not have the same scope of operations. We showed have careful delineation of comparable tasks followed by extensive data standardization and the heavy use of additional engineering cost information is necessary to make DEA modelling robust and relevant.

Let us close by pointing to a selection of issues that we consider relevant in future research on regulatory benchmarking.

One area is the combination of regulation with firm specific learning. Firms that are subject to regulatory benchmarking spend considerable resources collecting and standardizing data, and it is worthwhile to consider what added value this effort may bring to the daily operations of the firms. It can be useful to use the same data to support firm specific learning. We have run some such projects in parallel to the regulatory benchmarking exercises, and in the Danish water sector, the regulator supports and require such additional sector applications of the models, cf. Bogetoft (2012), but more analysis of this is clearly needed.

Based on our work on numerous regulatory projects, it also seems clear that we need more research on how to combine formal agency theory and DEA. This line of research is challenging since it combines two academic traditions, namely the detailed production economic approach of DEA with the more abstract micro

economic agency theory. A particular issue is how to compare and calibrate the incentive powers in SFA and DEA based schemes. A related but more of an empirical theme is how efficient different regulations really are. There are very few studies of the actual impact of introducing different types of benchmark based regulations.

In regulatory applications, it would also be useful with a better understanding of how to take a cautious approach without leaving excessive rents to the companies. In what sense is a best-off approach for example superior to an approach based on a single model or a single (DEA, SFA, etc) estimation technique? Making correct evaluations in practical applications, where the data and model specification may be incomplete, is complicated. In a recent paper, Banker and Zhang (2016), challenged regulatory use of constant or increasing (non-decreasing) returns to scale (CRS or IRS(NDRS)). Assuming that the underlying technology is in fact say IRS, it may still be better to assume VRS as a protection against misspecification of the model. Following their logic, then, it is insufficient to only use conceptual or natural monopoly arguments to motivate say an IRS assumption. Empirical tests should ideally be performed on the actual model specification. Specifically, we can expect the VRS assumption to be superior if the model does not properly capture the existence of different areas or customers with higher marginal costs than others.

Last but not least, there is a need for more serious academic discussions of what a good regulatory benchmarking model actually is. As illustrated in our cases, the actual choice of a model invokes a series of conceptual, statistical, and pragmatic criteria, and it is not clear how to prioritize and weight different concerns.

References

Agrell, P., Bogetoft, P., May 2014. International benchmarking of electricity transmission system operators. In: European Energy Market (EEM), 2014 11th International Conference on the. pp. 1–5.

Agrell, P. J., Bogetoft., March 2009. International benchmarking of electricity transmission system operators: e3grid project. Final report for ceer, SUMICSID Group.

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

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., Zhang, D., 2016. The returns to scale assumption in incentive rate regulation. Tech. rep., Temple University.

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.

34

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., 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., 2010. Benchmark package. Tech. rep., R.

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.

Brunekreeft, G., 2013. On the role of international benchmarking of electricity transmission system operators facing significant investment requirements. Competition and Regulation in Network Industries 14 (1), 2–24.

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

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.

Estache, A., Rossi, M., Ruzzier, C., 2004. The case for international coordination of electricity regulation: evidence from the measurement of efficiency in south america. Journal of Regulatory Economics 25 (3), 271–295.

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

Frontier Economics, Consentec, Sumicsid, 2013. E3grid 2012 - european tso benchmarking study. Report for ceer, Frontier Economics and Consentec and Sumicsid.

Glachant, J.-M., Saguan, M., Rious, V., Douguet, S., 2013. Incentives for investments: Comparing eu electricity tso regulatory regimes. Research report, European University Institute, Florence School of Regulation.

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

Haney, A., Pollitt, M., 2009. Efficiency analysis of energy networks: An international survey of regulators. Energy Policy 37 (12), 5814–5830.

Haney, A. B., Pollitt, M. G., 2013. International benchmarking of electricity transmission by regulators: A contrast between theory and practice? Energy Policy 62, 267–281.

Holland, P. W., Welsch, R. E., 1977. Robust regression using iteratively reweighted least-squares. Communications in Statistics-Theory and Methods 6 (9), 813–827.

ICF Consulting, October 2002. Unit costs of constructing new transmission assets at 380kv within the european union, norway and switzerland. Final report for dg tren, ICF Consulting.

Kuosmanen, T., 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the stoned method in the finnish regulatory model. Energy Economics 34 (6), 2189 – 2199.

Laffont, J.-J., Tirole, J., 1993. 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.

Perez-Arriaga, I. J., Montero, F. P., Oderiz, F. J. R., October 2002. Benchmark of electricity transmission tariffs. Report for dg tren, Universidad Pontifica Comillas.

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.

von Geymueller, P., 2009. Static versus dynamic dea in electricity regulation: the case of us transmission system operators. Central European Journal of Operations Research 17 (4), 397–413.

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.