Agrell, Per J.; Grifell-Tatjé, Emili. «A dynamic model for firm-response to non-credible incentive regulation regimes». Energy policy, Vol. 90 (2016), p. 287-299. DOI 10.1016/j.enpol.2015.12.029

A Dynamic Model for Firm-Response to Non-Credible Incentive Regulation Regimes

Per J. Agrell

Université catholique de Louvain
Center of Operations Research and Econometrics (CORE)
B-1348 Louvain-la-Neuve, Belgium.
Email: per.agrell@uclouvain.be

Emili Grifell-Tatjé

Universitat Autonòma de Barcelona,
Department of Business
ES-08193 Bellaterra, Spain.
Email: emili.grifell@uab.cat

Abstract

Economic network regulation increasingly use quantitative performance models (from econometrics and engineering) to set revenues. In theory, high-powered incentive regulation, such as revenue-caps, induces firms to cost-efficient behavior independent on underlying model. However, anecdotal evidence shows regulated firms occasionally maintaining cost-inefficiency under incentive regulation even under slumping profitability. We present a model for firm-level efficiency under a regime with a probability of failure explaining this phenomenon. The model is based on the hypothesis that the regulatory choice of method can be associated with intrinsic flaws leading to judicial repeal and replacement of it by a low-powered regime. The results show that the cost efficiency policy is proportional to the type of firm (cost of effort), value of time (discount factor) and the credibility of the method (risk of failure). A panel data set for 2000-2006 for 128 electricity distributors in Sweden is used to validate the model predictions (radical productivity slowdown, failing profitability and efficiency) at the launch and demise of a non-credible regulation method. The work highlights the fallacy of viewing incentive regulation as a method-independent instrument, a result applicable in any infrastructure regulation.

Final version published in *Energy Policy, vol 90*
(March), p. 287-299. DOI: 10.1016/j.enpol.2015.12.029

A Dynamic Model for Firm-Response to Non-Credible Incentive Regulation Regimes

Per J. Agrell

Université catholique de Louvain
Center of Operations Research and Econometrics (CORE)
B-1348 Louvain-la-Neuve, Belgium.
Email: per.agrell@uclouvain.be

Emili Grifell-Tatjé

Universitat Autonòma de Barcelona,
Department of Business
ES-08193 Bellaterra, Spain.
Email: emili.grifell@uab.cat

1. Introduction

Inefficient operations, imprudent expenditures, low staff productivity and excessive investments by regulated firms are all classical indications that the sector regulation is inadequate and in need of reform. Conventional wisdom would focus at the *incentive power* of the regulation, arguing that the methods and paths to reach a new situation are irrelevant to the final welfare effects. In this paper, we show theoretically that this may not be in the case of an imperfect regulation method and strategic firms. We will also estimate this phenomenon empirically with a data set for the detailed firm response under a disputed change in regulatory regime. Although the strategic player in our model is the firm, the policy backdrop for our paper is robust regulatory design.

Regulatory authorities attempt to achieve the dual objectives of assuring a comprehensive, continuous and environmentally compatible service as well as controlling for rent extraction through excessive direct tariffs or by discriminatory pricing of access to impede competitive entry. The National Regulatory Authorities (NRA) define the business perspectives of the regulated operators affecting the operations and economic conditions at a given time. But beyond this, their rulings also signal their commitment for future investments, entry and development of operators. The underlying task is further complicated by the existence of multi-output production (capacity provision, transport work and customer services) and heterogeneous input conditions (specific assets, geographical and systemic constraints, different interfaces) under a steady technological development. The NRA is facing an evident asymmetry of information with respect to the capacity, cost and capabilities of the regulated entities. This excludes a naive direct command and control approach to regulation, leaving the room to the traditional economic regulatory approaches; low-powered cost-recovery and high-powered incentive regulation, cf. Joskow (2011).

This paper contributes to the energy policy literature by challenging the conventional conclusions from regulatory economics according to which only the commitment and not the method of determination count for the incentive effects in regulation. We review the existing diverse methods in energy infrastructure regulation, including the models used for determining efficient costs. Noting that some failures can be linked to specific features in the regulatory models, we conjecture that rational firms observe the weaknesses of such models and anticipate their failure. We explore the properties of the model, derive optimal policies that extend the intuition for incentive regulation and state a set of formal results from the model. In competitive markets, an analogy can be made with the credit risks. It is there a well-known result that the incentives for client-specific investments decrease with increasing bankruptcy risks. The regulation literature is rarely addressing client or regulatory failures. We have not found any paper modelling the lack of credibility that could undermine the trust in a regulatory regime. We model this credibility as a failure probability and provide empirically verifiable hypotheses that apply to the case of energy network regulation. Using panel data for electricity distributors and a narrative for the failed introduction of a regulatory method in Sweden, we then validate the hypotheses. Our model, combined with a rigorous economic methodology to decompose the drivers of profitability, provides a rational explanation for the behavior of the firms and the demise of the regime. A second contribution lies in the detailed analysis of the sector-level impact in terms of profitability, productivity and efficiency resulting from a regulatory policy error of this type. To our knowledge, both the theoretical and the empirical contribution are seminal in the literature.

The outline of the paper is as follows. In the Section 2, we review energy network regulation models, in particular for electricity distribution. In Section 3 we present a decision model for firm-level response to non-credible regulation. In Section 4 we provide the methodological framework for the productivity analysis to test the model. In Section 5 we provide a narrative for the Swedish case, followed by Section 6 on the data and the activity model used for the estimation. Section 7 contains the analysis of a number of hypotheses derived from the theoretical results. The paper is closed in Section 8 with a discussion and some policy conclusions.

2. Regulation of electricity distribution

In Europe, the preamble to the Third Energy Market Directive European Commission (2009) implicitly supposes that incentive (high-powered) regulation of the revenue-cap type is implementable and effective for network regulation of the distribution system operators (DSO) and the transmission system operators (TSO). Empirical evidence shows that DSOs achieve cost savings above any *a priori* expectations, in particular in combination with private or mixed ownership (Cambini and Rondi, 2010). However, although the Directive prescribes the application of a well-defined method *ex ante*, no supra-national model is defined by the legislator. The detailed modus of regulation is then left to national legislators to decide. In practice, this has resulted in heterogeneity with respect to the modes, models and methods used by NRAs, see Haney and Pollitt (2009, 2013) for recent surveys.

The decentralized regulatory regimes for independent firms or decision makers are conventionally classified in order of their *incentive power* as either low- or high-powered regulation mechanisms. The incentive power represents the share of 1$ cost savings or cost increase that the firm can retain or must absorb, respectively. Cost-plus regulation is the classical low-powered alternative, with incentives for over-investment and inefficiency. Rate-of-return regulation is currently found in many countries, including the United States, as a low-powered option that determines the financial return of the industry. Here, any capital investment is covered by the regime, but operating expenditure should be capped under an allowance proportional to the regulatory asset base. The seminal work by Averch and Johnson (1962) points out the incentives for overcapitalization to increase the rate base under this regime.

The seminal RPI-X price cap from Littlechild (1983)) and its revenue cap variants are examples of high-powered regimes. Incentive regulation is widely applied to electricity transmission and distribution in Europe, e.g., in England and Wales (Pollitt, 1995). Liston (1993) shows that the fixed income induces cost efficiency by the agent's cost minimization. However, the CPI-X model is associated with several theoretical and practical problems; strategic behavior on behalf of the agents, fearing punishment in subsequent periods for productivity improvements, the ratchet effect (Weitzman, 1980; Freixas et al., 1985); sensitivity to, and lack of foundation for, the X-term (Cambini and Rondi, 2010); and inability to accommodate changes in the output profile.

The yardstick competition regime (Shleifer, 1985) is a promising addition to the regulatory arsenal. The main problem of the basic yardstick model is the comparability between agents and in particular its inability to accommodate variations in the output profiles and operating conditions between the agents, cf. Agrell et al. (2005a). Yardsticks also rely on econometric modeling in general, which requires skills and care not to introduce other errors and bias, cf. Cronin and Motluk (2007).

One instrument in the regulatory arsenal is the *engineering cost model*, also known as a technical normative model. In electricity distribution operations, engineering cost models are, or have been active, in Chile (Rudnick and Donoso, 2000; Recordon and Rudnick, 2002), in Spain (Grifell-Tatjé and Lovell, 2003) and in Sweden, as discussed Section 5. A normative cost model is based on an attempt to come closer to the true production frontier, or to draw on other information than merely the observations. The concept is tempting in regulation because of its potential profit reduction possibility and its integration in yardstick regulation. However, given the high cost of failure and service interruption in network services, the issue of feasibility in the normative estimation is primordial. In general, technical normative models are just special cases of engineering cost functions with varying level of information requirements. As such, they are used to prescribe rather than predict the optimal, or allowable, cost for a certain level of operation. Thus, the model's estimate can be made feasible by parameterization and construction.

Our model addresses the risk of regulatory failure explicitly. Our empirical illustration is based on the engineering cost model in Sweden, but other examples of failed regulation models have been documented in the Netherlands (Nillesen and Pollitt, 2007) and in Belgium (Agrell and Teusch, 2015). The qualitative findings in these cases are consistent with our predictions and results.

Theoretically, any implemented model will evoke the same response from the regulated firms. However, we argue that high-powered models must be economically and technically sound to result in the intended incentive effects. In the next section, we present a model to explain why firms, not even subject to the potential flaws of a given high-powered model, have incentives to stall investments in cost efficiency improvements under regimes based on unstable methods.

3. Strategic gaming in the failure of incentive regulation

The underlying assumption in incentive regulation is that the firm is profit maximizing. Facing an exogenous demand y for services, having private information about a cost function $c(y)$, the firm is maximizing profit $\pi(y) = R(y) - c(y)$ by minimizing cost if and only if the revenue function $R(y)$ is exogenously fixed. Theoretically and empirically, the effects of high-powered regulation are at large validated (Cambini and Rondi, 2010). The idea in this paper is that the *credibility* of the regulatory model is crucial to firm level reactions, in particular in jurisdictions with judicial recourse. To be a credible regulation regime, it has to assure *participation* and *information revelation*. The first condition implies that the regime should not force any efficient operator to exit. The second condition means that the regime should be able to generate all necessary information for its execution by information revelation, as opposed to strategic signaling or bargaining. Note that the failure of a regime is not instantaneous, its downfall is brought by sometimes lengthy judicial appeals and lobbying. In case of regulatory failure to maintain a given regime, the regulator logically cannot replace the model with a new high-powered model without recreating the necessary credibility and trust from industry and other stakeholders. By default, a low-powered regulation regime cannot be appealed based on methodological arguments, since it merely reimburses the observed costs[1]. In practice as well as in theory, this amounts to a period with a low-powered model (e.g. cost-plus or rate-of-return regulation) following any judicial backlash.

3.1. Theoretical model

A firm[2] facing a non-credible high-powered regime considers the probability of it falling to be $v \geq 0$ per year. As discussed above, we further assume that a failed regime is replaced by a low-powered regime in which the revenue of the firm is set to the actual cost. Let us assume ex-post verifiable cost in a sector with $i = 1,..,n$ operators each producing a vector of M outputs expressed as $y \in \mathbb{R}^M_0$ valued at output prices $p \in \mathbb{R}^M_{++}$ transforming a vector of N inputs $x \in \mathbb{R}^N_+$ with associated input prices $w \in \mathbb{R}^N_{++}$. The input prices include all expenses necessary to maintain the activity of the firm and its asset base. Furthermore, wx is the observed cost and $R = py$ denotes the revenue for the firm[3]. The production possibilities for the firm is defined as a set T of all feasible input-output combinations $T = \{(x,y)|x \text{ can produce } y\}$. Define $c(y,w)$ as the minimum cost to produce y outputs at input prices w that is $c(y,w) = \min_x \{wx : (x,y) \in T\}$. Let the associated minimum input vector for y be $x^\star$ such that $wx^\star = c(y,w)$.

The firm is maximizing expected horizon utility consisting of profits and *slack*, defined as the difference between observed and minimum cost. The one-period utility for inputs x at prices w and a regulated revenue R is

$$u(x,R) = (R - wx) + \rho(wx - c(y,w)), \tag{1}$$

where $\rho \in [0,1]$. The *slack* is here a convenient expression for the case of linear cost of *effort*, such that 1 euro of surplus cost (for example poor procurement policies, obsolescence, luxurious equipment) has the value of ρ euro in disposable profits (that could be paid as e.g. dividends). For a cost-plus policy with $R = wx$, equation (1) simplifies to

$$u(x,wx) = \rho(wx - c(y,w)). \tag{2}$$

The expected horizon utility is the sum of the discounted single-period utilities with a discounting factor $0 < \delta < 1$, meaning that 1 euro at the end of of the first period is worth δ euro at the start of the horizon. The reservation utility for the firm is normalized

[1]Note the possible appeals of the implementation of a low-powered model, i.e. the individual rulings and the due diligence exercised by the regulator in the cost-review process to define prudent cost behavior. However, such appeals refer only to the firm under scrutiny and their consequences do not jeopardize the viability of the regime as such.

[2]Firm is here used as the term for the concession holder or licensee, frequently called the *operator* of the regulated activity, excluding any concern for possible financial structures (holding companies) or (semi)-public enterprises or cooperatives.

[3]To simplify notation, all vector products are assumed to be done in the appropriate dimensions, i.e. suppressing the sign for the transposed vector.

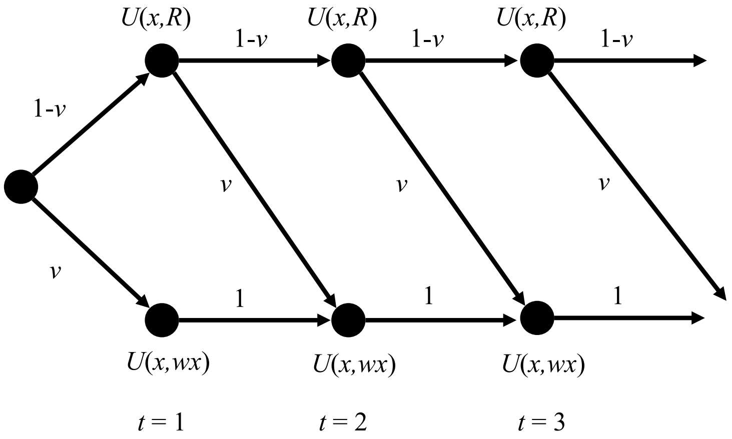

Figure 1: Dynamic regulation model with failure probability v.

to zero. The firm does not participate unless the expected utility is higher or equal to the reservation utility.

3.2. Dynamic game

The regulator and the firm are playing an infinite horizon game for a stationary demand y and fixed prices w. Initially, the regulator selects a regime $\{R, v\}$ consisting in a revenue level R derived using some methodology (e.g. average cost or ideal network models) associated with a failure risk v. The objectives of the regulator are to assure firm participation and to incite cost minimization by the participating firms. Observing the regime $\{R, v\}$ the firm chooses to participate in the game or to exit. Exiting gives reservation utility. If participating, the firm maximizes its expected horizon utility $EU(x)$ in (3) over an input vector $x \in [x^{\star}, \bar{x}]$ valid for the horizon[4]. The lower input bound $x^{\star}$ defines a fully *cost efficient* policy. Any policy such that $x > x^{\star}$ is defined as *cost inefficient* behavior. The payouts for a participating firm in the game are illustrated in Figure 1. In each period, the regime is repealed with probability v and the regulation resorts to cost-plus, i.e., the firm receives the reimbursement wx leading to a utility $u(x, wx)$. With probability $1 - v$, the regulator wins the appeal and enforces the payment R with the firm level utility $u(x, R)$. The failure of the regime is irreversible for the horizon, i.e., the regulator must continue paying wx for the rest of the horizon.

In this game, the expected horizon utility for a stationary input vector x for the firm

[4]The upper bound $\bar{x}$ for the support may be seen as existing observations or as the maximum value that the regulator could award without resetting the regulation.

is given as

$$\begin{aligned}
EU(x) &= \sum_{t=1}^{\infty} u(x,wx)v\delta^t + \sum_{t=2}^{\infty} u(x,wx)v\delta^t(1-v)^{t-1} + \sum_{t=1}^{\infty} u(x,R)\delta^t(1-v)^t \\
&= u(x,wx)\left[\frac{v\delta}{1-\delta} + \frac{v\delta^2(1-v)}{1-\delta(1-v)}\right] + u(x,R)\frac{\delta(1-v)}{1-\delta(1-v)} \\
&= \rho(wx-c(y,w))\left[\frac{v\delta}{1-\delta} + \frac{v\delta^2(1-v)}{1-\delta(1-v)}\right] + (R-wx+\rho(wx-c(y,w)))\frac{\delta(1-v)}{1-\delta(1-v)}.
\end{aligned} \tag{3}$$

The first-order condition of (3) with respect to x is obtained as:

$$\frac{dEU(x)}{dx} = \rho w\left[\frac{v\delta}{1-\delta} + \frac{v\delta^2(1-v)}{1-\delta(1-v)}\right] + w(\rho-1)\frac{\delta(1-v)}{1-\delta(1-v)}. \tag{4}$$

In the case of a perfectly *credible* regulation regime ($v=0$), the expected utility collapses to the expression

$$EU(x)_{v=0} = (R-wx+\rho(wx-c(y,w)))\frac{\delta}{1-\delta}. \tag{5}$$

Call a regulation regime such that $R \geq c(y,w)$ *feasible*, meaning that a cost efficient firm could participate in the game under a credible model ($v=0$). For such ideal regulation, the firm response would be full cost efficiency. On the other hand, consider a flawed regulation with certain demise ($v=1$). In this case, the expected utility becomes

$$EU(x)_{v=1} = \rho(wx-c(y,w))\frac{\delta}{1-\delta}. \tag{6}$$

Thus, the firm will follow an optimal policy leading to cost padding, selecting the upper bound $\bar{x}$, leading to the suboptimal cost $w\bar{x} > c(y,w)$.

3.3. Model results

We now proceed by deriving the results from the model in terms of firm behavior under various regulations, expressed here through the contextual parameters. The fundamental result postulates that the optimal firm response depends on intrinsic characteristics (time preferences and cost of effort), but also - which is a new result - on the credibility of the regulatory model, an exogenous parameter. We continue from the general case into a more representative situation where at least some firms have non-zero cost of effort. Based on this assumption, we derive the important Corollary 1 stipulating that in this case there will be a firm-specific threshold in credibility for cost-efficient firms. The next Corollaries 2 and 3 show the analogous thresholds for the types of firms in terms of time preferences and cost of effort. The last Corollary 4 gives the functional response for the firm behavior (efficiency) with respect to a given regulation. The theoretical results below will be used to formulate four empirically verifiable hypotheses that will be tested on empirical data in Section 7.

Proposition 1. *The optimal cost policy of a firm under multi-period regulation depends of the probability of regulatory failure v (*credibility*), the time preferences of the firm δ (*impatience*) and the utility value of inefficient cost ρ (*cost of effort*).*

Proof. Follows directly from the first-order conditions (4) and the three parameters above. ∎

Note that the level of the allowed revenue R does not affect the optimal cost policy, but only the participation constraint.

Corollary 1. *Assume a given cost of effort $\rho > 0$ and discounting factor δ. Then, there exists a finite failure risk $\hat{v}(\delta,\rho)$ above which cost-efficiency is a dominated policy.*

Proof. Follows from the first order condition (4) as a function $h(v,\delta,\rho)$. Define the critical failure rate $\hat{v}(\delta,\rho) = \{v : h(v,\delta,\rho) = 0\}$. For a failure rate $v \leq \hat{v}(\delta,\rho)$ cost-minimization is optimal and for $v > \hat{v}(\delta,\rho)$ the firm has a monotonously increasing utility in x. ∎

The function $\hat{v}(\delta,\rho)$ is illustrated in Figure 2. As a consequence of Corollary 1, the firm facing a credible regulation will select the input level $x^\star$, giving a cost efficient level $wx^\star = c(y,w)$. For a non-credible regime with a higher failure risk $v > \hat{v}(\delta,\rho)$, the firm will adopt an input maximization behavior, that is selecting the upper bound $\bar{x}$. The associated cost $w\bar{x} > c(y,w)$ implies cost inefficiency by the firm. As an example, imagine a firm with a low cost of effort ($\rho = 0.2$) and a discount factor $\delta = 0.99\%$. Then in Figure 2 the firm would be cost minimizing for any regime with a failure rate less than $\hat{v}(\delta = 0.99, \rho = 0.2) = 0.173$. On the other hand, a firm with a high cost of effort (e.g. $\rho = 0.8$) would only abstain from inefficiency if the policy were almost impossible to overturn, $v < \hat{v}(\delta = 0.99, \rho = 0.8) = 0.004$.

Corollary 2. *Assume a non-credible regime $v > 0$ and a given cost of effort ρ. Then, for any cost-efficient firm there exists an upper bound $\hat{\delta}$ for the discount factor.*

Proof. The bound is obtained from the fixed point $\hat{v}(\hat{\delta},\rho) = v$. ∎

Corollary 3. *Assume a non-credible regime $v > 0$ and a given discount factor δ. Then, for any cost-efficient firm there exists an upper bound $\hat{\rho}$ for the cost of effort.*

Proof. Follows directly from the inverse function of the critical failure rate $\hat{v}(\delta,\rho)$, the bound is obtained as $\hat{v}(\delta,\hat{\rho}) = v$. Note that $\hat{\rho}$ is unique and bounded for all $v > 0$. ∎

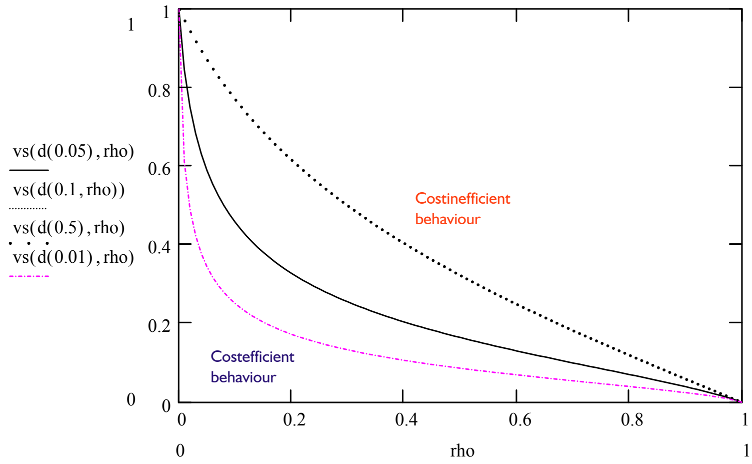

Figure 2: Critical failure probability $\hat{v}(\delta,\rho)$ for $\delta = \{0.99, 0.952, 0.909, 0.667\}$.

Corollary 4. *Assume a non-credible regime $v > 0$. The cost efficiency for a firm is then inversely proportional to the discount factor δ and the cost of effort ρ.*

Proof. Follows directly from Proposition 1 and the function for the first order conditions (4), $h(v,\delta,\rho)$. A necessary condition for a cost-efficient optimal policy by the firm is that $h(v,\delta,\rho) \leq 0$, inducing cost-minimization with the optimal solution $x = x^{\star}$. Corollary 3 gives that there is an upper bound $\hat{\rho}$ for h given v and δ. Corollary 2 provides analogously an upper bound $\hat{\delta}$ for h given v and ρ. Consider the curvature of h with respect to ρ:

$$\frac{d^2EU(x)}{dxd\rho} = w\left[\frac{v\delta}{1-\delta} + \frac{v\delta^2(1-v)+\delta(1-v}{1-\delta(1-v)}\right] > 0 \tag{7}$$

All terms are positive, meaning that $EU(x)$ is convex in ρ. Given that

$$lim_{\delta\to 0}\{h(v,\delta,\rho)\} = 0, \tag{8}$$

and

$$lim_{\delta\to}\{h(v,\delta,\rho)\} = \infty, \tag{9}$$

for given v and $\rho > 0$, the optimal policy will always be cost maximization for a sufficiently large δ. ∎

The implication of Corollary 4 is illustrated in Figure 3. The operator characteristics $\{\delta,\rho\}$ form a unit surface partitioned by an indifference curve $\hat{\delta}(\hat{\rho})$ derived for a given v from Corollary 4. The surface below the curve $\hat{\delta}(\hat{\rho})$ corresponds to the characteristics for cost-minimizing (cost efficient) firms, the area above corresponds to inefficient firms.

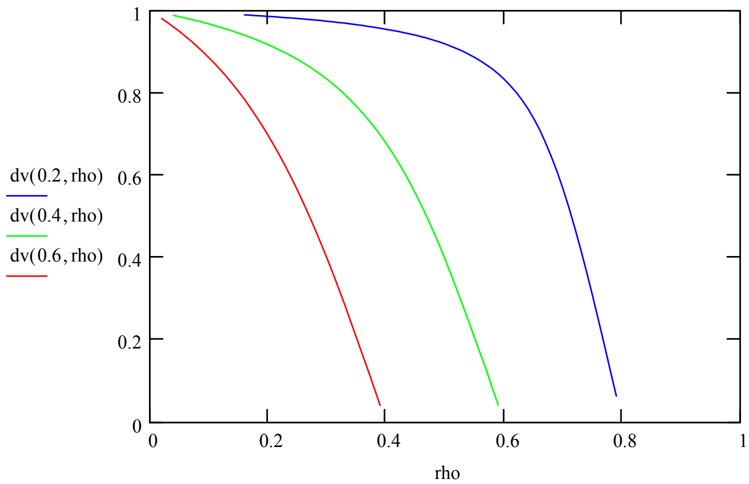

Figure 3: The indifference curve $\hat{\delta}(\hat{\rho})$ for $v = 0.2$ and $\rho = \{0.1, 0.2, ..., 0.9\}$

As the credibility of the regulation increases, v decreases and the set (area) in Figure 3 increases.

Remark 1. *Given n independent firms each having a cost of effort drawn from a distribution with density function $f(\rho)$ and cumulative density function $F(\rho)$ on the support $[0,1]$, then the probability that all firms are cost efficient under a non-credible regime is equal to $1-(F(\hat{\rho}))^n$.*

The intuition behind Remark 1 is clear: the hope of incentivizing all firms to efficiency in a weak regulation regime is thin. In practice, there will always be inefficient firms in the set of regulated operators, see Figure 4.

Thus, the empirical conjecture would then be a higher incidence of non-profit-maximizing behavior from firms that have a plausible case of a failing regulatory regime. In particular, firms with stable semi-public ownership can represent the case of long-range time preferences and high cost of effort. This is frequently the case for energy distribution in Europe. The opposite extreme would be a set of privately owned franchisees in countries with high inflation or political risks, here the time preferences is short-run and the relative cost of effort low. In the next section, we test the validity of our model on an interesting case of regulatory failure in Scandinavia, a region otherwise characterized by early adoption of market-oriented solutions, cf. Amundsen and Bergman (2007).

4. Empirical model

The previously presented model leads to a number of empirically verifiable hypotheses for non-credible regimes; in particular that firms exhibit lower cost efficiency fol-

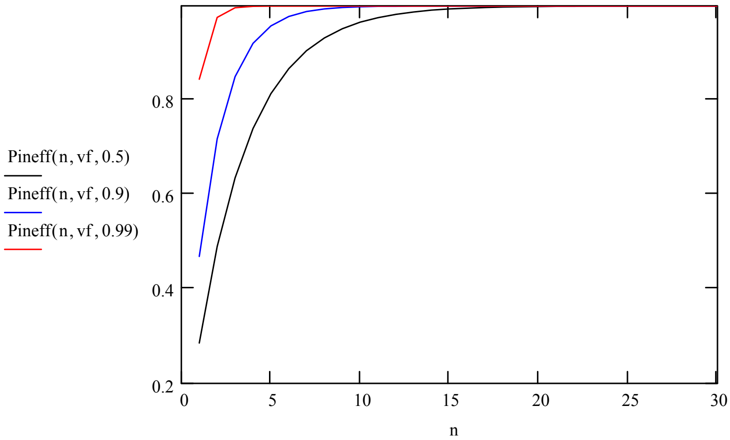

Figure 4: The probability of at least one inefficient observation for $n = 1,..,30$, $v = 0.2$ and $\delta = \{0.5, 0.9, 0.99\}$ when ρ is uniformly distributed.

lowing a flawed reform, that there is a stagnation of technological change and, as a consequence, that the productivity of the sector will suffer. In order to investigate this phenomenon, we need both a test-case with a failed regime and a methodology that is capable of differentiating between profitability, cost efficiency and dynamic effects for multi-input, multi-output production.

For the empirical case, we collected data for the period before and up to the failing of a transient regulation regime in Sweden, described in more detail in Section 5. The data and the activity model are presented in Section 6.

The methodology for the study of this empirical case is based on profitability[5], the firm financial indicator defined by the ratio revenue to cost $\Pi = py/wx$. Georgescu-Roegen (1951) introduced profitability, called return to the dollar, as a financial performance indicator into the economic literature. It is independent of the scale of production, a virtue not shared by cost, revenue or profit measures. This property of independence of the scale of production is particularly relevant in sectors with a wide range in the size of operation. Moreover, it allows for the direct comparison between the remuneration from the regulator ($R = py$) and the observed cost of the firm (wx). We are interested in the study of the evolution throughout time of the ratio revenue to cost, i.e. profitability change. It is defined as

$$\frac{\Pi^{t+1}}{\Pi^{t}} = \frac{p^{t+1}y^{t+1}/p^{t}y^{t}}{w^{t+1}x^{t+1}/w^{t}x^{t}}, \tag{10}$$

[5]See Grifell-Tatjé and Lovell (2015b)[Chapters 2-3] for an exhaustive exposition of this firm financial indicator.

which is equal to the ratio of revenue change to cost change. The next step is to identify the factors that cause changes in profitability. These factors are associated with changes in quantities and prices of individual outputs and inputs. Hence, we want to isolate the changes in prices of the changes in quantities, either of which influences profitability change. We can decompose cost change in (10) as

$$\begin{aligned}\frac{w^{t+1}x^{t+1}}{w^t x^t} &= \left[\frac{w^{t+1}x^{t+1}}{w^t x^{t+1}}\frac{w^{t+1}x^t}{w^t x^t}\right]^{1/2}\left[\frac{w^{t+1}x^{t+1}}{w^{t+1}x^t}\frac{w^t x^{t+1}}{w^t x^t}\right]^{1/2} \\ &= W_F(w^{t+1},w^t,x^{t+1},x^t)X_F(x^{t+1},x^t,w^{t+1},w^t)\end{aligned} \tag{11}$$

Expression (11) takes the geometric mean of the Laspeyres-Paasche index number pairs and creates a Fisher input price index and a Fisher input quantity index[6]. The last line of (11) defines a more compact notation as a Fisher input price index W_F and a corresponding Fisher input quantity index X_F.

Symmetric derivations for the revenue side obtain an analogous expression for revenue change as

$$\frac{p^{t+1}y^{t+1}}{p^t y^t} = P_F(p^{t+1},p^t,y^{t+1},y^t)Y_F(y^{t+1},y^t,p^{t+1},p^t), \tag{12}$$

where P_F is called a Fisher output price index and Y_F defines a Fisher output quantity index. Combining (11) and (12) yields an expression for the relative change in profitability,

$$\frac{\Pi^{t+1}}{\Pi^t} = \frac{P_F(p^{t+1},p^t,y^{t+1},y^t)}{W_F(w^{t+1},w^t,x^{t+1},x^t)}\frac{Y_F(y^{t+1},y^t,p^{t+1},p^t)}{X_F(x^{t+1},x^t,w^{t+1},w^t)}. \tag{13}$$

This profitability change is the product of a Fisher *price recovery* index P_F/W_F for output and input prices, respectively, and the Fisher *productivity* index Y_F/X_F for output and input quantities, respectively. The price recovery index compares the variation in the prices on the inputs with the corresponding input price changes between two periods. The interpretation of price recovery index is intuitive: a value lower than unity implies that the firm has increased the output prices less than the input price changes. Under high-powered regulation, where the output prices are fixed, this would be the expected outcome. Inversely, a price recovery index higher one than indicates an increase of the output prices compared to the input prices.

We are interested in the decomposition of the Fisher productivity index by its economic drivers. It has been object of attention by Diewert (2014) (see also Grifell-Tatjé

[6]The Fisher index has a set of appealing axiomatic properties, see Balk (2012)

and Lovell (2015a)), Grifell-Tatjé and Lovell (2003, 2015b); Kuosmanen and Sipiläinen (2009); Ray and Mukherjee (1996). We follow Grifell-Tatjé and Lovell (2015b) that can be related to the approach of Ray and Mukherjee (1996). A conventional approach to deal with the generic production possibilities set introduced in Section 3.1 is defined by the mathematical programming models in Färe et al. (1985), based on the Data Envelopment Analysis (DEA) technique introduced by Charnes et al. (1978). Its empirical form in period t is

$$T^t = \left\{ (x,y) : y \leq \sum_i \sum_{s \leq t} \lambda_i^s x_i^s, x \geq \sum_i \sum_{s \leq t} \lambda_i^s y_i^s, \lambda_i^s \geq 0 \right\}, \tag{14}$$

where x_i^s is the input vector and y_i^s is the output vector for firm i in period s, respectively. Note that the technology here is defined for constant returns to scale, which is a common assumption in energy network regulation (Jamasb and Pollitt, 2003). Additionally, the technology is bounded (above) by a piecewise linear surface, or frontier, over the data formed by the best observations in all the years from year 1 to year t inclusive. We observe that the technology is defined in a *sequential* way which does not allow for technical regress (Tulkens and Eeckaut, 1995). As before $c^t(y^t, w^t) = \min_x \{w^t x^t : (x^t, y^t) \in T^t\}$ where $w^t x^t \geq c^t(y^t, w^t)$ and cost efficiency is defined as $CE^t(y^t, w^t) = c^t(y^t, w^t)/w^t x^t \leq 1$[7]. The decomposition of the Fisher productivity index Y_F/X_F in (13) is based on the input quantity index $X_F(x^{t+1}, x^t, w^{t+1}, w^t)$ because the output quantities are considered exogenous. We focus initially at the Laspeyres input quantity index component of $X_F(x^{t+1}, x^t, w^{t+1}, w^t)$. We have

$$\frac{w^t x^{t+1}}{w^t x^t} = \frac{w^{t+1} x^{t+1}/c^{t+1}(y^{t+1}, w^{t+1})}{w^t x^t/c^t(y^t, w^t)} \frac{c^{t+1}(y^t, w^t)}{c^t(y^t, w^t)} \frac{w^t x^{t+1}/c^{t+1}(y^t, w^t)}{w^{t+1} x^{t+1}/c^{t+1}(y^{t+1}, w^{t+1})}, \tag{15}$$

and we can introduce this decomposition (15) into the Laspeyres productivity index of the Fisher productivity index in (13). This yields

$$\begin{aligned} \frac{\frac{p^t y^{t+1}}{p^t y^t}}{\frac{w^t x^{t+1}}{w^t x^t}} &= \frac{w^t x^t/c^t(y^t, w^t)}{w^{t+1} x^{t+1}/c^{t+1}(y^{t+1}, w^{t+1})} \frac{c^t(y^t, w^t)}{c^{t+1}(y^t, w^t)} \\ &\quad \left[\frac{\frac{p^t y^{t+1}}{p^t y^t}}{\frac{c^{t+1}(y^{t+1}, w^t)}{c^{t+1}(y^t, w^t)}} \frac{w^{t+1} x^{t+1}/c^{t+1}(y^{t+1}, w^{t+1})}{w^t x^{t+1}/c^{t+1}(y^{t+1}, w^t)} \right] \end{aligned} \tag{16}$$

[7]The calculations are made in R (R Core Team, 2015) partly using the ”Benchmarking” package (Bogetoft and Otto, 2015).

which attributes Laspeyres productivity change to three economic drivers. The first term at the right hand side of (15) is the *cost efficiency change*. It can take values higher, equal or lower than one. If the firm is more, equal or less cost efficient in period $t+1$ than in period t, the implication for productivity is increasing, maintaining or decreasing the current level. The second expression on the right hand side of (15) measures the impact of *technical change* on the productivity of the firm. It can take values higher or equal to one, since the technology here is defined as sequential, without regress. A value higher than one means positive technical change. Then, the cost of producing y^t at prices w^t, is lower with the technology of period $t+1$ than using that of period t. Finally, the third expression in (16) in brackets measures the impact on productivity of two effects; *change of size* and *allocative efficiency*. For simplicity we call the entire term *size change*[8].

A similar decomposition to (16) can be done based on the Paasche productivity component of the Fisher productivity index in (13). Here we obtain the same components, i.e. the cost efficiency change, the technical change and the size change[9]. Thus, taking the geometric mean of the decomposition of the Laspeyres productivity index and the Paasche productivity index produces a decomposition of the Fisher productivity index. We obtain

$$\begin{aligned}\frac{Y_F(y^{t+1},y^t,p^{t+1},p^t)}{X_F(x^{t+1},x^t,w^{t+1},w^t)} &= \frac{w^t x^t/c^t(y^t,w^t)}{w^{t+1}x^{t+1}/c^{t+1}(y^{t+1},w^{t+1})} \\ &\quad \left[\frac{c^t(y^t,w^t)}{c^{t+1}(y^t,w^t)}\,\frac{c^t(y^{t+1},w^{t+1})}{c^{t+1}(y^{t+1},w^{t+1})}\right]^{1/2} \\ &\quad \left[\frac{\frac{p^t y^{t+1}}{p^t y^t}}{\frac{w^t x^{t+1}/c^{t+1}(y^t,w^t)}{w^{t+1}x^{t+1}/c^{t+1}(y^{t+1},w^{t+1})}}\;\frac{\frac{p^{t+1}y^{t+1}}{p^{t+1}y^t}}{\frac{w^t x^t/c^t(y^t,w^t)}{w^{t+1}x^t/c^t(y^{t+1},w^{t+1})}}\right]^{1/2} \\ &= \Delta CE\cdot\Delta TC\cdot\Delta SC\end{aligned} \tag{17}$$

The first term, the cost efficiency change ΔCE indicates whether the cost efficiency

[8]Without output mix changes and under constant returns to scale $y^{t+1}=\lambda y^t$, $\lambda>0$ and $c^{t+1}(y^{t+1},w^t)/c^{t+1}(y^t,w^t)=\lambda$, making the first term in the brackets of (16) equal to 1. Consequently, the first term in brackets of (16) only measures the impact on productivity of output mix changes. Grifell-Tatjé and Lovell (2015b) show that the second component in brackets of (16) is an input allocative effect that can be expressed as $AE^{t+1}(x^{t+1},w^{t+1},y^{t+1})/AE^{t+1}(x^{t+1},w^t,y^{t+1})$. It is bounded above by unity if x^{t+1} is more allocatively efficient relative to (y^{t+1},w^{t+1}) than to (y^{t+1},w^t) under the period $t+1$ technology, which corresponds to the expected behavior by the firm.

[9]This footnote is similar to the previous one. The change of size effect is also equal to 1. Additionally, we have a component that can be interpreted as an input allocative efficiency effect but, in this case, it is bounded below by unity (Grifell-Tatjé and Lovell, 2015b, p.145).

has changed between periods t and $t+1$. The second term, the technological change ΔTC gives information about the shift of the cost function, i.e. the frontier, between periods t and $t+1$. As before, since the frontier can only shift outwards, ΔTC is bounded below by unity. Under the assumptions of this study the third term, the size change ΔSC, takes a unit value[10]. Hence, the evaluation of the drivers behind productivity growth can be limited to cost efficiency change ΔCE and technological change ΔTC. In the following, we will use these terms to assess whether there has been a behavioral change in productivity in the sector following the change of regulation regime.

5. Case: the engineering cost model in Sweden 2003-2006

Prior to the deregulation in 1996, the Swedish utilities were predominately integrated and the market dominated by the publicly owned Vattenfall network (market share 35% of total delivered energy). Although a regulatory commission was appointed, the incumbent regime was basically self-regulation by industry associations and larger firms as price-setters. The unbundling following the deregulation and the application of the Electricity Act (1996) changed the power balance in the new market for electricity distribution between the regulatory authority *Swedish Energy Agency* (STEM)[11] and the firms.

Supported by a revised Electricity Act (2001), the regulator launched an original initiative in 1998 to develop an engineering cost model, called the Network Performance Assessment Model (NPAM), to be used to review tariffs for electricity distribution. NPAM is used to calculate a surrogate output measure independent of historical costs. It thus gives an assessment of absolute performance according to a technical norm.

The NPAM model corresponded to a significant investment on behalf of the regulator[12]. The participating firms fully carried the costs of GIS-positioning[13], information

[10]As we have seen in footnotes above under the assumption of constant returns to scale, the size change captures changes in allocative efficiency and output mix. Concerning allocative efficiency AE, we expect a value close to one for the geometric mean of two terms, each bounded below and above by one. In the case of infrastructure provision, the output mix changes per period are likely to be small and consequently the contribution of the output mix will be equal or close to one. As a conclusion, the product of the two effects must be equal to one, or close to it. In the empirical part of the study in Section 7, this is the case throughout the period.

[11]STEM is the acronym for *STatens EnergiMyndighet*, translated as the *Swedish Energy Agency*. In 2008, the regulatory department was made autonomous under the name the *Swedish Energy Market Inspectorate*, EI.

[12]A cautious estimate 1998-2002 for the direct cost for the regulator could be 84 man-months of internal resources and some 60 man-months of external resources, excluding costs for software, coding, administration of two pilot runs with live data, information material and dissemination.

[13]The model requires the geographical location and type of all low-, medium- and high-voltage connections and the transformer stations from the transmission system operator.

processing, validation and reporting. Internal estimates by the regulator indicate that these fixed administrative costs can be relatively heavy in the budget of smaller networks with non-GIS administered connections. The annual reporting cost for the operators was estimated by the Swedish Government to 230 MSEK (27 M€) (Government of Sweden, 2009).

5.1. Critique

The NPAM model is a conscientious attempt to model a complex planning problem, without duplicating the industry planning systems. An obvious criticism is the limited information value of the model compared to its costs. Neither the generated grid, nor the budget elements have any prescriptive value other than in the final cost expression. In spite of its official status, a quarterly news brief, several workshops held by STEM and numerous courses by various associations, the model was poorly documented (Larsson, 2003, 2005; Lantz, 2003; Gammelgård, 2004).

Industry associations (Svensk Energi, 2002) criticized the model for its lack of investment incentives for delivery quality. The quality provisions in the model were oriented towards observed interruptions, whereas the industry associations suggested more refined measures for the stochastic outcomes. Even the operators ranked as fully efficient by the NPAM gave the regime a cold shoulder (Törnqvist and Persson, 2006).

Early academic critique Lantz (2003) questioned the incentive properties of the NPAM model as a yardstick instrument, given that information collected from a firm will contribute to the setting of its individual score. The model also ignores the time perspective involved, the distinction between short-run operating efficiency with given grids and long-run technical efficiency, based on ex ante information. The scant scientific analyses of the method highlighted the unclear feasibility of the cost norm (Jamasb and Pollitt, 2008; Jamasb and Söderberg, 2010), as well as its lack of robustness (Wallnerstrom and Bertling, 2008). However, Turvey (2006) recognized the weaknesses while still finding the appraoch in NPAM superior to econometric models. Notwithstanding, the regulatory strategy for the exact use and the calibration of the model were never adequately defined and anchored, not even within the regulatory agency.

5.2. Demise and fall

Formally implemented Jan 1, 2003, NPAM was not operational until June 2004 due to problems in parametrization, testing and documentation. The application of NPAM turned out to be very complex. The assessment of the operating year 2003, disclosed in five rulings from June 2005 to June 2006[14] obliged in total 21 operators to lower their tariffs with in total 644 MSEK (76.3 M€). All concerned concession operators filed appeals of the rulings to the administrative court. The firms attacked a number of

[14] June 21, 2005; Dec 8, 2005; Jan 26, 2006; Feb 10, 2006 and June 1, 2006

procedural issues in addition to the principal question of the feasibility of the cost norm. After two rounds[15] in the administrative court, the regulator backed down on the high-powered principle and defined an upper bound for revenue cuts through a calculation of a reasonable return of capital. However, rather than solving the feasibility of the NPAM, the focus was now enlarged to the definition of the cost of capital in the new regime. Gradually in the judicial process, the requirements for revenue cuts was reduced to 198 MSEK (23 M€) in 2007, dismissing the appeals from all but eight operators. In spite of the concessions, it becomes clear to the government and part of the regulatory authority that NPAM would not withstand the judicial process. Replacing the previous regulator in January 2008, the newly appointed director scrapped the earlier regulatory policy and settled the regulatory claims out of court in bilateral negotiations with the eight remaining firms, further reducing their revenue cuts to 140 MSEK (16.5 M€). From Jan 2009, it was formally decided that NPAM is no longer a regulatory regime and an interim cost-recovery regime was established until 2012[16].

6. Data

We analyse a balanced panel of annual data collected by the regulator STEM for 132 to 137 DSOs operating 241 to 247 concession areas for 2000 to 2006. Observations with erroneous or missing data for inputs or outputs were deleted from the analysis set, corresponding to 1.5% of the revenue base. The mergers of distribution networks that occur during the period are partly actual horizontal integration, partly coordinated reporting of multiple concession areas under the same ownership and control. Since the purpose of the paper is not to study the differences in efficiency between merged, coordinated and non-merged entities, the entities constituting the former two categories are consolidated prior to the merger event to create a balanced panel. Thus, we recreate retroactively the merged entity prior to the merger by consolidating data in the panel. The final consolidated data set contains 128 DSOs active in all years 2000-2006, in all 896 observations. Some descriptive statistics for the data set are given in Table 1.

6.1. Activity model

The activity analysis model for the sector contains two outputs and four inputs. The outputs are the main revenue drivers; distributed electricity at low voltage and high voltage, respectively. The output prices are derived as the ratio of total revenue per low voltage and high voltage divided by the volumes distributed per voltage level, respectively. The total costs are decomposed into four cost types with respective prices;

[15] October 2007 and May 30, 2008

[16] From January 2012, a new regime is a low-powered rate-of-return regime, based on full cost-recovery for all firms, later complemented with a 1% productivity target for 25% of the controllable operating expenditure (Ek, 2010).

transmission costs, costs for energy losses, operating expenditure (OPEX, excluding energy) and capital expenditure (CAPEX). The corresponding four inputs are total energy volume for transmission, total energy losses in the distribution network, total network length weighted with the number of low-voltage connections (as a driver of OPEX) and the nominal network capital. Average input prices are obtained through the total costs divided by the input quanta. By definition, the activity model explains all revenues and costs for the operations.

Table 1: Descriptive statistics and model variables.

Category	Unit	Definition	mean	median	SD
Revenue $R = py$	kSEK	Total revenue	137,764	49,967	387,118
	kSEK	Revenue LV	118,394	41,876	335,470
	kSEK	Revenue HV	19,371	6,707	53,213
Costs wx	kSEK	Total cost (TOTEX)	119,515	46,483	346,036
	kSEK	Cost transmission	33,791	13,285	100,420
	kSEK	Cost energy losses	7,878	2,864	21,395
	kSEK	Operating expenditure (OPEX)	46,766	18,615	130,483
	kSEK	Capital expenditure (CAPEX)	31,082	8,602	102,922
Outputs y					
	MWh	Energy delivered low voltage (LV)	488,052	204,662	1,235,396
	MWh	Energy delivered high voltage (HV)	221,633	71,037	623,509
Output prices p					
	kSEK/MWh	Price per energy delivered LV	0.228	0.226	0.043
	kSEK/MWh	Price per energy delivered HV	0.109	0.104	0.057
Inputs x					
	MWh	Energy transported, total	742,112	281,796	1,913,920
	MWh	Energy losses, total	32,427	11,952	86,027
	km	Connection-weighted network LV+HV	41,415	14,198	121,128
	kSEK	Network capital, total	458,831	100,737	1,521,204
Input prices w					
	kSEK/MWh	Transmission price	0.049	0.048	0.019
	kSEK/MWh	Cost per energy losses	0.260	0.252	0.120
	kSEK/km	OPEX per connection-line meter	1.379	1.332	0.543
	%	Cost of capital	0.086	0.083	0.033

Notes: km = 1,000 meters, kSEK = 1,000 Swedish crowns (SEK), SD = standard deviation.

7. Analysis

Initially, we observe that the facts in the case correspond relatively well to those of the theoretical model. First, given the mixed ownership situation among the operators (50 per cent privately owned[17], 40 per cent publicly owned[18] and 10 per cent consumer-cooperatives, (Agrell et al., 2005b)), we can safely assume that there is some heterogeneity with respect to the cost of effort ρ in the sample. At least some firms should have a strictly positive ρ, meaning that they are not exclusively profit maximizing. Second, given the lukewarm reception of the new regulatory model and the industrial and academic critique raised against it, we safely assume that all firms observed a high probability of regime failure $v >> 0$. Without adventuring into what value each firm (or group of firms) attributed to v, it suffices to recall our results from Remark 1 where adverse effects appear already from modest levels of v.

For clarity, we restate our hypotheses and proceed into the analysis step by step.

First, from Corollary 1 we have that the optimal response for firms with non-zero cost of effort ($\rho > 0$) should be inefficiency for a sufficiently weak regime. We formulate this as a sector prediction, based on the reasoning above concerning the sample:

Hypothesis 1. *Firms exhibit a lower cost efficiency CE during a non-credible regime* $v > 0$.

The average cost efficiency CE^t for the operators by year is listed in Table 2 below. We note a clear tendency of decreasing efficiency during the period, from on average 74.5 per cent before the NPAM to 71.9 per cent during the NPAM. In fact, the overall fall in cost efficiency from the initial year (2000) to the last year (2006) is 5.4 per cent. The difference in mean cost efficiency (2.6 per cent) is statistically significant. Additional support for this finding is found in Figure 5, where the red curve shows the mean CE^t by operator before the NPAM and the black curve depicts the analogous cost efficiency after the NPAM introduction. As seen in the Figure, the fall in cost efficiency is generalized, except for some initially highly inefficient operators[19]. Hence, we conclude that Hypothesis 1 is not rejected by the data.

Second, we extend the analysis to the multi-period setting. Assuming that Hypothesis 1 is valid for any period, then the prediction would be that the firms regress collectively over time, leading to no technical change:

Hypothesis 2. *The technical change of the firms is stagnating for the duration of a non-credible regime* $v > 0$.

The technical change ΔTC and the variation in cost efficiency ΔCE from the method-

[17] Including international firms such as *EDF, Electricite de France*, owner of *Graninge*, and the inter-Nordic operator *Birka*.

[18] Mostly municipal utilities with the exception of the state-owned operator *Vattenfall*.

[19] An inefficient policy under a credible regulatory regime may be optimal at very high cost of effort ρ, in which case the policy is independent of the credibility of the regime. Our results are robust to this assumption.

Table 2: Profitability Π^t and cost efficiency CE^t, mean per year, 2000-2006.

	year							period		
	2000	2001	2002	2003	2004	2005	2006	2000-02	2003-06	Diff
Π^t	1.150	1.149	1.141	1.128	1.128	1.086	1.079	1.147	1.105	-0.042***
CE^t	0.762	0.732	0.741	0.732	0.723	0.713	0.708	0.745	0.719	-0.026***

Notes: $***p < 0.001$; $**p < 0.05$; $*p < 0.01$.

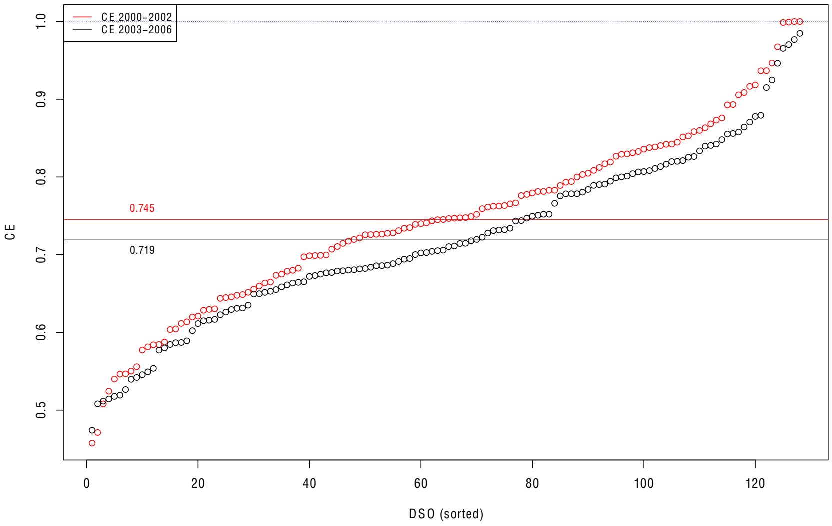

Figure 5: Cost efficiency CE^t, average per DSO, development before and after NPAM.

ological framework, expression (17), are calculated and reported in Table 3[20]. The pre-reform technological change rate is strong, on average 4.8 per cent per year. At the launch of the NPAM, there is a radical drop in technological change, the value is virtually at its floor of unity (0.1 per cent). The difference in mean technical change is significant and of the expected sign. Looking closer at the data in Figure 6 shows a striking difference at firm level. In the anticipation of the fall of the NPAM model, almost no operator shows technological progress (the black curve). As predicted by the model, the absence of incentives is leading to a stand-still of the investments in new technology and processes, resulting in the observed stagnation of technological change. Hypothesis 2 is not rejected by the data.

Third, continuing with the dynamic analysis we turn our attention to the productivity development. As a direct consequence of the previous two Hypotheses, the productivity change for the average firm (not only the frontier firm) would stagnate or decline for a

[20]The results for ΔSC from (17) are equal to unity throughout the period without any significant differences.

Table 3: Cost efficiency ΔCE and technology change ΔTC, before and after NPAM.

	All		Pre NPAM		Post NPAM			
n	768		384		384		384	
period	2000-2006		2000-2002		2003-2006			
	Mean	SD	Mean	SD	Mean	SD	Diff	p-value
ΔCE	0.990	0.065	0.989	0.082	0.991	0.043	0.002	0.778
ΔTC	1.024	0.033	1.048	0.033	1.001	0.009	-0.047***	< 0.001

Notes: ***$p < 0.001$; **$p < 0.05$; *$p < 0.01$.

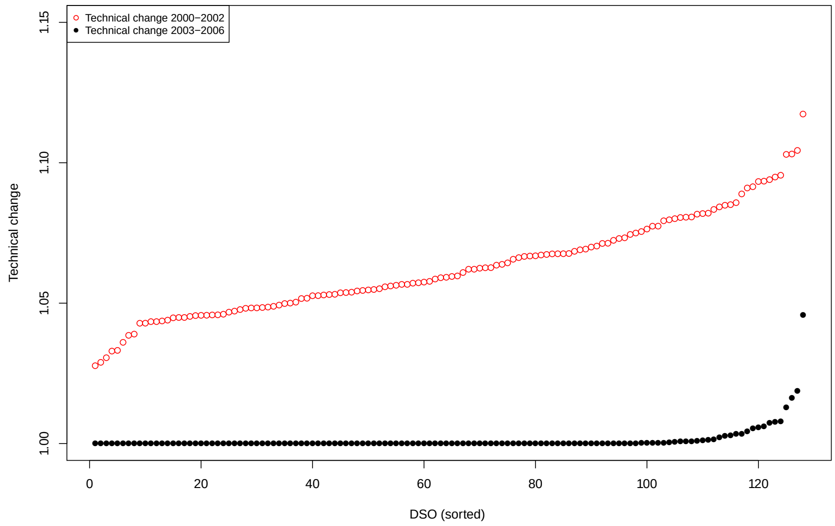

Figure 6: Technical change ΔTC, average per DSO, before and after NPAM.

sufficiently weak regime:

Hypothesis 3. *The productivity change of the firms is low or nil for the duration of a non-credible regime $v > 0$.*

The productivity change, calculated as the right-hand side of expression (13), is presented in Table 4. As a consequence of the previous results in terms of cost efficiency and technological change, the productivity change prior to the reform was positive and strong, on average 3.5 per cent per year. After the introduction, the productivity change is on average negative (-0.7 per cent). The difference between the means of the two periods is significant and of the right sign. Presented in more detail in Figure 7, the findings are without any ambiguity. Here we note that the confidence intervals of the

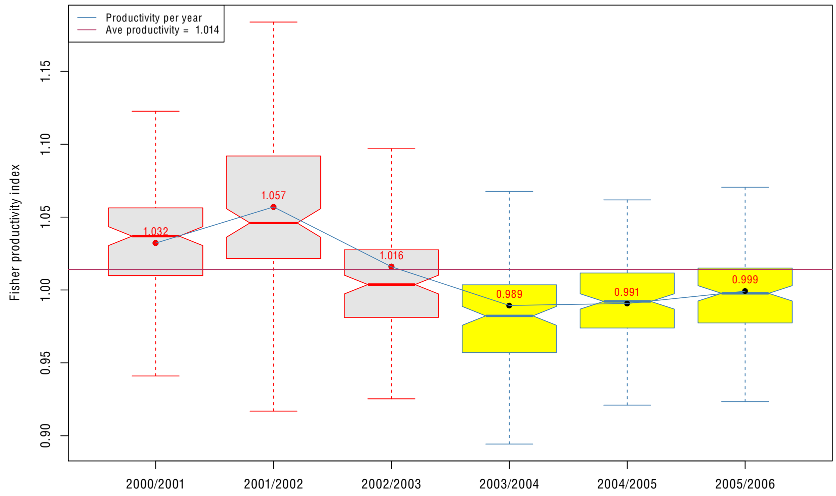

Figure 7: Productivity development 2000-2006.

three years during the reform are below the horizon average productivity change (1.4 per cent). Thus, we find that Hypothesis 3 cannot be rejected.

Table 4: Profitability variation, price recovery and productivity change, before and after NPAM.

	All		Pre NPAM		Post NPAM			
n	768		384		384		384	
period	2000-2006		2000-2002		2003-2006			
	Mean	SD	Mean	SD	Mean	SD	Diff	p-value
Profitability variation	0.994	0.097	0.997	0.080	0.991	0.111	-0.006	0.470
Price recovery	0.987	0.137	0.973	0.149	1.001	0.123	0.028**	0.005
Productivity change	1.014	0.084	1.035	0.102	0.993	0.053	-0.042***	< 0.001

Notes: $***p < 0.001$; $**p < 0.05$; $*p < 0.01$.

Fourth, we now address the paradoxical firm-level incentives under non-credible regulation. Conventional intuition would postulate that the firms would adopt an efficient policy when facing a high-powered regulation, allowing for non-negative profits. However, in our model the firm may very well incur losses for a certain time without resorting to efficient operations, even when given the option to do so. Thus, the participation constraint is not binding: it is not even active during the period. We formulate this as our last hypothesis:

Hypothesis 4. *The profitability of the firms is lower on average, and decreasing throughout the duration of a non-credible regime $v > 0$.*

The average profitability Π^t by firm and year is presented in Table 2 above. The findings document a significant and monotonous fall in profitability over the period,

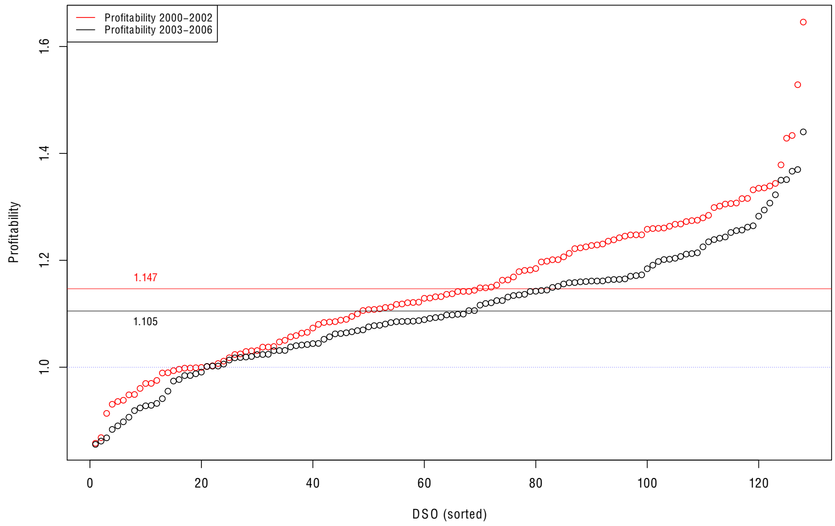

Figure 8: Profitability Π^t, average per DSO, before and after NPAM.

from on average 14.7 per cent prior to the NPAM to 10.5 per cent after the reform. At operator-level, the results are illustrated in Figure 8. Hypothesis 4 is not rejected by the empirical data.

The result can be explained by the previously presented methodology. Profitability variation is driven by productivity change and price recovery, as shown in expression (13). We have already discussed the results for the productivity change above, so our attention is turned to price recovery. The results for price recovery are presented in Table 4. An interesting pattern is revealed: the initial price recovery is negative (-2.7 per cent), meaning that the operators did not fully pass on input price changes into tariff changes. However, they maintained profitability through productivity improvements. This translates into sharing the productivity gains with the captive clients at a stable profitability level (see Table 2). After the reform, however, the average price recovery is about one, meaning that input price changes (e.g. inflation) are entirely passed on to the customers. On the other hand, as shown above, the productivity change is lower than one, implying a decrease in profitability. The scenario is indeed gloomy: the customers end up paying more and the firms' owners earning less on the operations.

Remains then the question of a counter-factual for the model results. The empirical results could be distorted if there was a general slowdown of productivity in the period, or if the period just before the reform was a "golden age" with abnormal productivity development. In Table 5 we list the results for some previous studies for electricity distribution in Scandinavia[21].

Table 5: Cumulative productivity development, electricity distribution, 1970-2004.

Paper	Country	n	Period	M	TC
Hjalmarsson and Veiderpass (1992)	Sweden	298	1970-78	1.56	1.42
Hjalmarsson and Veiderpass (1992)	Sweden	298	1978-86	1.22	1.39
Kumbhakar and Hjalmarsson (1998)	Sweden	108	1970-90	-	0.019 - 0.022/yr
Førsund and Kittelsen (1998)	Norway	150	1983-89	1.12	1.11
Edvardsen et al. (2006)	Norway	98	1996-03	1.15	-
Agrell et al. (2015)	Norway	198	1995-04	1.24	1.25
Kumbhakar et al. (2014)	Norway	127	1998-10	-	0.01/yr
Miguéis et al. (2011)	Norway	127	2004-07	1.00	1.04

Notes: M = Malmquist index, TC = Technical change, n = average no of obs per year.

Starting with the pre-reform situation: The seminal study by Hjalmarsson and Veiderpass (1992), documents an average productivity growth (TC) 1970-1986 for Swedish electricity distribution of 5 per cent annually. In Kumbhakar and Hjalmarsson (1998), a parametric approach resulted in stable technical change estimates around 2 per cent per year for Sweden 1970-90. The Swedish DSOs had clearly a stable productivity development prior to the reform.

Next, a reasonable counterfactual can be found in the Norwegian electricity DSOs of similar size and ownership distribution. The regulation in Norway was based on a frontier analysis model implemented as a revenue cap. The regulation was not disputed and must be characterized as robust and credible. A study by Førsund and Kittelsen (1998) show a technical change in the order of 1.8 per cent per year for the period 1983-89. Later work by Edvardsen et al. (2006) reports an average annual Malmquist productivity increase of 2.1 per cent for the period 1996-2003, confirmed by Agrell et al. (2015) for the period 1995-2004 with larger dataset as 2.8 per cent average annual technical change. Miguéis et al. (2011) used a non-parametric model and weight restrictions on the Norwegian DSOs 2004-2007, yielding an average technical change of 1.2 per cent per year. These results are also robust for parametric methods. Kumbhakar et al. (2014) used parametric distance functions to estimate scale efficiency effects for Norwegian DSOs 1998-2010, finding proof of an average technical change rate of 1 per cent per year.

Thus, we are facing a scenario where comparable operators continue to show positive productivity growth, whereas the regulatory crisis brings the development to a halt.

[21]The hypothesis that the Swedish operators are structurally different from those in Scandinavia is rejected in the activity analysis in Agrell and Bogetoft (2010)

8. Discussion

The inspiration for this paper comes from an apparent empirical contradiction to the logic of incentive theory: why would firms stall their productivity development, increase user prices and endure lower profitability when revenue is fixed beforehand anyway in incentive regulation? The model in this paper brings rationality to observed firm behavior when taking into account the fact, usually neglected in regulation economics, that the method for deciding the revenue or price caps must be economically sound and incentive compatible. If this is not the case, such as with the Network Performance Assessment Model (NPAM) in Sweden, the firms can no longer trust the long-term commitment from the regulator on the reimbursement, since the model is likely to be rejected. This implies that each operator must engage in a gamble. On the one hand, investing in efficiency improvements may be very costly if the model is retracted and replaced with a cost-plus regime, in which case no profit can be harvested from efficiency. On the other hand, staying at an inefficient cost level in spite of fixed tariffs will drive profitability down if the regime endures.

In testing the model predictions, we note that the parameters *prima facie* are non-observable. Moreover, in a general setting the failure risk v is not common knowledge but a firm-level estimate. However, the case of NPAM in Sweden is particularly well suited since the failure risk was higher than the thresholds derived in the theoretical model, such that all but the most inept[22] firms would find efficiency to be a dominated policy. Indeed, the wide participation in the judicial class actions by operators of different size and governance endorses the universality of the vulnerability assessment. Another advantage with the NPAM case is the full adoption of a cost-recovery regime following the collapse of the regime. Since by definition there are no incentives for efficiency in low-powered regulation, it is not necessary to validate their post-NPAM policies. One may also speculate whether the strong criticism in itself influenced the credibility, questioning the exogeneity of the parameter and opening for the question of collusion. We conjecture that the failure prospect v in this particular case was so convincing that no collusion was necessary; the regime was likely to fall by its own demerits rather than by socio-economic pressure.

The policy conclusions of the models can be derived at two levels. First, the choice of incentive regulation *per se* is only the initial step in a process where the method of determining the parameters and levels of the reimbursements is equally important. In the current model, we have chosen not to frame the model in a social welfare context as the perspective is limited to a given pool of firms under global cost minimization. However, a policy maker must also consider market entry and industrial restructuring as valid options to restore efficiency and to introduce heterogeneity in firm response. The trade-off between the costs carried by the regulator versus those covered by the firms is

[22] Or the most impatient, but the Swedish operators can be assumed to be homogeneous in this regard.

of principal, rather than applied, interest in this case.

Finally, a word about the generality of the contribution in this paper. Economic regulation is omnipresent, not only in energy networks and more generally in utilities (water, energy, transport), but also in many other public services of quasi-monopoly character (airports, public safety, hospitals, education, transport infrastructure). Contracts in these areas commonly include incentive components that are high-powered in order to promote efficiency and obtain lower costs. As the budgets available for such activities dwindle, it is likely that the regulators, or equivalent authorities, will seek to estimate the costs closer to the actual best-practice. If this is not judiciously done and the services tendered are essential and capital intensive, then behavior similar to that predicted by the model may occur among the service providers. Thus, we believe that this contribution provides a justified and important complement to the basic theory of incentive provision in mechanism design, for the benefit of regulators, firms and consumers.

Agrell, P., Bogetoft, P., 2010. Harmonizing the nordic regulation of electricity distribution. Energy, Natural Resources and Environmental Economics, 293–316.

Agrell, P., Bogetoft, P., Grammeltvedt, T. E., May 2015. The efficiency of the regulation for horizontal mergers among electricity distribution operators in norway. In: 12th International Conference on the European Energy Market (EEM). IEEE, pp. 1–5.

Agrell, P., Bogetoft, P., Tind, J., 2005a. Dea and dynamic yardstick competition in scandinavian electricity distribution. Journal of Productivity Analysis 23 (2), 173–201.

Agrell, P. J., Bogetoft, P., Bjørndalen, J., Vanhanen, J., Syrjänen, M., 2005b. Nemesys subproject a: system analysis. Tech. rep., Nordenergi and Sumicsid.

Agrell, P. J., Teusch, J., 2015. Making the belgian distribution system fit for the energy transition. Reflets et Perspectives de la Vie Économique 54 (1), 157–174.

Amundsen, E. S., Bergman, L., 2007. Integration of multiple national markets for electricity: The case of norway and sweden. Energy Policy 35 (6), 3383–3394.

Averch, H., Johnson, L. L., 1962. Behavior of the firm under regulatory constraint. The American Economic Review 52 (5), 1052–1069.

Balk, B. M., 2012. Price and quantity index numbers: models for measuring aggregate change and difference. Cambridge University Press, New York, USA.

Bogetoft, P., Otto, L., 2015. Benchmarking with DEA and SFA. R package version 0.26 Edition.

Cambini, C., Rondi, L., 2010. Incentive regulation and investment: evidence from european energy utilities. Journal of Regulatory Economics 38 (1), 1–26.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European journal of operational research 2 (6), 429–444.

Cronin, F. J., Motluk, S. A., 2007. Flawed competition policies: designing markets with biased costs and efficiency benchmarks. Review of Industrial Organization 31 (1), 43–67.

Diewert, W. E., 2014. Decompositions of profitability change using cost functions. Journal of Econometrics 183 (1), 58–66.

Edvardsen, D. F., Førsund, F. R., Hansen, W., Kittelsen, S. A., Neurauter, T., 2006. Productivity and regulatory reform of norwegian electricity distribution utilities. Performance Measurement and Regulation of Network Utilities, 97–131.

Ek, G., 2010. Förhandsregleringens krav på effektiviseringar - intäktsramen för löpande kostnader. (The Ex Ante Efficiency Requirements - the Revenue Cap for Operating Expenditure) EI R2010:11, Energimarknadsinspektionen.
URL www.ei.se

European Commission, 2009. Directive 2009/72/ec of the european parliament and of the council of 13 july 2009-concerning common rules for the internal market in electricity and repealing directive 2003/54/ec.

Färe, R., Grosskopf, S., Lovell, C. K., 1985. The Measurement of Efficiency of Production. Dordrecht, Kluwer-Nijhoff, Boston.

Førsund, F. R., Kittelsen, S. A., 1998. Productivity development of norwegian electricity distribution utilities. Resource and Energy Economics 20 (3), 207 – 224.

Freixas, X., Guesnerie, R., Tirole, J., 1985. Planning under incomplete information and the ratchet effect. The review of economic studies 52 (2), 173–191.

Gammelgård, M., 2004. The network performance assessment model. Ph.D. thesis, Royal Institute of Technology KTH, Stockholm, Sweden.

Georgescu-Roegen, N., 1951. The aggregate linear production function and its applications to von neumann's economic model. In: Koopmans, T. C. (Ed.), Activity analysis of production and allocation, cowles commission monograph Edition. Vol. 13. John Wiley & Sons Inc., New York, Ch. IV, pp. 98–115.

Government of Sweden, 2009. Regeringens skrivelse 2008/09:206 Regelförenklingsarbetet, in swedish Edition. Vol. 2008/09:206. Government of Sweden.

Grifell-Tatjé, E., Lovell, C., 2003. The managers versus the consultants. The Scandinavian Journal of Economics 105 (1), 119–138.

Grifell-Tatjé, E., Lovell, C., 2015a. Decompositions of profitability change using cost functions: A comment. Working Paper WP03/2015, Center for Efficiency and Productivity Analysis (CEPA). School of Economics, University of Queensland, Australia.

Grifell-Tatjé, E., Lovell, C. K., 2015b. Productivity Accounting: The Economics of Business Performance. Cambridge University Press, New York, USA.

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.

Hjalmarsson, L., Veiderpass, A., 1992. Productivity in swedish electricity retail distribution. The Scandinavian Journal of Economics 94, 193–205.

Jamasb, T., Pollitt, M., 2003. International benchmarking and regulation: an application to european electricity distribution utilities. Energy policy 31 (15), 1609–1622.

Jamasb, T., Pollitt, M., 2008. Reference models and incentive regulation of electricity distribution networks: An evaluation of sweden's network performance assessment model (npam). Energy Policy 36 (5), 1788–1801.

Jamasb, T., Söderberg, M., 2010. The effects of average norm model regulation: The case of electricity distribution in sweden. Review of Industrial Organization 36 (3), 249–269.

Joskow, P., 2011. Incentive regulation in theory and practice: electricity distribution and transmission networks. University of Chicago Press, Chicago.

Kumbhakar, S. C., Amundsveen, R., Kvile, H. M., Lien, G., 2014. Scale economies, technical change and efficiency in norwegian electricity distribution, 1998–2010. Journal of Productivity Analysis 43 (3), 295–305.

Kumbhakar, S. C., Hjalmarsson, L., 1998. Relative performance of public and private ownership under yardstick competition: electricity retail distribution. European Economic Review 42 (1), 97–122.

Kuosmanen, T., Sipiläinen, T., 2009. Exact decomposition of the fisher ideal total factor productivity index. Journal of productivity analysis 31 (3), 137–150.

Lantz, B., 2003. Nätnyttomodellens regleringsprincip. Tech. Rep. 2003-398, School of Management, Gothenburg University.

Larsson, M., 2003. The network performance assessment model–regulation with a reference network. In: Proceedings of the Elforsk Market Design Conference, Stockholm. Vol. 17.

Larsson, M., 2005. The network performance assessment model: a new framework of regulating the electricity network companies. Licenciate thesis, Royal Institute of Technology KTH.

Liston, C., 1993. Price-cap versus rate-of-return regulation. Journal of Regulatory Economics 5 (1), 25–48.

Littlechild, S. C., 1983. Regulation of British Telecommunications' profitability: report to the Secretary of State, February 1983. Department of Industry.

Miguéis, V. L., Camanho, A., Bjørndal, E., Bjørndal, M., 2011. Productivity change and innovation in norwegian electricity distribution companies. Journal of the Operational Research Society 63 (7), 982–990.

Nillesen, P., Pollitt, M., 2007. The 2001-3 electricity distribution price control review in the netherlands: regulatory process and consumer welfare. Journal of Regulatory Economics 31 (3), 261–287.

Pollitt, M. G., 1995. Ownership and performance in electric utilities: the international evidence on privatization and efficiency. Oxford University Press, Oxford, UK.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org/

Ray, S., Mukherjee, K., 1996. Decomposition of the fisher ideal index of productivity: A non-parametric dual analysis of us airlines data. Economic Journal 106 (439), 1659–78.

Recordon, E., Rudnick, H., 2002. Distribution access pricing: application of the oftel rule to a yardstick competition scheme. Power Systems, IEEE Transactions on Power Systems 17 (4), 1001–1007.

Rudnick, H., Donoso, J., 2000. Integration of price cap and yardstick competition schemes in electrical distribution regulation. IEEE Transactions on IEEE Transactions on Power Systems 15 (4), 1428–1433.

Shleifer, A., 1985. A theory of yardstick competition. The RAND Journal of Economics, 319–327.

Törnqvist, J., Persson, L., 2006. Elnätbolagens anpassning till nätnyttomodellen. Msc thesis, School of Business, Economics and Law.

Tulkens, H., Eeckaut, P. V., 1995. Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. European Journal of Operational Research 3 (80), 474–499.

Turvey, R., 2006. On network efficiency comparisons: Electricity distribution. Utilities Policy 14 (2), 103–113.

Wallnerstrom, C., Bertling, L., 2008. Investigation of the robustness of the swedish network performance assessment model. IEEE Transactions on Power Systems 23 (2), 773–780.

Weitzman, M. L., Spring 1980. The ratchet principle and performance incentives. Bell Journal of Economics 11 (1), 302–308.