Incentives in Regulatory DEA Models with Discretionary Outputs
The Case of Danish Water Regulation

Heesche, Emil; Bogetoft, Peter

Document Version
Final published version

Published in:
Decision Analytics Journal

DOI:
10.1016/j.dajour.2022.100049

Publication date:
2022

Citation for published version (APA):
Heesche, E., & Bogetoft, P. (2022). Incentives in Regulatory DEA Models with Discretionary Outputs: The Case of Danish Water Regulation. *Decision Analytics Journal*, *3*, Article 100049. https://doi.org/10.1016/j.dajour.2022.100049

Incentives in Regulatory DEA Models with Discretionary Outputs
The Case of Danish Water Regulation

Heesche, Emil; Bogetoft, Peter

Document Version
Final published version

Published in:
Decision Analytics Journal

DOI:
10.1016/j.dajour.2022.100049

Publication date:
2022

Decision Analytics Journal 3 (2022) 100049

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

Incentives in regulatory DEA models with discretionary outputs: The case of Danish water regulation

Emil Heesche [a], Peter Bogetoft [b,*,1]

[a] University of Copenhagen, Department of Food and Resource Economics, Rolighedsvej 23, 1958 Frederiksberg C, Denmark
[b] Copenhagen Business School, Department of Economics, Porcelaenshaven 16A, 2000 Frederiksberg, Denmark

ARTICLE INFO

Keywords:
Data Envelopment Analysis
Incentives
Regulation
Discretionary outputs
Water sector

ABSTRACT

Data Envelopment Analysis (DEA) based cost norms are widely used to regulate natural monopolies like water, electricity, and gas networks. In the typical application, demand is considered fixed and non-controllable (non-discretionary), and the challenge is to incentivize the monopoly to provide the demanded services at the lowest possible costs. In this paper, we investigate the incentives of a DEA based regulation when some of the demand dimensions, the cost drivers, can, in fact, be controlled by the monopoly. In such cases, the classical DEA based regulation may lead to suboptimal incentives. Specifically, we examine both analytically and numerically the impacts of including a discretionary quality indicator in the benchmarking model used to regulate Danish water firms. We show that the catch-up period allowed in this regulation gives strong incentives to reduce costs since the firms can keep possible cost reductions for several years before the cost norm fully internalizes the cost reduction potentials. However, on the other hand, this scheme also provides weak quality incentives since it takes several years before the extra cost of increasing quality is fully internalized in the cost norm.

1. Introduction

Natural monopolies are not subject to the disciplining forces of a competitive market and are, therefore, often assumed to provide services at too high costs. Network firms in the water, electricity and gas industries are examples of such natural monopolies due to the high fixed costs associated with the construction of the networks. Most countries, therefore, have regulations on the services provided and the tariffs that such firms are allowed to charge. In Europe, it is common to use benchmarking-based regulations: the allowed revenues of the individual firms are determined by benchmarking its costs for the services provided against the best practice costs that can be inferred for all network firms. If the best practice costs of a given firm's service level are lower than the firm's actual costs, the firm is forced to reduce costs over time. In this way, the regulation relies on model-based pseudo competition. A commonly used benchmarking approach is Data Envelopment Analysis (DEA), sometimes combined with Stochastic Frontier Analysis (SFA). For recent references, see, e.g., Agrell and Bogetoft [1,2], Agrell et al. [3,4], Bogetoft [5],Bogetoft and Otto [6], Dai and Kuosmanen [7] and Ramanathan et al. [8]. For more on the theory of when DEA-based regulation may be optimal, see Bogetoft [5,9,10,11,12,13] and Bogetoft and Otto [6].

A key characteristic of typical network regulation is that the demand for different services is largely insensitive to prices, at least for the range of prices naturally allowed by regulators. That is, demand can be considered exogenous and fixed, and the challenge is mainly to determine the minimal costs of providing these services.

The regulatory problem can formally be formulated as an agency problem, cf. the references above; the aim is to find reimbursements to firms that are individually rational and incentive-compatible. Firms must have private incentives to reduce costs, and the allowed income must be sufficient to make production profitable, i.e., reimbursements must exceed the true underlying minimal costs. The advantage of DEA is precisely that it offers the smallest upper bound of the minimal cost when there is considerable initial uncertainty about the underlying cost function.

However, if demand is not exogenous, things become more complex, and DEA-based regulation may not be optimal. When a regulated firm can affect the services demanded, i.e., the cost drivers in the usual regulatory cost benchmarking model, there is a risk that the firm may use this strategically. By inducing either a very low or very high demand for certain services or, in other ways inducing consumers to demand a less common mix of services, it will be more difficult to find comparators in the benchmarking model. This means that the bias of the DEA estimated cost, i.e., the difference between the cost norm determined by benchmarking and the true underlying cost, increases, and the firm can extract extra information rents. We will illustrate this

* Corresponding author.
E-mail addresses: ehe@kfst.dk (E. Heesche), pb.eco@cbs.dk (P. Bogetoft).

1 Financial support from Independent Research Fund Denmark, Grant 9038-00042A, is greatly appreciated.

https://doi.org/10.1016/j.dajour.2022.100049
Received 19 February 2022; Received in revised form 2 April 2022; Accepted 13 April 2022
Available online 22 April 2022

in Section 2 below. The key insight is, however, simple. A regulated firm will try to maximize its profits, and this entails a search for levels of discretionary outputs where the benchmarking is more lenient. This also implies that the firm will not be guided by the potential value to consumers of the discretionary outputs. Therefore, the resulting outcomes are no longer socially optimal.

A good example of a potentially discretionary output is quality. Consumers may be willing to forgo quality, e.g., live with some services not always being delivered if the price reductions are sufficiently large. Regulation of quality is often handled by some add-on regulation rewarding (penalizing) the firms if services are delivered more (less) steadily than some threshold. A good example of such a scheme is the Norwegian so-called KILE system used to incentivize steady service delivery in electricity distribution, cf. NVE [14]. The use of add-ons to traditional cost benchmarking can make it easier to control the quality incentives and can be a natural approach, particularly if quality is a property of all the services being provided. In contrast to a more traditional add-on approach, in this paper, we investigate the direct integration of discretionary services such as quality into the cost benchmarking model. There are also examples of this approach in network regulation, e.g., in the Brazilian Electricity Regulatory Agency (ANEEL) model (Pessanha & Melo, 2021) to regulate electrical distribution firms.

While it is relatively straightforward to show the challenges of discretionary outputs in benchmarking-based regulations in general, as we will do in Section 2, we need more specific settings, i.e., more specific regulations and regulatory cost norms, to gauge the size and significance of the problem. Therefore, the bulk of this paper examines a specific regulatory framework, namely, that used to regulate Danish water firms.

Danish drinking and wastewater firms are natural monopolies and subject to two different regulations: economic and environmental.

Economic regulation is intended to reduce firms' costs and, thereby, prices. The Danish Water Regulatory Authority (DWRA), set a revenue cap for each firm. DWRA's goal is to set a revenue cap equal to the cost of the most efficient firms. This will force inefficient firms to reduce their costs. However, DWRA needs to consider different underlying conditions and allow a reasonable time for inefficient firms to catch up to best practices. Therefore, they need to use more advanced benchmarking models to compare the firms. DWRA uses both a DEA and SFA as part of this.

Environmental regulation is intended to secure a high quality of water and reduce pollution. This is typically done by setting minimum requirements for the firms. If the firms do not fulfil the requirement, the environmental regulator starts a dialogue with the firm to get the firm back on track. For some environmental parameters, such as pollution, firms need to pay a fee. However, this fee does not give high economic incentives because firms are allowed to charge the full amount of such fees to consumers in the economic regulation. Hence, the interaction between the two regulations is suboptimal, which has recently led to criticism.

In a perfect regulation, firms have incentives to reduce their costs and choose the level of quality that is socially optimal. This requires, however, good information about society's utility, viz., consumers' willingness to pay for quality. In Denmark, such information is still not available for most quality parameters. Moreover, even when it becomes available, it may not be simple to find the best balance between production costs and consumer preferences. Recall that this has to take place in a second-best world where the regulator has inferior information about the underlying cost function. Varying the services provided may come with extra information rents to the firms since they may become less comparable, making the benchmarking less effective, cf. also Bogetoft and Eskesen [15]. In fact, extra information rents that can be extracted when service mixes change may preclude such adjustment and make it second best optimal to stick to past production mixes, cf., e.g., Antle and Bogetoft [16].

In our analysis of Danish water regulation, we will therefore not look at the demand for different qualities but rather look at the supply side. We investigate the incentives firms have to change quality levels when quality is part of the cost drivers in the benchmarking model used to set the revenue cap.

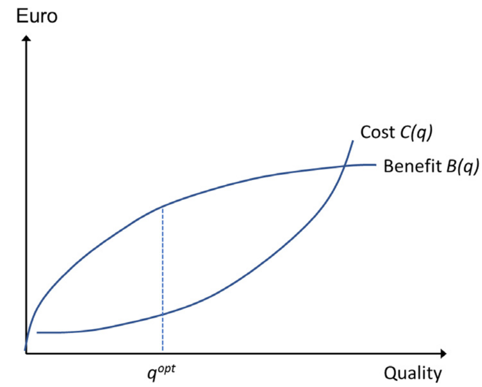

Fig. 1. Underlying cost and benefits of quality.

We first investigate the cost reduction incentives i*n the present regulation. Firms have incentives to reduce costs.* Assuming that quality is costly, *firms have incentives to limit quality* when it is not directly rewarded.

We next examine the incentives *if the regulator uses the same revenue cap formula but includes quality measures in the benchmarking model*. In this case, the quality incentives are more complicated. *Inefficient firms may still have strong incentives to limit their quality, but for some firms, a high-quality strategy may also become attractive since it may protect the firm against reductions in the revenue cap*. The details depend on the specific context of the firms and their placement in the production space.

The rest of this paper is structured as follows. Section 2 outlines the principal problems of using DEA-based revenue cap schemes when some of the cost drivers are discretionary. In Section 3, we first describe the current benchmarking model used to regulate Danish water firms and discuss the incentives. Next, we assume that quality is incorporated in the benchmarking model and show how this affects incentives. In Section 4, we show an application based on data from the Danish water sector, which confirms the discussion in the previous sections. In Section 5, we discuss the difference between static and dynamic incentives, taking into account the long-run ramifications of changes in costs or quality levels. In Section 6, we discuss the different limitations and extensions of our analyses. Concluding remarks are given in Section 7.

2. Regulation with discretionary outputs

To illustrate the complications of using a revenue cap scheme to incentivize the choice of an optimal quality level, let us assume that the true underlying benefits of quality to the consumers and the true underlying costs of quality to the firm are as illustrated in Fig. 1.

2.1. Perfect information

Now, if the regulator has perfect information about the underlying costs and benefits, he can determine the optimal quality level q^{opt}. In principle, the regulator, in this case, can simply demand that the firm implements this quality level, perhaps by revoking the firm's concessions rights if this quality level is not realized. A version of this idea is illustrated in Fig. 2.d.

There are, of course, many other schemes that could be used. For example, a series of workable revenue cap schemes are illustrated in Fig. 2. Faced with these schemes, the firm has private incentives to pick the optimal quality level since it maximizes its profits.

One possibility is to pay the firm whichever benefits $B(q)$ it creates plus a lumpsum amount A (subfigure a). Another is to use a two-price

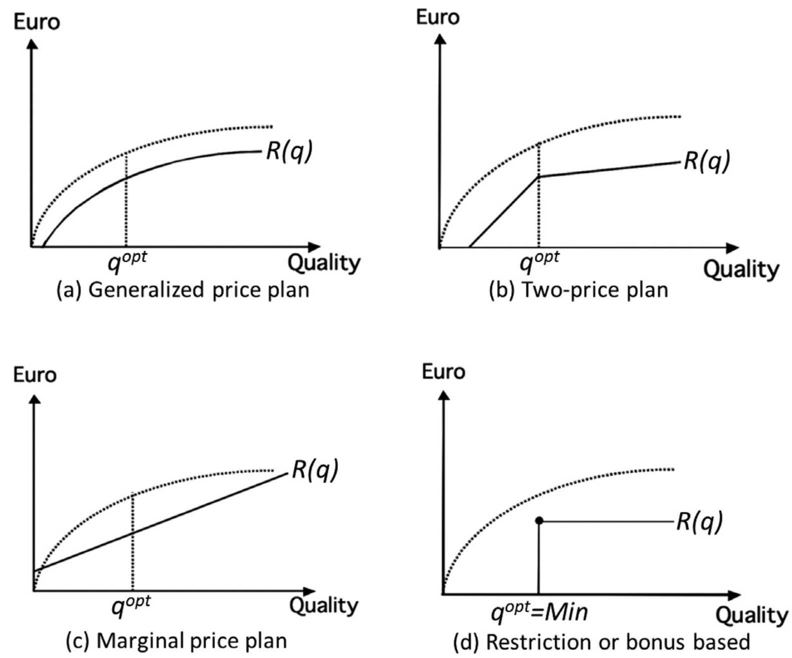

Fig. 2. Alternative incentive schemes.

schedule where the marginal revenue is relatively high, p_1 until the desired quality level and lower $(p_1 - p_2)$ for higher values of quality (subfigure b). A third option is to simply work with a marginal reward scheme with a slope p equal to the slope of the benefits (and cost) curves at the optimal quality level (subfigure c). Finally, as mentioned above, an option is to simply demand the required quality and to penalize if quality is below by, for example, revoking the firm's right to operate (subfigure d). Formally, we can express the revenue functions, $R(q)$ as:

- Generalized price plan $R(q) = A + B(q)$
- Two-price plan $R(q) = A + p_1 q – p_2 max\{q - q^{opt}, 0\}$
- Marginal price plan $R(q) = A + pq$
- Restriction or bonus based $R(q) = A \ if \ q \geq q^{opt}, = 0 \ otherwise$

It is worth noting that all of the above revenue cap schemes, except the generalized price plan or benefit-sharing rule (subfigure a), requires some knowledge of the underlying cost function. Only in this way can the q^{opt} value be determined, and hereby the last three schemes can be defined.

2.2. Imperfect information

Now, consider the case of imperfect information where the firm knows its cost function, but the regulator does not. The regulator only knows the cost and quality levels of a selection of other firms in the sector. The core of benchmarking-based revenue caps is that the regulator uses such observations to set the revenue cap by estimating the underlying costs of providing different service levels. The C^{DEA} function in Fig. 3 below may illustrate the resulting revenue cap if DEA is used to approximate best practice costs. Hence, in this case, the DEA estimated cost function C^{DEA} serves as the revenue cap function R above.

In such a setting, the firm's profit (information rent) becomes the difference $C^{DEA}(q) - C(q)$. A profit-maximizing firm will, in this case, find the quality level that leads to the largest distance between $R = C^{DEA}$ and C, i.e., between the red and the blue curves in Fig. 3.

Three important observations can be made based on this.

First, we note that the resulting revenue cap function $R = C^{DEA}$ is now convex, while it was concave in all cases in Fig. 2 above.

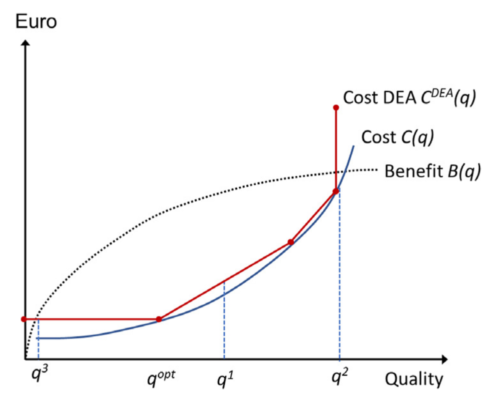

Fig. 3. Using a DEA-based revenue cap.

Therefore, the firm will maximize the difference between a convex revenue $C^{DEA}(.)$ and a convex cost function $C(.)$.

Second, there may be many solutions to the firm's profit maximization problem, and it is unlikely that the optimal quality level will be implemented. In the illustration, the estimated cost function must provide a close approximation of the true cost function around the optimal production level to support near-optimal quality levels such as q^1.

Third, by using the fact that the cost function will typically have horizontal and vertical parts – or more general non-full facets – it is clear that it may sometimes be optimal to simply choose the minimal possible quality level such as q^3, while in other cases, it may be optimal to choose a very high-quality level, such as q^2.

In summary, it is intuitively clear that using a DEA-based revenue cap is likely to lead to suboptimal endogenous choices of the quality level. DEA-based regulations work well when the outputs are exogenous but not when outputs are discretionary.

We will examine this intuitive insight further below, where we look at a regulation in more detail and examine the incentives to increase quality both analytically and empirically.

3. The Danish regulation of water firms

The specific regulation of Danish water firms has a series of institutional details. For a precise description, see (Danish Water Regulatory Authority, 2020). In our analysis, we ignore some of the institutional details since they are not consequential to the incentives.[2]

The core of the regulation is simple and is based on a traditional RPI-X[3] regulatory framework. The allowed revenue to a firm in period t is determined as

$$R_t = R_{t-1} \cdot \left(1 - X^{ge}_{t-1:t} - X^{sp}_{t-1}\right) \tag{1}$$

$$X^{sp}_{t-1} = \frac{\left(\frac{R_{t-1}-\hat{C}_{t-1}}{R_{t-1}}\right)}{P} \tag{2}$$

where R_t is the revenue cap, $X^{ge}_{t-1:t}$ is a general efficiency requirement to account for expected price developments and productivity growth from time $t-1:t$, and X^{sp}_{t-1} is an individual efficiency requirement that forces inefficient firms to catch up to best practice. The individual efficiency requirement is lagged one year because the regulator cannot obtain the relevant information for the calculations at time t. Finally, $\hat{C}_{t-1}$ is the firm's best practice cost level as determined by the benchmarking model and the data from all regulated firms. We will often refer to this simply as the cost norm.

The individual efficiency requirement is divided by P because DWRA as most regulators assumes it takes some time to catch up to best practices. DWRA assumes that it takes eight years to meet the full efficiency requirement. Therefore, DWRA uses $P = 8$. To make the analyses more general we will work with a catch-up period of $P > 1$ except for the specific simulations on Danish data where we of course use $P = 8$.

The specific requirement X^{sp}_{t-1} is not, as is often the case, calculated by comparing the actual costs last period, x_{t-1} and the cost norm $\hat{C}_{t-1}$ and allowing for a certain period to catch up. Instead, the maximum allowed costs, the revenue cap R_{t-1}, are compared with the cost norm and are gradually reduced. In effect, this will force actual costs down as well.

For simplicity, let us suppress the time notation for now. This does not significantly influence the one-period conclusions.[4] We return to the dynamic incentives in Section 5. Rewriting the revenue cap formula, we can therefore express the allowed revenue as

$$R = R^* \cdot (1 - X^{ge}) - \frac{1}{P}R^* + \frac{1}{P}\hat{C} \tag{3}$$

where R^* is the last period's revenue cap and, as such, a constant. The only way that the revenue cap is affected in a given period is, therefore, via the cost norm $\hat{C}$.

The cost norm is calculated in a so-called "best-of-two" benchmarking model. DWRA operates with both a DEA and a SFA model, and for each firm, the highest best-practice cost from the two models is used as the cost norm. Our analysis in this paper applies in most cases for both types of models, but for simplicity, we only use DEA in the numerical examples.

DWRA sees the production of water firms as a transformation of 1 input into 2 outputs. The input, x, is the firm's controllable costs (hereafter costs), and the outputs, y, are so-called net volumes. The net volumes can be thought of as aggregations of all the tasks the firm needs to undertake. The tasks are split into operational and capital tasks, giving two separate net volumes. For a detailed description, see [17] . For this paper, an important property of the net volumes is that DWRA assumes that they are fixed and not directly controllable by the firms. The firms can, therefore, not influence the outputs but only try to reduce the costs.[5] Therefore, the DEA and SFA models used by DWRA also focus on input-oriented efficiencies from which the cost norms can be calculated. Finally, it should be mentioned that DWRA assumes constant returns to scale in their production models.

In the following, we will assume that the firms seek to maximize profits

$$\Pi = R - x \tag{4}$$

In regulatory studies, this is a common assumption — although one that is not always in complete accordance with reality.[6]

Before closing this introduction to the core of the regulatory setting, let us note that we use a simplified notation. The cost norm $\hat{C}$ is really a function of the outputs delivered, and its structure is determined by the costs and outputs of all the regulated firms. Assume that there are n firms, the firm that we study, and $n-1$ other firms numbered $i = 1, \ldots, n-1$. The firm we study has used input x to produce outputs y, while the other firms have used input x^i to produce outputs $y^i, i = 1, \ldots, n-1$. Now let the vector of inputs be $\boldsymbol{x} = (x^1, x^2, \ldots, x^{n-1}, x)$ and matrix of outputs be $\boldsymbol{y} = (y^1, y^2, \ldots, y^{n-1}, y)$. In this case, the benchmarking model determines an estimated cost function $C(. \mid \boldsymbol{x}, \boldsymbol{y}) : R^2 \to R$. The cost norm for the firm that we study is then $C(y \mid \boldsymbol{x}, \boldsymbol{y})$, which we have simply denoted $\hat{C} = C(y \mid \boldsymbol{x}, \boldsymbol{y})$.

3.1. Cost reduction incentives

If a firm with a fixed output changes its cost x, it will directly reduce profit by the same amount. It will also have an indirect effect by affecting revenue. From the definition of the revenue cap, we see that

$$\frac{\partial R}{\partial x} = \frac{1}{P} \cdot \frac{\partial \hat{C}}{\partial x} \tag{5}$$

Increasing the cost may lead to a higher cost norm and, thereby, less of a reduction in the revenue. Since the difference between the revenue cap and cost norm is only planned to be eliminated over P periods, the possible revenue gain is only $\frac{1}{P}$ of the increase in the cost norm. Summing up, we have

$$\frac{\partial \Pi}{\partial x} = \frac{1}{P} \cdot \frac{\partial \hat{C}}{\partial x} - 1 \tag{6}$$

Hence, *firms have strict cost reduction incentives* as long as

$$\frac{\partial \hat{C}}{\partial x} < P \tag{7}$$

When the cost norm is determined by a DEA model, this is always the case. Inefficient firms do not affect the frontier, and for an inefficient firm, we thus have $\frac{\partial \hat{C}}{\partial x} = 0$. Efficient firms may directly set the cost norm, in which case we have $\frac{\partial \hat{C}}{\partial x} = 1$. In the few cases where a firm is efficient but not strictly superefficient, i.e., it is located on a facet

[2] Only one caveat is important to mention here. The specific efficiency requirement may, in theory, be negative if industry costs increase or if a firm increases the services it delivers. This possibility is, however, excluded in the regulation. The regulator uses $\max(X^{sp}_{t-1}, 0)$ instead of X^{sp}_{t-1}. However, a negative efficiency requirement is rare and will therefore be ignored for now.

[3] RPI-X refer to a regulation that allows the revenue to increase with the retail price inflation (RPI) but at the same time requires revenue to be reduced due to expected improvements in the productivity, i.e., except for a general and sometimes an individual efficiency requirements (X).

[4] The only 'mistake' is a time lag. The revenue in a given period does not depend on this period's costs, but possibly the costs of the last period. We ignore the need for a one-period discounting to account for this.

[5] In reality, firms can influence the net volumes to some extent, e.g., by introducing spare capacity via extra assets that are not necessary, but we ignore such blunt approaches to playing the regulations. Instead, we will later focus on the introduction of another output, quality, which we will assume is controllable by the firm.

[6] In reality, the Danish water companies are not allowed to withdraw profit, but they can use any excess revenues inside the firms. Firms can, for example, save the profit to buffer themselves against future costs, or they can use it to allow themselves some slack. The firms are in general municipalities or cooperatively owned, and some firms argue that they would rather pay back the profit to consumers because they think it is the socially responsible thing to do.

spanned by other efficient firms, we even have $\frac{\partial \hat{C}}{\partial x} = 0$. Either way, firms have a good incentive to reduce costs when a DEA cost norm is used.[7]

We also note that the strong cost reduction incentives result from the regulator allowing for a P-years catch-up period. The efficient firms have incentives to reduce their costs only because DWRA divides the efficiency requirement by P. If DWRA did not do this, $\frac{\partial \hat{C}}{\partial x}$ would need to be less than 1 to ensure the firms have strict cost-reducing incentives. In this case, the efficient firms would be neutral to a change in their costs as the marginal change in the profit would be zero.[8] If slack, therefore, has some value to the firms, the cost reduction incentives would vanish when the catch-up period is one. As we will show in Section 6, however, even with strong preferences for slack, the present regulation would give strong cost reduction incentives when DWRA allows for an 8-year catch-up period.

It is interesting to note how *the catch-up period strengthens the cost reduction incentives*. The catch-up period is usually thought of as a reflection of the time needed to implement the technical, managerial and organizational changes necessary to implement best practices. The catch-up period, however, also plays another role. It allows the firm to reap the gains from cost reductions for a period of time before the norm catches up. In other words, it serves to reduce the so-called ratchet effect [18] that may destroy incentives.

3.2. Quality expansion incentives

Let us now turn to quality incentives. To make the quality decisions explicit,[9] let us assume that, in addition to the non-discretionary outputs y, there is now a discretionary quality variable q. Additionally, let us assume that there is a cost associated with higher qualities. Let the true underlying costs of providing quality q when the non-discretionary outputs are y be $c^q(q \mid y)$. We assume that larger qualities lead to higher costs, i.e., $\frac{\partial c^q}{\partial q} > 0$.[10]

In reality, it may of course be difficult to determine the cost of quality. The companies do not choose the exact level of quality but rather a given risk for lack of quality. For example, if a company has an old pump, there might be a high risk that this pump will suddenly break down and the water flow stop. However, the company do not know the exact risk. To reduce the risk of low quality, the company can buy an extra pump in reserve. The company can easily calculate the price for having an additional pump but can only estimate the likelihood of how this will improve the quality. It should be possible for big companies with many resources and good asset management to get a good estimation. For smaller companies, this is not trivial to do. In the following analyses, we will therefore not make any more specific assumptions about the cost of quality. Instead, we will focus on how the allowed revenue depends on the quality. As will be clear, companies in many cases do not need to know their precise costs of quality. We will provide bounds on the (marginal) cost of quality that suffices for the companies to decide whether to increase or decrease quality.

Since quality is not part of the present cost norm, $\hat{C} = \hat{C}(y|\boldsymbol{x}, \boldsymbol{y})$, it is obvious that firms will have no incentives to provide quality; it will only increase costs, and as we know from Section 3, increasing costs will decrease profit for both efficient and inefficient firms. More formally, we have that the derivative of profit with respect to quality is

$$\frac{\partial \Pi}{\partial q} = \frac{\partial R}{\partial q} - \frac{\partial x}{\partial q} = \frac{\partial R}{\partial x}\frac{\partial x}{\partial q} - \frac{\partial x}{\partial q} = \left(\frac{\partial R}{\partial x} - 1\right)\frac{\partial x}{\partial q} = \left(\frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial x} - 1\right)\frac{\partial x}{\partial q} < 0 \tag{8}$$

Hence, as long as quality q does not affect the cost norm, firms will try to reduce quality. Therefore, we can conclude that DWRA's *present regulation gives the desired incentive to reduce costs but gives the wrong quality incentives*, which is no surprise since quality is not part of the cost norm.

Let us now see what happens if quality is included in the cost norm, i.e., when $\hat{C} = \hat{C}(y, q|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q})$. In this case, the revenue depends on both the costs and the quality, $R(x, q)$, and since x depends on q, we obtain

$$\frac{\partial \Pi}{\partial q} = \frac{\partial R}{\partial q} + \frac{\partial R}{\partial x}\frac{\partial x}{\partial q} - \frac{\partial x}{\partial q} \tag{9}$$

The first term is the direct effect of quality on the revenue cap. The second term is the effect on the revenue from the cost increase necessary to increase quality. The last is the direct effect on profit from the extra cost of producing higher quality. We can rewrite this as

$$\frac{\partial \Pi}{\partial q} = \frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial q} + \frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial x}\frac{\partial x}{\partial q} - \frac{\partial x}{\partial q} \tag{10}$$

or similarly

$$\frac{\partial \Pi}{\partial q} = \frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial q} + \left(\frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial x} - 1\right)\frac{\partial x}{\partial q} \tag{11}$$

The factor in parentheses is the marginal profit resulting from a marginal cost increase of 1, as we also saw in Section 3. We know that it is negative. In the quality formula, it is multiplied by the marginal cost of quality. The first term is the direct impact of quality on the cost norm. It is positive, but it only enters the equation with a factor of $\frac{1}{P}$.

The factor $\frac{1}{P}$ on the $\frac{\partial \hat{C}}{\partial q}$ term points to another interesting impact of the assumed P periods to catch. This leads to much weaker quality incentives since an increase in the estimated underlying costs $\hat{C} = \hat{C}(y, q|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q})$ from an increase in quality is only expected to be fully accounted for in P periods. *Therefore, the length of the catch-up period negatively affects the incentives to increase quality. On the other hand, it strongly rewards reductions in quality*. The regulator only requires the cost to be reduced by $\frac{1}{P}$ of what the cost norm suggests will be saved by reducing the quality level. In addition, there is an indirect benefit if the actual costs decrease. In this case, firms keep the cost reduction except that there will be a small reduction in the cost norm due to the $\frac{1}{P}\cdot\frac{\partial \hat{C}}{\partial x}$ factor.

Consider as a small example a case where a firm can increase its quality level by 1 if it spends 1 DKK[11] extra. Likewise, it can save 1 DKK if it reduces the quality by 1. Additionally, let us assume that the cost norm model perfectly depicts these changes and that the catch-up period is $P = 8$ as in the Danish regulation. We then have the following changes in profit:

Increase quality: $\frac{1}{8}\cdot 1 + \left(\frac{1}{8}\cdot 1 - 1\right)1 = -\frac{6}{8}$

Decrease quality: $\frac{1}{8}\cdot(-1) + \left(\frac{1}{8}\cdot(-1) - 1\right)(-1) = 1$

This shows how using a long catch-up period increases the incentives to lower quality. *The reduction in quality will be partially forgiven in the first seven periods. Similarly, if the firm was to increase quality, it would take seven periods before this was fully accommodated in the revenue cap.*

[7] In the case of an SFA norm, it is theoretically feasible that the marginal impact on the cost norm may be larger than eight and hence deprive the firms of incentives to reduce costs. It is, however, not a likely outcome since it would require a very large change in the estimated SFA function. In the present Danish SFA model, it never happens.

[8] Note that this is only true in DEA. In SFA, the functional form is not changed too much; the efficient companies typically have $\frac{\partial \hat{C}}{\partial x} < 1$ because part of the cost increase is interpreted as an increase in bad luck (noise term). In addition, the SFA frontier is calculated based on all the companies and not only the efficient ones. Therefore, a cost increase for a single firm will only influence the frontier to some extent.

[9] We can think of quality as an output and, thereby, one of the y-dimensions.

[10] We normally think of quality as being positively correlated with costs, but it could also be negatively correlated if the costs to repair quality exceed the costs to achieve high quality. It could, for example, be more expensive to repair a broken pipeline than to properly maintain it. The correlation between quality and costs is discussed in [17] .

[11] DKK is the Danish currency

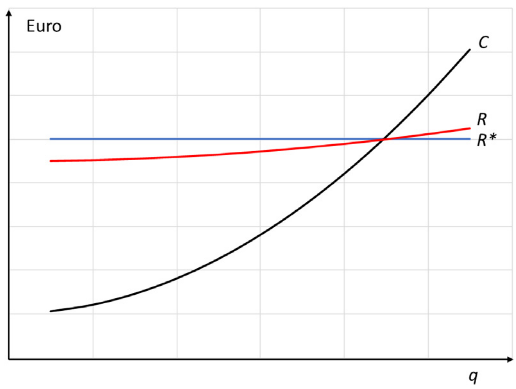

Fig. 4. Illustrative example of the relationship between costs and the revenue cap.

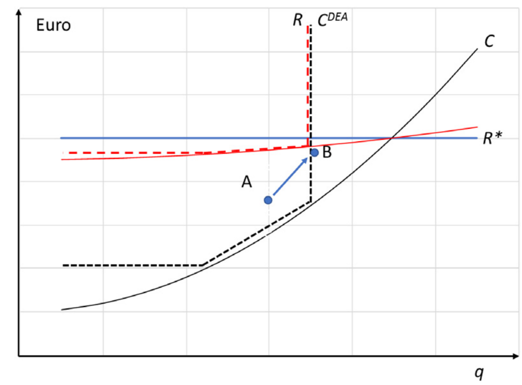

Fig. 5. Illustrative example of the relationship between costs and the revenue cap with DEA-based cost norms.

In the next sections, we illustrate these incentives to lower costs and quality using first a simple example and next a DEA model akin to the model used to define the cost norms of Danish water firms. In these illustrations, we stick to the Danish use of $P = 8$.

3.3. Graphical illustrations

To further illustrate the analytical insights above, let us consider the relationship between quality, cost and revenue cap. If the general productivity requirement X^{ge} is 0, we have

$$R = R^* \cdot (1-0) - \frac{1}{8}R^* + \frac{1}{8}\hat{C} = \frac{7}{8}R^* + \frac{1}{8}\hat{C} \tag{12}$$

where R^* is the last period's revenue cap and, as such, a constant. In Fig. 4, we keep the net volume constant and only consider the impact of changing the quality level. We assume that the cost norm is a perfect approximation of the true cost function, $\hat{C} = C$. We see that the marginal change in revenue when the quality increase is $\frac{1}{8}$ of the change in the cost norm. *Therefore, if the cost norm provides a reasonably good approximation of the true costs, all firms will provide the minimal possible quality level.*

In reality, the cost norm is not a perfect replication of the true, underlying cost function. Instead, it provides an upper bound on the true costs. Assuming a variable returns to scale (VRS) DEA model, the approximation may look like the piecewise linear black dotted function in Fig. 5 below.

In this case, the allowed revenue will be slightly larger, as illustrated by the piecewise linear red dotted function. It is not, in general, going to lower the incentives to reduce quality. A firm that increases quality is still only rewarded with $\frac{1}{8}$ of the estimated best practice changes in costs.

In some cases, however, an alternative strategy for a firm may be to choose a quality level slightly above the highest quality level of the previous period. This could correspond to a movement from A to B in Fig. 5 below. By doing so, a firm can obtain a much larger marginal increase in revenue since the cost estimate is upwards biased to an extreme degree in B. Hence, in some cases, there may actually be strong incentives to increase quality due to the bias in the estimated cost norm.

In Fig. 6 below, we illustrate the above cases using 2019 data from the Danish waterworks. Summary statistics are given in Table 1. We use water wastage as the quality indicator, which is presently not part of the regulation. Because water wastage is an undesirable output[12], we use 100 minus water wastage as a percentage. A value of 92 means that the end consumers receive 92% of the water fed into the network.[13]

[12] An undesirable output is characterized as an output that is negatively correlated with the input. Hence, it is essentially an input that can be substituted with other inputs.

[13] It can be problematic to use percentage in DEA [19], but we ignore this, as the specifics are not very important to examine the companies' incentives.

In Fig. 6, we only illustrate the firms with a net volume below 60. As can be seen from the summary statistics, this nevertheless covers the vast majority of the firms.

The green facets in Fig. 6 are the fully dimensional efficient facets[14], and the grey facets are those that are not fully dimensional. The blue dots show the efficient firms, the red dots show the inefficient firms and the red line shows an inefficient firm gradually adjusting its quality. We see that most firms are projected in the cost direction on fully dimensional efficient facets.

Imagine now an individual firm considering whether to change its quality level. As illustrated by the red curve, we assume that net volume is fixed and that changing quality comes with a cost.

We start by examining the area around A1. A1 is defined as the lowest level of quality where the firm is benchmarked against a fully dimensional efficient facet.

If the firm chooses to marginally reduce its quality from here, it will now be benchmarked against a non-fully dimensional efficient facet, where the cost norm is constant. The firm, therefore, profits from further reducing its quality — the revenue cap stays constant while the cost reduces. At A3, the firm's costs will be equal to the cost norm, and continuing to lower the quality level will lead to the firm establishing its own cost norm. This means that the revenue cap will also decline, but since the revenue cap is only affected by $\frac{1}{8}$ of the cost norm reduction, it still pays to reduce quality.

If the firm instead considered increasing the quality from A1, it is benchmarked against the fully dimensional efficient facets up to point A2. Between A1 and A2, the incentives are ambiguous. It is possible that the marginal increase in cost is low and that the cost norm increases enough to make the marginal revenue higher than the marginal change in cost. This seems unlikely but not impossible since we know that the DEA cost frontier is biased, and the more so, the fewer firms are located in the neighbourhood of A1 and A2, cf., e.g., [21]. Moreover, in reality, the firms' costs of providing quality are not a smooth function of the quality level. Quality provision may require both fixed and variable costs. A firm may therefore need some starting assets to increase its quality. It will hereafter become marginally less expensive to improve quality until the assets reach their maximum capacity. Hereafter, the firm may need new assets, and so on. In such cases, having optimal quality levels somewhere between A1 and A2 becomes more likely since the marginal quality costs may be low, and the fixed costs may already be sunk.

Now assume that the firm is considering quality levels around A2, where the firm again meets a non-fully dimensional efficient facet. In this case, the firm may have incentives to increase its quality. If the firm

[14] A fully dimensional efficient facet is a facet that is estimated based on a convex combination of the observed data together with the assumption of returns to scale but not the assumption of free disposability [20].

Table 1
Summary statistics.

Statistics	N	Mean	St. Dev.	Min	Pctl. (25)	Pctl. (75)	Max
Costs (in mill. DKK)	71	29.46	29.51	7.03	13.29	31.71	174.61
Quality	71	92.42	3.27	81.92	90.76	94.42	98.63
Net volume (in mill. DKK)	71	29.00	26.79	7.70	14.28	31.93	163.51

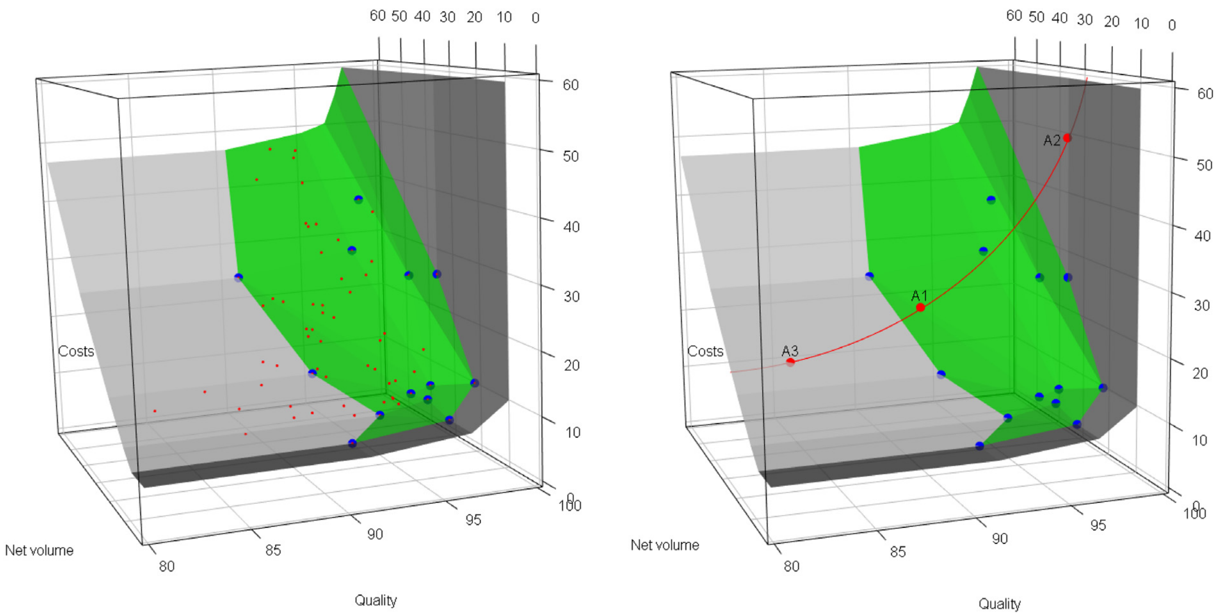

Fig. 6. DEA-VRS frontier based on reel data and seen from the inside.

increases the quality up to A2 and thereby is identified as fully efficient, its cost norm will likely make a discrete jump. The firm, therefore, goes from having a low cost norm compared to its actual costs to suddenly having a cost norm that is equal to its costs. Even though the jump in the cost norm is only internalized by the fraction $\frac{1}{8}$, this may suffice to make the increase in quality worthwhile if the marginal quality cost is not too large. It is, therefore, possible that firms located close to the non-fully dimensional efficient facet will have incentives to increase quality. It is, however, also clear that it will not have incentives to increase quality any further, since after A2, the firm will determine its own cost norm, and only the fraction $\frac{1}{8}$ of the actual marginal cost will be internalized in the revenue cap.

In summary, *we know that the firm benefits from lowering quality when it starts at A1. The best option, in this case, is to choose the lowest possible quality level. Between A1 and A2, the firm may have marginal incentives to increase quality if the DEA norm provides a bad approximation of the cost of changing quality. Close to A2, it is likely that the firm may have incentives to increase quality to the A2 level, but it will have no incentives to increase quality above A2.*

In our discussions above, we have made a few simplifications compared to the model and the regulation used by DWRA. First, we have assumed that the revenue cap can increase above the old revenue cap. In reality, this is not the case, i.e., the revenue is always capped at R*. This, of course, limits the incentives to increase quality further. Second, we have assumed that the DEA-based cost norm relies on variable returns to scale. In reality, however, the model used by DWRA assumes constant returns to scale. Hence, there are no parts of the cost function that go to infinity.[15] However, due to the existence of non-fully dimensional efficient facets, it is still possible to have a large discrete jump in the cost norm.

[15] To see this, assume, namely, that a firm has demonstrated the possibility of producing (y^0, q^0) at the cost of x^0. Now consider a firm producing (y, q). By the CRS assumption and free disposability, the costs of doing so can never exceed $x^0 \cdot \max\left(\frac{y}{y^0}, \frac{q}{q^0}\right)$.

4. Simulating marginal and discrete quality incentives in Danish water

In this section, we examine in more detail when Danish water firms have incentives to make marginal and discrete changes in the quality level. As in the previous section, we use the firms' costs as input, the sum of the net volumes as a fixed output and water wastage as the quality proxy, which is also an output. Also, since we now simulate specifically on the Danish data, we assume $P = 8$. We use VRS for illustrative purposes.

4.1. Marginal quality incentives

Recall from Eq. (11) above that

$$\frac{\partial \Pi}{\partial q} = \frac{1}{8} \cdot \frac{\partial \hat{C}}{\partial q} + \left(\frac{1}{8} \cdot \frac{\partial \hat{C}}{\partial x} - 1\right) \frac{\partial x}{\partial q} \tag{13}$$

If a firm is inefficient, it does not affect the best practice cost norm, i.e., $\frac{\partial \hat{C}}{\partial x} = 0$, and therefore, the firm has incentives to increase quality only if

$$\frac{\partial x}{\partial q} \leq \frac{1}{8} \cdot \frac{\partial \hat{C}}{\partial q} \tag{14}$$

i.e., only if the firm's marginal cost of increasing quality is at most one-eighth of the marginal cost according to the cost norm. If the firm is fully efficient, we have $\frac{\partial \hat{C}}{\partial x} = 1$, and therefore marginal quality improvements are attractive as long as $\frac{\partial x}{\partial q} \leq \frac{1}{7} \cdot \frac{\partial \hat{C}}{\partial q}$. This does, however, not hold for any of the efficient companies if we assume that their marginal costs with respect to quality are equal to the marginal estimated cost norm with respect to quality, $\frac{\partial x}{\partial q} = \frac{\partial \hat{C}}{\partial q}$.

In Fig. 7, left panel, we show the cost norm for an efficient firm as its quality increases. The function is derived simply by an iterative process where quality changes in small steps while the net volume and costs are kept fixed. Additionally, in the right panel, we have calculated

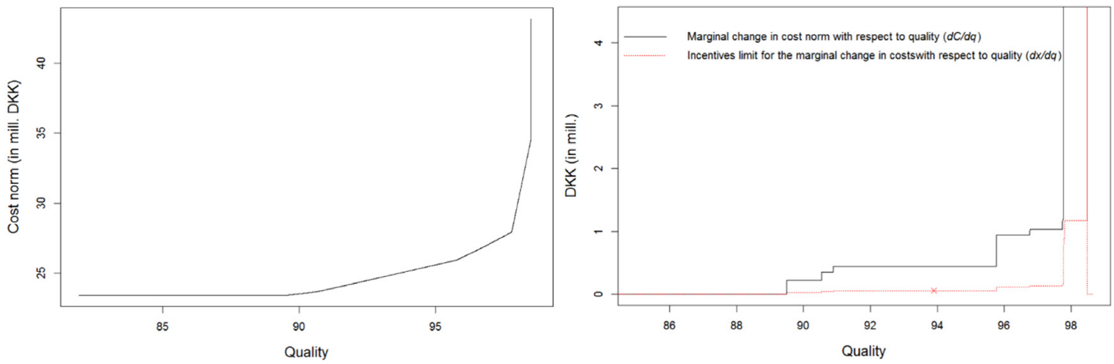

Fig. 7. Relationship between quality, cost norm, marginal changes in cost norm, and marginal quality cost.

$\frac{\partial \hat{C}}{\partial q}$. Note that the costs of the firm will likely change with the quality as well, but we do not know how much, and we can ignore this as long as the firm is inefficient since then the cost norm is not affected. We stop the iterative process when a firm becomes efficient. To calculate when the firm becomes efficient, we assume that its marginal costs are equal to the marginal costs on the frontier.[16] This assumption will only influence the illustration when the firm becomes efficient.

Quality goes from the lowest observable level of quality to the highest among all firms. We observe that the difference between the highest and lowest cost norms is approximately 20 mill. DKK. This means that if the firm chooses the highest level of quality (the exact level where they become fully efficient), it will have a cost norm that is 84 per cent higher than if it had chosen the lowest level of quality. Therefore, the level of quality in this model has a huge effect on this firm's cost norm and, thereby, potentially also on the profit.

In the right panel, we show the marginal cost norm $\frac{\partial \hat{C}}{\partial q}$ in black. In red (and dashed), we show the maximal marginal quality costs $\frac{\partial x}{\partial q}$ for which the firm has incentives to marginally increase quality. As shown above, the red dashed curve is $\frac{1}{8}$ of the black curve since the firm has incentives to increase quality as long as $\frac{\partial x}{\partial q} \leq \frac{1}{8} \cdot \frac{\partial \hat{C}}{\partial q}$.

For the first many iterations, the cost norm is constant, $\frac{\partial \hat{C}}{\partial q} = 0$. This is because the firm is benchmarked against a non-fully dimensional efficient facet. At approximately a quality level of 89, the cost norm starts to increase. Hereafter, the firm is benchmarked against several different facets, which results in a steeper and steeper slope.

Finally, we observe a huge jump in the cost norm when quality reaches approximately 98. This is because the firm is now identified as being efficient. The size of the jump here depends on the marginal cost of quality and, for the sake of the illustrations, has been assumed to be equal to the marginal costs in the cost norm, $\frac{\partial x}{\partial q} = \frac{\partial \hat{C}}{\partial q}$.

In the right panel, if the firm has a $\frac{\partial x}{\partial q}$ that is lower than the red dashed line, it will have incentives to marginally increase quality in the specific area.[17]

The firms' present quality level is marked with the red cross. This specific firm has a quality level of 93.9, and it needs a marginal quality level of $\frac{\partial x}{\partial q} < 55,280$ to have incentives to marginally increase quality from its given level.

The greater the firm wants to increase quality, the higher the marginal cost of quality $\frac{\partial x}{\partial q}$ can be. Immediately before it becomes fully efficient (where quality is approximately 98.5), the firm has incentives to increase quality even if $\frac{\partial x}{\partial q}$ is extremely high.

The marginal change in the cost norm differs between the firms based on their net volumes and level of quality. Therefore, the incentives differ as well. The discussion above is an example of the marginal change in the cost norm and the corresponding incentives for one firm. In Fig. 8, we show the incentives for four other firms. The firms differ in size based on their net volumes. They are selected to represent the 20th, 40th, 60th and 80th percentiles of the net volume values.

The first two firms (DMU 1 and DMU 2) are both benchmarked to a non-fully dimensional efficient facet. They, therefore, do not have incentives to marginally increase quality. The incentives to marginally increase quality in the last two firms (DMU 3 and DMU 4) depend on their marginal cost of quality $\frac{\partial x}{\partial q}$. As explained above, it must be below the red dashed line to give strict quality improvement incentives.

Note that the first three firms (DMU 1-3) first become efficient when they have the highest observed quality. Therefore, for these firms, we cannot see the red dashed line going down to zero subsequently, which is, of course, the case if they could increase quality even more.

4.2. Discrete quality incentives

Above, we examined the incentives to marginally increase quality for five firms.

Unfortunately, *the setting does not have the standard textbook regularity of concave revenue and a convex cost function. It, therefore, does not suffice to look at local incentives to determine global incentives*. The marginal revenue from quality curves and the marginal cost of quality curves may cross at multiple quality levels.

Truly, we can assume that the cost of increasing quality for an individual firm is convex, but the revenue cap is essentially

$$R = R^{*} \cdot (1 - X^{ge}) - \frac{1}{8} R^{*} + \frac{1}{8} \hat{C} \tag{15}$$

and therefore, convex as well. This means that a firm considering varying its quality should ideally look at the possible gains from all reductions and expansions of quality to determine the optimal quality level. Put differently, let us assume that a firm initially produces q^{*} and now considers changing to another quality level q. If it can estimate the change in its cost of producing quality as Δc and the corresponding change in revenue cap as ΔR, where

$$\Delta c = c(q) - c(q^{*}) \leq \frac{1}{8} (\hat{C}(q) - \hat{C}(q^{*})) = \Delta R \tag{16}$$

and $c(.)$ is the firm's cost of quality for the given net-volume level; then, it is attractive to move from q^{*} to q. Put differently, the optimal quality level for the firm is determined as the solutions to

$$\max_{q} \frac{1}{8} \hat{C}(q) - c(q) - \left(\frac{1}{8} \hat{C}(q^{*}) - c(q^{*}) \right) \tag{17}$$

If a firm is inefficient and if we believe it is no more efficient in producing quality than the best practice firms are, then there are only two potentially optimal solutions: one is to set quality as low as possible, and the other is to make a discrete increase in quality such

[16] That is, we assume that $\frac{\partial x}{\partial q} = \frac{\partial \hat{C}}{\partial q}$. We do this to get a better approximation of the quality level where efficiency is obtained. As soon as we find that the firm is efficient, we stop examining the incentives. In this way, we still only examine the incentives for an inefficient firm, but we find a more realistic level of quality before it is identified as being efficient.

[17] Note that we have zoomed in, leaving out the highest defined $\frac{\partial \hat{C}}{\partial q}$ and the last jump from Fig. 7 where the firm goes from being inefficient to efficient.

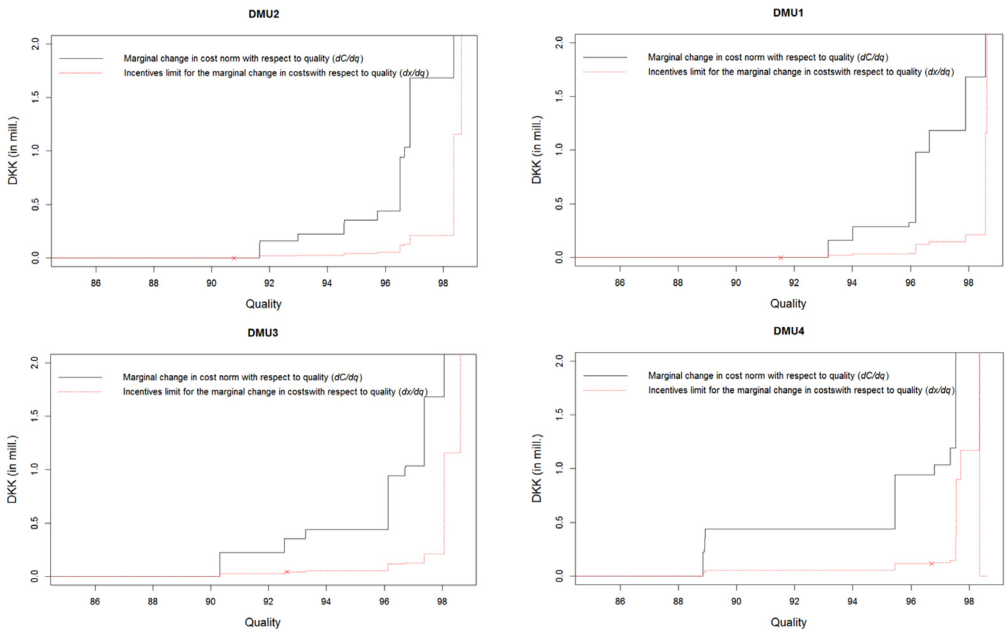

Fig. 8. Marginal changes in cost norms and marginal quality thresholds for four inefficient firms.

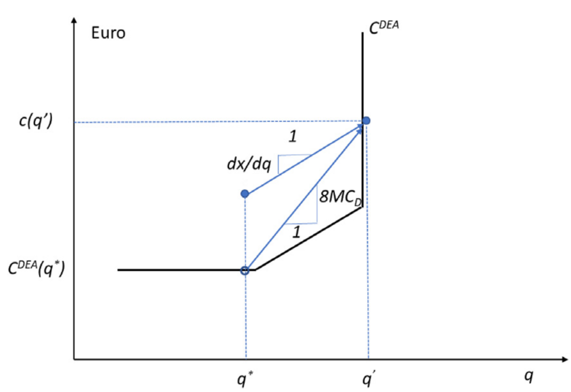

Fig. 9. Gains from a discrete quality increase.

that the best practice cost of the other firms is infinite, as in Fig. 5. The advantage of the latter approach is that the best practice cost will now be the costs of the firm in question. Hence, the firm may obtain a discrete jump in the cost norm large enough to justify its extra costs of quality.

To illustrate this idea, we have calculated the increase in quality, say Δq, necessary for each firm to become fully efficient, even though the cost of quality, like the best practice costs, increases up until the last non-Archimedean infinitesimal quality addition. Let us denote the cost level for the firm at this increased quality level q' as $c(q')$. The corresponding increase in the revenue cap is now

$$\Delta R = \frac{1}{8}\left(c\left(q'\right) - \hat{C}(q^*)\right) \tag{18}$$

and is generated by a change in the quality of $\Delta q = q' - q^*$. We can therefore say that if it is possible to increase quality from q^* to q' at an average marginal cost less than

$$MC_D = \frac{\Delta R}{\Delta q} \tag{19}$$

then it is attractive to make this discrete jump in quality. The idea is illustrated in Fig. 9.

We calculated these values for the different Danish waterworks and illustrated the results in Fig. 10.

The red dot for a given firm shows the thresholds for the firm's marginal costs below which it is worthwhile to make marginal increases in quality. This value corresponds to the red dashed line at the red crosses in Figs. 7 and 8. The black dots show the discrete marginal cost thresholds, MC_D. Suppose a firm can maintain an average marginal cost below MC_D while increasing quality to make the firm efficient. In that case, the firm will have incentives to make this discrete increase in quality. The size of this discrete jump in quality is illustrated with the blue cross.

The figure shows that low-quality firms should not be willing to pay much to increase quality. This is no surprise because they are most likely benchmarked against a non-fully dimensional efficient facet, making marginal gains zero. In addition, their quality should increase considerably for their costs to become efficient, i.e., Δq in the MC_D formula is large.

For some firms, the distance between the black and red dots is high. This occurs when the firm is currently being benchmarked on a facet with low $\frac{\partial \hat{C}}{\partial q}$ but where a relatively small increase in quality Δq suffices to make it efficient. In other words, a small marginal change in quality will only increase the cost norm slightly, but a slightly bigger discrete improvement in quality will change the cost norm considerably. Another reason could be that such a firm is highly inefficient with high costs. If such a firm can become efficient by increasing quality, it will likely have incentives to do this, as it will let the cost norm be equal to its high costs.

5. Dynamic incentives

Thus far, we have focused on firms' myopic, single-period incentives to reduce costs and expand quality. We will now consider the dynamic incentives. What are the incentives to reduce costs and expand quality in a given period when we take into account the effects on the revenue cap in later periods?

For an inefficient firm, the myopic and dynamic incentives are the same since the firm does not influence cost norms and, hereby, the development of the revenue cap.

For an efficient firm, things are more complex. If an efficient firm improves its performance in one period, it will affect the future revenue caps downwards. Therefore, the firm faces a so-called Ratchet effect - by improving in one period, it makes its own future harsher, which lowers the incentives to improve in the first place.

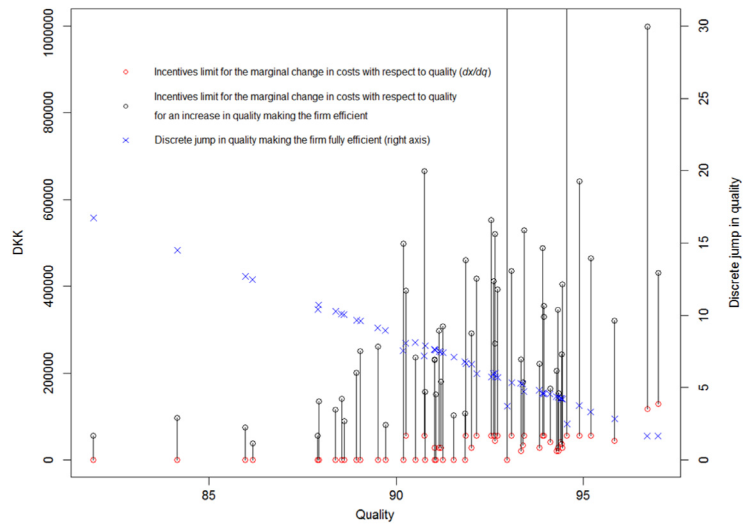

Fig. 10. Thresholds on marginal quality costs making marginal and discrete quality changes attractive.

To analyse the dynamic effects, we can ignore the general productivity requirements since they affect the revenue cap irrespective of the firms' behaviour. We can therefore focus on the key updating rule

$$R_t = \alpha R_{t-1} + \beta \hat{C}_t \tag{20}$$

wherein the case of Danish water, we have $\alpha = \frac{7}{8}$ and $\beta = \frac{1}{8}$.

Assume, first, that an efficient firm lowers its cost with Δx in period 1 and hereafter returns to its original cost level. We will now investigate how this affects the revenue cap. In period 1, the revenue cap is reduced with $\beta\Delta x$. In the next period, the reduction is $\alpha\beta\Delta x$; in the third year, it is $\alpha^2\beta\Delta x$; in the fourth year, it is $\alpha^3\beta\Delta x$; etc. In summary, the reduction in the sum of future revenue caps is

$$\beta\Delta x + \alpha\beta\Delta x + \alpha^2\beta\Delta x + \alpha^3\beta\Delta x + \cdots = \beta\Delta x\frac{1}{1-\alpha} \tag{21}$$

as long as $\alpha < 1$. If we discount future gains and losses with $\gamma \leq 1$, we obtain a loss in future revenue cap values of

$$\beta\Delta x\frac{1}{1-\gamma\alpha} \tag{22}$$

The gain to the firm was a cost reduction of Δx in period 1, and therefore, the total discounted profit effects of the one-period cost reduction are

$$(1-\beta\frac{1}{1-\gamma\alpha})\Delta x \tag{23}$$

Note that when $\beta = 1-\alpha$, as in the Danish water model, and we assume $\gamma = 1$, the accumulated loss over an infinite time horizon is precisely equal to the gain in accumulated costs, Δx. *Hence, when there is no discounting, the firm does not have strict incentives to reduce costs in a single period. When the discount factor is less than one, $\gamma < 1$, then the firm has strict incentives to reduce costs, although the incentives are much less than in the myopic model* since it takes 'forever' to approach the old revenue cap.[18]

If, instead of a one-period reduction in cost, the firm introduces a cost reduction from period 1 onwards, we obtain the same conclusions. Hence, the incentives to reduce in period 2 (following a period 1 reduction) are the same as the incentives to reduce in period 1. The reason is that the period 2 cost reduction will trigger a similar reduction in the cost norm followed by a gradual increase in the cost norm towards the starting point.

In summary, we see that inefficient firms have the same incentives in a multiperiod model as in a single-period model — as long as they stay inefficient. Efficient firms, however, have much weaker incentives to reduce costs and may have no strict incentives to do so if the discount factor is 1.

Now, consider the quality incentives.

As in the case of pure cost reductions, for an inefficient firm, the myopic and dynamic quality incentives are the same since the firm influences neither the cost norm nor the revenue cap in later periods.

Consider now an efficient firm that increases its quality with Δq in the first period and hereafter returns to its original quality. The corresponding increase in costs along the efficient frontier, Δx, will only be recovered with the β factor the first year. Later, however, further gains will materialize since the revenue cap will gradually grow. More precisely, in the first year, the revenue cap increases with $\beta\Delta x$, and in subsequent periods, it increases with $\alpha\beta\Delta x$, $\alpha^2\beta\Delta x$, etc. corresponding to an aggregated gain in revenue cap of $\beta\Delta x\frac{1}{1-\gamma\alpha}$ when the discount factor is ρ. Hence, the extra cost of increasing quality is barely recouped over the coming years, assuming that the future gains are not discounted (γ). In other words, even with long-run gains, the quality incentives are very weak.

Reducing quality, however, is more attractive. Truly, it comes not only with a cost reduction of Δx in the first period but also with an accumulated loss in revenue caps of $\beta\Delta x\frac{1}{1-\gamma\alpha}$. Hence, in the very long run, and when there is no discounting, the gains and losses cancel out.

In summary, the very negative (positive) incentives towards quality improvements (reductions) from our single-period analysis are softened when a dynamic perspective is introduced.

6. Limitations and extensions

Our analyses above come with several limitations and relevant extensions. Before concluding the paper, let us point to a few limitations and interesting extensions.

We have examined the incentives to increase quality in a model where quality is either entirely absent from the benchmarking model or is included directly in the benchmarking model. It is important to mention that several other approaches can be used. The most common approach is to have a strict minimum requirement on the quality or to *incentivize quality via an add-on payment/penalty to the cost-focused revenue cap scheme*. We discussed some alternatives in Section 2.

[18] As mentioned earlier, in our analysis, we have disregarded the 'no-revenue-cap-increase' restriction, which is also part of the Danish water regulation. In that case, the revenue cap reductions will not increase, i.e., the efficient firm will experience a reduction in revenue cap of $\beta\Delta x$ in each period. It follows that a one-shot cost reduction is not attractive unless $\Delta x \geq \beta\Delta x + \rho\beta\Delta x + \rho^2\beta\Delta x \ldots$, i.e., a one-shot reduction is only attractive if $\frac{\beta}{1-\rho} \leq 1$, which is equivalent to $\rho \leq 1-\beta$. With $\beta = \frac{1}{8}$, this corresponds to an interest rate of at least $\frac{1}{7} \approx 14.3\%$.

There are, however, further alternatives that are more akin to the inclusion of quality in the benchmarking model. One possibility is to adjust for quality in a *second-stage analysis* or to perform a quality adjustment of the traditional volume-based cost drivers such as the net-volume measures in the Danish water regulation. Again, it requires more specific assumptions to examine incentives in such cases, but it seems a worthwhile topic for future research.

Another limitation of our analysis is that we have assumed all firms to have *full information*. This is not the case in reality. The benchmarking results differ considerably from year to year; therefore, firms cannot determine the exact consequences of changing their level of quality. In our analysis, for example, it will be extremely difficult to find the exact level of quality that will make them efficient. If they try to exploit the possibility of obtaining a discrete "jump" in cost norms where they go from being inefficient to efficient, as discussed in Section 4.2, they risk losing if they do not become efficient.

A few other extensions are also worth mentioning.

First, from the point of view of incentive theory, we have used a relatively naïve model in the sense that we did not have any cost of effort associated with cost reduction. One easy way to remedy this is to *assume that firms like slack*, i.e., excess spending of resources beyond what is strictly necessary to produce the services. Specifically, we might assume that a firm producing outputs y by spending costs x does so by adding slack s to the underlying true minimal cost $C(y)$,

$$x = C(y) + s \tag{24}$$

and to assume that they are not just interested in maximizing profits but also in benefitting from slack, i.e., by assuming that the firm's objective is to maximize firm utility

$$U = \Pi + \rho \cdot s \tag{25}$$

where the marginal value of slack compared to profit is $\rho < 1$. A similar approach has been used in several other papers on regulations, cf., e.g., Agrell et al. [4] and Bogetoft [9,12,13]. *It is intuitively clear that the gains from slack dampen the incentives to reduce costs and thereby also dampen the incentives to reduce quality since these incentives are derived mainly by cost reductions.* To see the first effect, it is easy to see that the marginal gains from changing costs x are now

$$\frac{\partial U}{\partial x} = \frac{\partial \Pi}{\partial x} + \rho = \frac{1}{P} \cdot \frac{\partial \hat{C}}{\partial x} - 1 + \rho \tag{26}$$

The firm, therefore, has incentives to reduce costs as long as

$$\frac{\partial \hat{C}}{\partial x} < P(1-\rho) \tag{27}$$

i.e., less often than previously where the condition was $\frac{\partial \hat{C}}{\partial x} < P$.

Another interesting extension would be to work with *super efficiencies* instead of efficiencies. In our analysis, we found that efficient firms had weaker incentives in general since a cost reduction leads to a decline in revenue cap. If a firm is only compared to best practices among the other firms, then there is no ratchet effect, i.e., a firm is not penalized in later periods from doing well in a given period. Again, the advantage of superefficiency from an incentive perspective has been emphasized in several of the papers cited above.

A final extension worth considering is benchmarking actual costs rather than the revenue cap. The Danish water regulation in each period finds the required savings by benchmarking the revenue cap of the last period again the cost norm,

$$x^{spDW} = \frac{\left(\frac{R-\hat{C}}{R}\right)}{P} \tag{28}$$

In more traditional benchmarking-based revenue cap schemes, the required savings are determined by comparing the actual costs x with the cost norm

$$x^{spTR} = \frac{\left(\frac{x-\hat{C}}{x}\right)}{P} \tag{29}$$

The impact of using the more traditional approach x^{spTR} depends on the revenue cap R. If the revenue cap is above the actual costs, x, the Danish water model requires a larger reduction x^{spDW} of the revenue cap than the traditional approach. If the revenue cap is below the actual costs, x, the Danish water model requires a smaller reduction herein than the traditional approach. It seems, therefore, that the special Danish water variant puts extra cost reduction incentives on very profitable firms while lowering the pressure on firms that are already struggling with a negative profit.

7. Conclusion

In this paper, we have examined how benchmarking-based revenue cap regulations fare when some of the cost drivers are discretionary, i.e. controllable by the firms. As an example, we considered the inclusion of a discretionary quality parameter in the benchmarking model and showed that the incentives to choose socially optimal quality levels are very limited.

At the abstract level, the challenge is that the benchmarking-based revenue cap is a convex function. Hence, a regulated firm faces a convex cost function and a convex revenue function, which does not generally lead to a unique optimal solution. Another general problem is that a regulated firm may choose quality strategically by searching for levels of quality where the benchmarking is more lenient.

To obtain more insight, it is useful to analytically and numerically examine a more specific regulation. We, therefore, examined the incentives in the Danish regulatory framework for water firms. As with any other regulation, this regulation has particular features, and we showed how these features affect the incentives.

As in the case of most regulations, the Danish water regulation allows for a catch-up period. We showed that the catch-up period provides strong incentives to reduce costs since firms can keep possible cost reductions for several years before the cost norm fully internalizes the cost reduction potentials. On the other hand, it also gives very weak quality incentives since it takes eight years before the extra cost of increasing quality is fully internalized in the cost norm.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Agrell, P. Bogetoft, Regulatory benchmarking: Models, analyses and applications, Data Envel. Anal. J. (2017) 49–91.
[2] P. Agrell, P. Bogetoft, Theory, techniques and applications of regulatory benchmarking, in: E. Grifell-Tatjé, C.K. Lovell, R.C. Sickles (Eds.), The Oxford Handbook of Productivity Analysis, Oxford University Press, 2018, pp. 523–555.
[3] P. Agrell, P. Bogetoft, J. Tind, Incentive plans for productive efficiency, innovation and learning, Int. J. Prod. Econ. (2002) 1–11.
[4] P.J. Agrell, P. Bogetoft, J. Tind, DEA and dynamic yardstick competition in scandinavian electricity distribution, J. Prod. Anal. 23 (2005) 173–201.
[5] P. Bogetoft, Performance Benchmarking: Measuring and Managing Performance, Springer, New York, 2012.
[6] P. Bogetoft, L. Otto, Benchmarking with DEA, SFA, and R, Springer, 2011.
[7] X. Dai, T. Kuosmanen, Best-practice benchmarking using clustering methods: Application to energy regulation, Omega 42 (1) (2014) 179–188.
[8] R. Ramanathan, U. Ramanathan, Y. Bentley, The debate on flexibility of environmental regulations, innovation capabilities and financial performance – a novel use of DEA, Omega 75 (2018) 131–138.
[9] P. Bogetoft, Incentive efficient production frontiers: An agency perspective on DEA, Manage. Sci. 95 (1994a) 9–968.
[10] P. Bogetoft, Non-Cooperative Planning Theory, Springer-Verlag, Berlin, 1994b.
[11] P. Bogetoft, Incentives and productivity measurements, Int. J. Prod. Econ. 6 (1995) 7–81.
[12] P. Bogetoft, DEA-based yardstick competition: The optimality of best practice regulation, Ann. Oper. Res. 27 (1997) 7–298.
[13] P. Bogetoft, DEA and activity planning under asymmetric information, J. Prod. Anal. (2000) 7–48.

[14] NVE, NVE, 2019, Retrieved from https://www.nve.no/reguleringsmyndigheten/okonomisk-regulering-av-nettse-lskap/om-den-okonomiske-reguleringen/kile-kvalitetsjusterte-inntektsrammer-ved-ikke-le-vert-energi/.
[15] P. Bogetoft, A. Eskesen, Balancing Rent Extraction and Service Differentiation, Working Paper, CBS, 2021.
[16] R. Antle, P. Bogetoft, Mix stickiness under asymmetric cost information, Manage. Sci. 65 (6) (2019) 2787–2812.
[17] E. Heesche, M. Asmild, Incorporating Quality in Economic Regulatory Benchmarking, Omega 110 (2020) 102630.
[18] P. Milgrom, J. Roberts, Economics, Organization, and Management, Prentice-Hall, Englewood Cliffs, 1992.
[19] R.G. Dyson, R. Allen, A.S. Camanho, V.V. Podinovski, C.S. Sarrico, E.A. Shale, Pitfalls and protocols in DEA, European J. Oper. Res. 24 (2001) 5–259.
[20] O.B. Olesen, N.C. Petersen, Indicators of ill-conditioned data sets and model misspecification in data envelopment analysis: An extended facet approach, Manage. Sci. 15 (1996) 7–306.
[21] P.W. Wilson, L. Simar, Statistical inference in nonparametric frontier models: The state of the art, J. Prod. Anal. 13 (2000) 49–78.