

Louvain School of Management

Catch-up speed in network regulation a decomposition of cost efficiency

Research Master's Thesis submitted by
Nathan Grognat

With the view of getting the degree in
Master en ingénieur de gestion, à finalité spécialisée

Supervisor
Prof. Per Joakim Agrell

Academic Year 2016-2017

UNIVERSITE CATHOLIQUE DE LOUVAIN

Abstract

Louvain School of Management

Master of Science in Business Engineering

Catch-up speed in network regulation: a decomposition of cost efficiency

by Nathan Grognet

Practitioners of network regulation are extensively using benchmarking measures in defining their incentive regulatory schemes and—more specifically—the speed at which those schemes require inefficient network utilities to catch-up to efficiency through cost minimisation. This catch-up speed is usually set based on the incumbent inefficiency of operators and the period of time they are granted to remove it. However, its computation lacks research and foundation as the allowed period to reach target performance is not firm-specific and often set based on negotiated speculations on potential productivity improvement. We propose an approach to set a firm-specific catch-up period based on historical cost efficiency change rates and the decomposition of an operator's cost efficiency into technical and allocative efficiency. We demonstrate that this decomposition can be used to assess the source of a firm's excess costs and thus the strategies that need to be implemented to eliminate those costs. An econometric analysis is used on a panel of 143 Swedish electricity distributors to validate the decomposition of cost efficiency and a model is proposed to compute the catch-up period according to this decomposition.

Keywords: incentive regulation, electricity networks, efficiency analysis, catch-up speed, revenue cap.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Per Joakim Agrell who provided me with insightful feedback and comments throughout the making of this Master's thesis. His patience and expertise helped me overcome the toughest obstacles. Furthermore I would like to thank my loved ones for their love and morale support. I will be forever grateful to all of you. Thank you so much.

Contents

Abstract	ii
Acknowledgements	iii
List of Figures	vii
List of Tables	ix
Abbreviations	xi
1 Introduction	1
2 Methodology	3
2.1 Regulation of network utilities	3
2.1.1 Regulatory mechanisms	5
2.1.2 Incentive regulation	6
2.2 Benchmarking	8
2.2.1 The technology set T	8
2.2.2 Efficiency measures	11
2.2.2.1 Farrell efficiency	11
2.2.2.2 Cost efficiency	12
2.2.2.3 Dynamic efficiency	14
2.2.3 Frontier models	16
2.2.4 Data Envelopment Analysis	17
2.2.4.1 Models for technical and cost efficiency	17
2.2.4.2 Non-discretionary variables	18
2.2.5 Benchmarking in regulation	20
2.3 Frontier-based revenue cap	22
2.3.1 X-factor	23
2.4 Practical cases	25
2.4.1 Austria - E-Control	26
2.4.2 Germany - Bundesnetzagentur	26
2.4.3 Norway - Norwegian Water Resources and Energy Directorate . .	27
2.4.4 Spain - Comisión Nacional de Energía	27
2.4.5 Sweden - Energy Market Inspectorate	28

2.5	Summary and review	29
3	Model	31
3.1	Decomposition of cost efficiency	31
3.1.1	Technical efficiency	32
3.1.2	Allocative efficiency	33
3.1.3	Assessment of efficiency profile	35
3.2	Complexity factor ξ_i	37
3.3	Firm-specific catch-up period T_i	40
3.3.1	Default catch-up period T_0	41
3.3.2	Sensitivity parameter α	43
3.4	Discussion	44
4	Data	47
4.1	Dataset	47
4.2	Model variables	48
4.2.1	Outputs	48
4.2.2	Inputs	48
4.2.3	Non-discretionary variables	49
5	Analysis	53
5.1	Significance of non-discretionary variables	53
5.2	Benchmarking results	54
5.2.1	Efficiency scores	54
5.2.2	Efficiency profile η	57
5.2.3	Efficiency changes	58
5.3	Model application	60
5.3.1	Hypothesis	60
5.3.2	Complexity factor ξ	63
5.3.3	Simulation	65
5.4	Summary of the results	68
6	Conclusion	69
A	Additional figures	71
A.1	Non-discretionary variables	71
A.2	Regression diagnostics	72
B	CD/USB content	75
	Bibliography	77

List of Figures

2.1	Technology set T under constant (top left), variable (top right), and decreasing returns to scale (bottom) assuming a DEA-based frontier (Bogetoft and Otto (2010))	10
2.2	Graphical representation of input Farrell efficiency E in a two-inputs setting (Bogetoft and Otto (2010))	11
2.3	Graphical representation of the different efficiency measures (Bogetoft and Otto (2010))	12
2.4	Ex ante regulation (Agrell and Bogetoft (2013))	22
2.5	Elements of the EI regulation (Schweinsberg et al. (2011))	28
3.1	Complexity ξ vs. efficiency profile η (cost efficiency of 60%)	39
5.1	Density of efficiency scores (2000-2005)	56
5.2	Efficiency scores by year (2000-2005)	56
5.3	Cost efficiency by year without and with z-variables	57
5.4	Efficiency profile η	58
5.5	Histogram for cost efficiency change	59
5.6	Efficiency profile η and complexity ξ	64
5.7	Density plot and histogram of complexity ξ	65
5.8	Catch-up requirement χ_i vs. Cost efficiency	68
A.1	Cost efficiency vs. Line proportion	71
A.2	Cost efficiency vs. High voltage proportion	72
A.3	Density plot of the standardised residuals of model (5.7)	72
A.4	Q-Q plot for the standardised residuals of model (5.7)	73
A.5	Standardised residuals vs. Fitted values of model (5.7)	73

List of Tables

2.1	Fixed-price contract vs. Cost-of-service contract	6
2.2	Some European regulation regimes (Agrell and Bogetoft (2013))	8
2.3	Returns to scale assumptions	10
2.4	Technical vs. allocative efficiency	14
2.5	Frontier models (Bogetoft and Otto (2010))	16
3.1	Example - Efficiency profile η_i	36
3.2	Example - Efficiency profile η_i and proportions ζ_{TE_i} and ζ_{AE_i}	38
3.3	Example - Complexity ξ_i , catch-up period T_i , and catch-up requirement χ_i	44
4.1	Weighted network length OLS results	49
4.2	Descriptive statistics of the model variables	51
5.1	Efficiency scores without z-variables (2000-2005)	53
5.2	Analysis of the z-variables significance	54
5.3	Efficiency scores with z-variables (2000-2005)	55
5.4	Mean efficiency scores by year	55
5.5	Example - Efficiency profile η of operators SEVAB Nät AB and Ystad Energi AB (2004)	57
5.6	Efficiency changes (2000-2005)	58
5.7	Mean efficiency change by year	59
5.8	Summary of the regression model from (5.7)	63
5.9	Proportions ζ_{TE} and ζ_{AE} , and complexity ξ	64
5.10	Results (efficiency scores, ζ_{TE_i} , ζ_{AE_i} , and ξ_i) (2006)	66
5.11	Catch-up (T_i , χ_i)	67

Abbreviations

AE	A llocative E fficiency
AEC	A llocative E fficiency C hange
CAPEX	C apital E xpenditure
CE	C ost E fficiency
CEC	C ost E fficiency C hange
CM	C ost-Malmquist Index
CPI	C onsumer P rice I ndex
CRS	C onstant R eturns to S cale
DEA	D ata E nvelopment A nalysis
DSO	D istribution S ystem O perator
DRS	D ecreasing R eturns to S cale
DMU	D ecision-Making U nit
EC	E fficiency C hange
ES	E fficiency S core
MI	M almquist Index
NPAM	N etwork P erformance A ssessment M odel
OPEX	O perating E xpenditure
TC	T echnological C hange
TE	T echnical E fficiency
TEC	T echnical E fficiency C hange
RPI	R etail P rice I ndex
RTS	R eturns T o S cale
SFA	S tochastic F rontier A nalysis
TFP	T otal-Factor P roductivity
TOTEX	T otal E xpenditure
VIF	V ariance I nflation F actor
VRS	V ariable R eturns to S cale

Chapter 1

Introduction

Due to the high fixed costs that network infrastructure requires, distribution system operators—or DSOs—manifest themselves as natural monopolies. Even though those firms are operating legally they pose a risk for social welfare as they can abuse their market position to gain excess profit at the expense of consumers. Because of this risk the state has to designate a regulatory authority whose purpose is to represent society and offer the regulated utilities a compensation scheme that will negate the risk of monopoly pricing. Since network operators set their tariffs based on their allowed revenue, by controlling the compensation they are granted the state can determine the tariffs consumers will be facing.

In Europe, most regulators have adopted incentive-based regulation schemes such as fixed-price contract which goal is to align the profit-maximising objectives of network operators with those of the market (i.e. optimal quality at minimal cost). The most widely used fixed-price scheme is revenue cap regulation which sets the revenue a DSO is allowed to receive each year. The cap is corrected yearly for inflation and an expected productivity improvement X —or X -factor—which purpose is to promote cost minimisation by gradually decreasing the allowed revenue. The X -factor is at the heart of incentive regulation as it is this factor which provides regulated entities with the proper incentives for efficient processes and operations.

Because the regulator and DSO have asymmetric information on the DSO's true cost function, estimating the revenue cap—or compensation—is a non-trivial exercise requiring extensive benchmarking analyses. In particular, the regulator has to impose a firm-specific required improvement X accounting for both the expected technical improvement of the industry as a whole (i.e. technological progress) and the inefficiency reduction that he expects inefficient firms to demonstrate each year (i.e. catch-up to efficiency). This second component represents the inefficiency (i.e. inefficient costs) that

the regulator wants to reclaim each year in the name of society. Due to the capital intensity of network operators the regulator can not require them to catch-up their incumbent inefficiency overnight as this could halt the energy distribution altogether. Regulated operators are thus given a number of years—or regulatory periods—to completely eliminate their inefficiency.

The motivation of this thesis stems from the lack of methodology regarding the computation of the catch-up speed in network regulation. To determine the speed at which network operators will be required to improve, authorities usually decide upon a certain number of years—or catch-up period—that DSOs will be granted to reach efficiency. Nevertheless the period over which inefficient operators must reach efficiency lacks foundation and a considerable element of negotiation between the regulator and the firms usually influence its decision.

The primary limitation of the approach currently used by regulators originates from the fact that the catch-up period of their regulatory schemes is not firm-specific. The catch-up requirement is usually computed based on the firm's inefficiency and the number of years firms have to reach optimal performance.

The objective of this thesis is to provide a review of the tools and methods currently used in setting the catch-up requirement and to propose an approach and methodology to assist the regulators in determining the catch-up that their fixed-price contracts should require from the regulated DSOs. Our research question is the following.

How can regulatory authorities make use of the available benchmarking tools and measures to set the required firm-specific catch-up speed in the fixed-price regulation of network utilities?

Even though our work can contribute to other fields than network regulation it is important to state that we are here focusing on electricity distribution.

The outline of the thesis is as follows. In Chapter 2 we present the fundamentals of the incentive regulation of network utilities and the benchmarking measures commonly used in regulation. In Chapter 3 we propose an approach and model to set the catch-up requirement based on the decomposition of cost efficiency into technical and allocative efficiency. Chapter 4 presents the empirical data that will be used in our analysis. In Chapter 5 we perform the regulator's exercise of efficiency assessment before applying our model to the data and empirically demonstrating the hypothesis upon which it relies. We conclude the research in Chapter 6.

Chapter 2

Methodology

In this chapter we first describe the need for regulation of network operators. The regulatory tools currently in use are briefly introduced and defined with a focus on incentive regulation. The use of benchmarking approaches in setting the improvement requirement X in the context of the frontier-based revenue cap regulation is presented and reviewed. Several European applications of incentive regulation are also reviewed.

2.1 Regulation of network utilities

Network companies such as electricity distributors are characterised by high fixed costs of network capital and low marginal operating costs. Because of high fixed and start-up costs for infrastructure and operations, distributor system operators manifest themselves as natural monopolies. For each geographic region, distribution cost is minimised by having a unique distributor supplying all the consumers. This leads to high barriers to entry for potential new rivals in electricity distribution. Natural monopolies pose a risk for social welfare as they can abuse their market position to gain excess profit at the expense of consumers (Carrington et al. (2002)). As stated in Agrell and Bogetoft (2013), the electricity distribution sector is thus subject to the risk of monopoly pricing, where a firm can charge to consumers a price much higher than the one they would have been charged with in a competitive market, while providing poorer services. Because monopolies are not facing any competition they do not have the proper incentives to reduce costs. There is also not any real substitute to the electricity and services provided by the DSOs, which means that the demand is quite inelastic¹.

Even though those natural monopolies are provided with licenses to operate as legal monopolies, they represent a risk for social welfare (Agrell and Bogetoft (2013)). This is

¹An increase or decrease in price does not have a significant impact on demand.

why DSOs—whether they are state- or privately-owned—need to be regulated in order to ensure that the energy is delivered at a fair price and good quality for consumers. Failing to do so represents a risk of market failure. The government thus needs to appoint a regulatory authority that will purchase the services of the distributors in the name of all the consumers. The act of network regulation is “*the prerogative by law or decree to vest the authority to define, apply, monitor and enforce the technical and economic conditions and rules for a set of operators*”. (Agrell (2015))

“*The economic regulation of electricity networks attempts to achieve two major goals. On the one hand, it is important to stop a network business from exploiting its natural monopoly position by setting prices well in excess of efficient costs. On the other hand, retaining strong incentives for network businesses to pursue profits is important for driving efficiency in production and management decisions.*” (Commission et al. (2013))

Indeed, besides removing the risk of monopoly pricing the goal of regulation is also to provide the regulated entities with the proper incentives for efficient costs. Theory and empirical evidence suggest that weak competition leads to poor internal managerial incentives for cost minimisation, leading to X-inefficiency (Commission et al. (2013)). The concept of X-inefficiency was first introduced in Leibenstein (1966) and represents the difference between the assumed efficiency of firms under perfect competition and their actual efficiency under imperfect competition. This business under-performance results from poor efficiency incentives.

“*The desire to create a competitive environment is now prevailing in infrastructure industries, and where competition is limited the search for efficiency gains is at the core of the regulation debate.*” (Coelli (2003))

The main challenge regulators are faced with when trying to enforce cost minimisation is the information advantage of regulated entities. Network companies possess exclusive information and details on their cost function and opportunities, leading to *information asymmetry*. Regulated firms have a tendency to use this exclusive information to their advantage to increase their profit to the disadvantage of social welfare. The regulator has imperfect information on the DSO’s true underlying costs and improvement potential and can only observe the realised costs in hindsight.

The fact that regulatory authorities are not completely informed about the specifics and attributes of a regulated firm’s costs leads to two main issues (Armstrong and Sappington (2007)). The first one is *moral hazard*, which is an issue arising when “*individuals engage in risk sharing under conditions such that their privately taken actions affect the probability distribution of the outcome*” (Hölmstrom (1979)). Because of asymmetric

information between the DSO and the regulator the individual actions of the regulated entity can not be observed and therefore contracted upon. In the presence of high realised costs it is not clear for the regulatory authority whether or not those are due to poor managerial effort in the firm's operations. The second issue arising from information asymmetry is *adverse selection* which refers to a situation where the service provider (i.e. the distribution system operator) has more information on the true value of its services than the consumers—represented by the regulator (Jullien (2000)). In the case of electricity distribution this can lead to DSOs providing electricity at prices that are excessively high relative to the actual costs of distributing it.

2.1.1 Regulatory mechanisms

Following Laffont and Tirole (1993), two extreme regulatory mechanisms can be implemented for electricity distributors (Joskow (2008)).

The first one is an ex ante² fixed-price contract³ that determines the price that the regulated firm will be permitted to charge its consumers. This price is fixed throughout a predetermined period of time and adjusts for changes in input price indexes (i.e. inflation) and other chosen exogenous factors. The formula of the fixed-price regulation for operator k can be expressed as follows (Joskow (2008)).

$$P^k(t) = P^k(0)(1 + RPI - X) \quad (2.1)$$

where P^k is the allowed price, RPI is the rate of input price accounting for changes in inflation, and X is a target productivity change factor.

Since the price is fixed ex ante for the whole regulatory period, the price cap regulatory mechanism does not respond to changes in efficiency (i.e. managerial effort). This promotes cost reduction by allowing the regulated firms to keep the gains from potential managerial improvement (Liston (1993)). The main drawback of such a scheme is the fact that the regulator is uncertain about the true cost opportunities the firms are actually faced with. The fixed price therefore needs to be high enough for the distributors to cover their realised costs. If firms are incurring high costs despite being efficient, a fixed price set too low may prevent them to break even and thus jeopardise the distribution of electricity altogether. Because of this the fixed price tends to be biased upward leading to poor informational rent extraction from distributors to consumers.

²“A priori.”

³Suggested initially in Littlechild (1983).

The second one is a pure ex post⁴ cost-recovery contract that compensates the regulated firm for all the costs it has actually incurred through production. Following Agrell and Bogetoft (2013) the reimbursement in a given period t for distributor k can be set through the following formula.

$$R^k(t) = C_{Opex}^k(t) + D^k(t) + (r + \delta)K^k(t) \quad (2.2)$$

where R^k is the allowed revenue, C_{Opex}^k the operating expenditure, D^k the capital depreciation, r the interest rate of capital, δ a mark-up, and K^k the capital.

Unlike the fixed-price contract, this mechanism fully extracts informational rent from the firms for the benefit of society. Distributors are breaking even and can not accrue excess profits through efficiency gains. Even though rents are completely removed there is not any incentive for efficiency improvement under this mechanism. Since they know their realised costs will be fully reimbursed by the regulator, DSOs do not have any interest in exerting optimal managerial effort leading to inefficiently high costs. Despite the fact that distributors are prevented from extracting profit from the market, this managerial slack—or X-inefficiency—leads to higher prices for consumers. Those schemes also provide strong incentives for over-investment in capital. (Agrell and Bogetoft (2013))

TABLE 2.1: Fixed-price contract vs. Cost-of-service contract

Contract type	Advantage	Drawback
Fixed-price	Incentives for efficiency	Poor informational rent extraction
Cost-recovery	Alignment of prices and costs	No incentive for cost reduction

Table 2.1 summarises the main advantage and drawback of those two regulatory mechanisms. Those are two extremes schemes and most regulatory schemes will lie somewhere in-between them.

2.1.2 Incentive regulation

The incentive regulation of network firms is designed to align the individual profit-maximising objectives of distribution system operators with the goals of public policy: optimal pricing and cost in electricity supply (Hackett (2010)). Regulatory schemes aiming at achieving this are called “*incentive-based regulations*”. Incentive regulation can be defined as the “*implementation of rules that encourage a regulated firm to achieve desired goals by granting some, but not complete, discretion to the firm*” (Uri (2001)).

⁴“In hindsight.”

Following this definition and looking at the two polar mechanisms presented in Section 2.1.1 we can see that they do not have the same incentive power.

Cost-recovery regimes such as cost-of-service, cost-plus, or rate-of-return regulations⁵ do not exhibit the proper incentives and result in poor performance. Those are called *low-powered* regimes since they demonstrate weak incentive power. Fixed-price regimes like price cap, revenue cap, or CPI-X regulations are *high-powered* regimes as they allow the regulated firms to retain realised efficiency gains. In those regimes to maximise profit DSOs need to minimise costs which is the key to incentive provision. (Agrell and Bogetoft (2013))

Another interesting high-powered regulatory scheme worth mentioning is the *yardstick* regime. This scheme is aimed at mimicking the incentives of a real competitive market. The cost function used to determine the allowed revenue of a firm is updated in each period with real observations from other firms. Instead of relying on exogenous parameters to predict cost developments (i.e. parameters RPI and X in (2.1)), this scheme uses real yearly data. Each firm is therefore in pseudo-competition with other firms operating under similar conditions. A yardstick scheme formula based on average cost⁶ can be defined as follows (Agrell and Bogetoft (2013)).

Assuming a set of firms $h = 1, \dots, k-1, k+1, \dots, K$ the allowed revenue of firm k is defined as

$$R^k(t) = \frac{1}{K-1} \sum_{h \neq k} C^h(t) \quad (2.3)$$

where K is the number of firms, and C^h the realised cost of firm h .

Two difficulties arise from this scheme concerning its implementation. For this mechanism to be fair the K distributors of the set need to be comparable in terms of contextual constraints and operating conditions. Also if the set of K firms is too small there may be a risk of collusion. (Agrell and Bogetoft (2013))

⁵More details on those regulatory schemes can be found in Agrell and Bogetoft (2013) and Agrell (2015).

⁶Yardstick scheme can also rely on best practice (optimal cost) or take other forms.

TABLE 2.2: Some European regulation regimes (Agrell and Bogetoft (2013))

Country	Regulatory scheme
Austria	Revenue cap
Belgium	Revenue cap
Switzerland	Cost recovery
Germany	Revenue cap
Denmark	Revenue cap
Spain	Revenue cap
Finland	Revenue cap
France	Cost recovery
Great Britain	Revenue cap
Greece	Cost recovery
Hungary	Price cap
Ireland	Price cap
Netherlands	Yardstick competition
Norway	Yardstick competition
Sweden	Revenue cap

Table 2.2 shows us the regulatory schemes used in several European countries. We can see that most of these countries (i.e. 10 out of 15) are using fixed-price mechanisms with a preference for revenue cap⁷ over price cap regulation.

2.2 Benchmarking

“Relative performance evaluations or—using modern terminology—benchmarking is the systematic comparison of the performance of one firm against other firms.” (Bogetoft and Otto (2010))

Benchmarking allows us to assess and compare the relative performance—or efficiency—of a firm against the rest of the industry. While it can be used to make intra-organisational comparisons (i.e. comparisons of sub-units from a same company or organisation), benchmarking is most often used for inter-organisational comparisons (i.e. comparisons of firms from a same industry).

2.2.1 The technology set T

Efficiency can be defined as the use of the fewest resources (i.e. inputs) to generate to most products or services (i.e. outputs). In order to properly define efficiency we first

⁷Revenue cap is developed in more details in Section 2.3.

need to define the concept of *technology*. Assuming a firm i is producing a vector of m outputs $y_i \in \mathbb{R}_+^m$ with a vector of n inputs $x_i \in \mathbb{R}_+^n$ the technology set T represents all the feasible input-output combinations (i.e. production plans) (Bogetoft and Otto (2010)). Roughly, it tells us “what is possible” when converting resources x to products or services y .

$$T = \{(x, y) \in \mathbb{R}_+^n \times \mathbb{R}_+^m \mid x \text{ can produce } y\} \quad (2.4)$$

Several axioms are essential when defining the possibility set T (Daraio and Simar (2007), Shepherd (2015)). We need to assume that the set T satisfies the following conditions.

Condition 1 - Free disposability The free disposability assumption is defined as follows.

$$(x, y) \in T, x' \geq x, 0 \leq y' \leq y \Rightarrow (x', y') \in T \quad (2.5)$$

This means that if the combination (x, y) belongs to the technology set T , then the combination (x', y') that produces less output with more input is also in T ⁸.

Condition 2 - Convexity The set T is assumed to be convex, which means that if two points (x_1, y_1) and (x_2, y_2) belongs to T then every combination in-between those two is also feasible (i.e. in T). For $\alpha \in [0, 1]$, we have

$$(x_1, y_1), (x_2, y_2) \in T \Rightarrow (x, y) = \alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) \in T \quad (2.6)$$

The convex combination (x, y) is a mixture of input-output combinations (x_1, y_1) and (x_2, y_2) . The convexity assumption states that—if (x_1, y_1) and (x_2, y_2) are feasible— (x, y) is feasible too.

Condition 3 - Returns to scale The shape of set T is defined by the assumed *returns to scale* (RTS). RTS expresses “the relation between a proportional change in inputs to a productive process and the resulting proportional change in output” (Daraio and Simar (2007)).

$$(x, y) \in T \Rightarrow (\lambda x, \lambda y) \in T \quad \text{where } \forall \lambda \in \Lambda(r) \quad (2.7)$$

⁸This assumption is sometimes called the “*possibility of destroying goods without costs*” (Daraio and Simar (2007)).

The RTS assumptions (r) most commonly used in regulation are constant (crs), variable (vrs), and decreasing returns to scale (drs)⁹.

TABLE 2.3: Returns to scale assumptions

Returns to scale (r)	Assumption
Constant returns to scale	$\Lambda(crs) = \mathbb{R}_0^+$
Variable returns to scale	$\Lambda(vrs) = \{1\}$
Decreasing returns to scale	$\Lambda(drs) = [0, 1]$

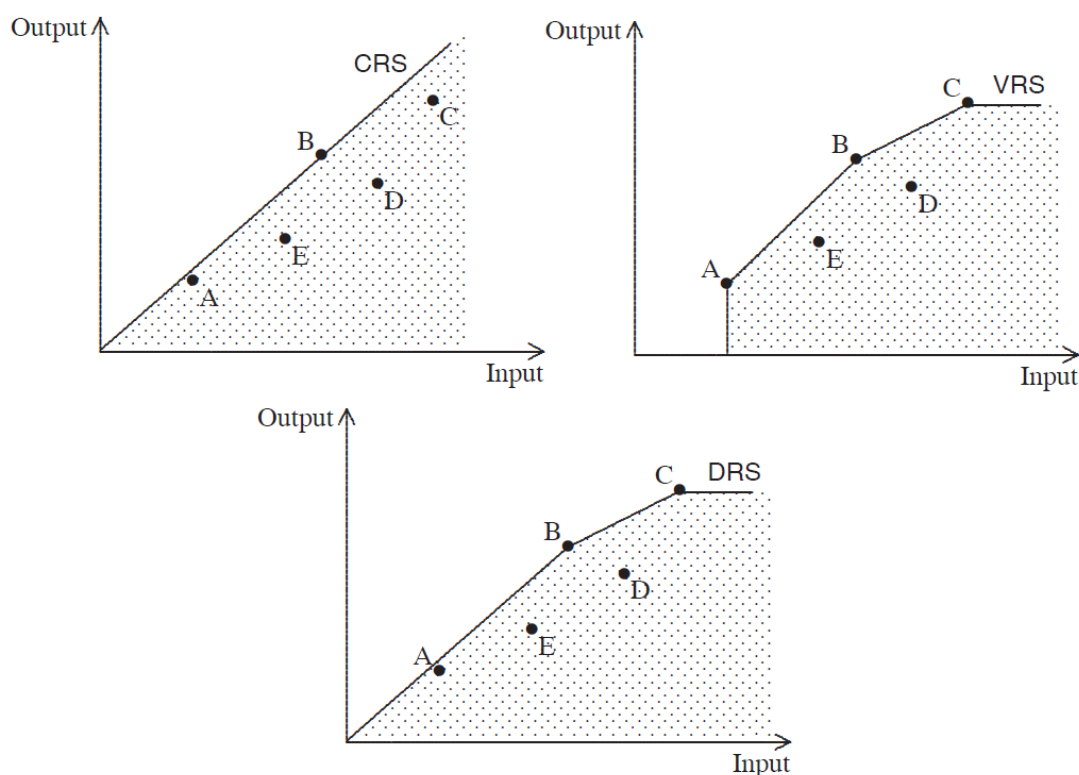


FIGURE 2.1: Technology set T under constant (top left), variable (top right), and decreasing returns to scale (bottom) assuming a DEA-based frontier (Bogetoft and Otto (2010))

Figure 2.1 provides a graphical representation of the set for a single-input single-output technology. We can see that the estimated technology is largest under the crs assumption and smallest under the drs assumption, while the set under vrs has a moderate size. Thus, depending on the chosen returns to scale assumption the *efficient frontier*—and therefore the size of the technology set T —will be different.

⁹There is a total of six different returns to scale assumptions: free disposability hull (fdh), free replicability hull (frh), variable returns to scale (vrs), decreasing returns to scale (drs), increasing returns to scale (irs), and constant returns to scale (crs). See Bogetoft and Otto (2010) for additional details and comparisons.

2.2.2 Efficiency measures

Now that we have defined the efficient frontier we can define the efficiency of a firm by evaluating the distance between this firm's production plan and the frontier. Debreu (1951) proposed a coefficient of resource utilisation and—based on this idea—Farrell (1957) suggested to compute the efficiency of a firm by using the radial distance of its production plan to its projection on the frontier.

2.2.2.1 Farrell efficiency

Farrell efficiency—or *technical efficiency*—can be used in a multi-input multi-output setting by assessing to what extent inputs can be reduced without changing the outputs. As defined in Bogetoft and Otto (2010) the *input-based* Farrell efficiency of an input-output combination (x, y) is the smallest θ such that it is still possible to produce y with θx .

$$E(x, y) = \min\{\theta \mid \theta x \text{ can produce } y\} \quad (2.8)$$

E is the maximal proportional contraction of all inputs x that can produce y .

Figure 2.2 shows a graphical representation of the input-based Farrell efficiency when considering two inputs¹⁰.

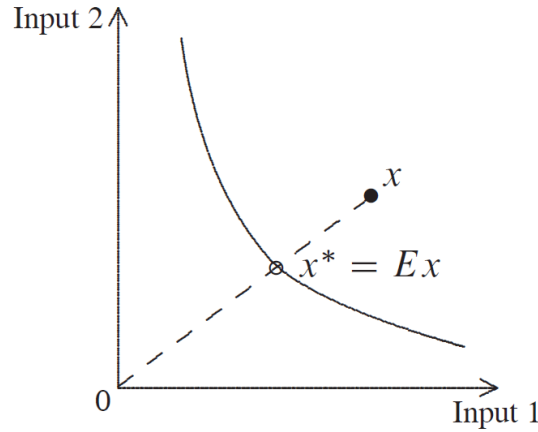


FIGURE 2.2: Graphical representation of input Farrell efficiency E in a two-inputs setting (Bogetoft and Otto (2010))

¹⁰The input isoquant is the set of Farrell-efficient production plans corresponding to the output level that the firm is producing.

2.2.2.2 Cost efficiency

A limitation of the approach of Section 2.2.2.1 is that it is not considering the relative importance of the firm's inputs when computing its efficiency. When input prices are available they can be used to weight the input levels (i.e. quantities) and provide more insight on the efficiency of the firm. Weighting inputs x with their respective prices w allows us to estimate a firm's input-based *cost efficiency*¹¹.

Assuming we have the vector of input prices w we can consider the cost-output combination (c, y) ¹² instead of the input-output combination (x, y) . It is harder for a firm to be efficient under the (c, y) model (i.e. cost efficient), this can be illustrated graphically¹³.

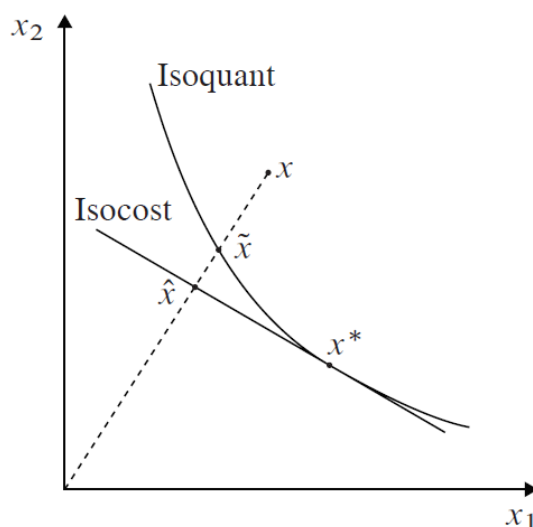


FIGURE 2.3: Graphical representation of the different efficiency measures (Bogetoft and Otto (2010))

Figure 2.3 considers two inputs x_1 and x_2 . Assuming a firm currently producing at production plan x , the isoquant¹⁴ represents the input-based technical efficient frontier—or set of best practices—for the level of output currently being produced by that firm.

Cost efficiency CE can be calculated as the ratio of the minimal cost c^* for the current output level and the actual cost incurred by the firm c .

$$CE = \frac{c^*}{c} = \frac{wx^*}{wx} \quad (2.9)$$

¹¹One could also make no distinction between input levels and prices and consider the costs as inputs (aggregation of quantities and prices). In this case cost efficiency is the Farrell efficiency of costs and can not be decomposed into technical and allocative efficiency.

¹²Where $c = wx$.

¹³The explanation that follows is taken from Bogetoft and Otto (2010).

¹⁴Set of technically efficient production plans for the level of output produced by the firm.

where c^* is the optimal minimal cost and x^* its corresponding input combination for the output produced.

Because the technical efficiency TE is the maximal proportional reduction of input levels it can be calculated as

$$TE = \frac{\|\tilde{x}\|}{\|x\|} = \frac{w\tilde{x}}{wx} \quad (2.10)$$

where $\tilde{x}$ is the projection of x on the technical efficient frontier keeping the ratio of inputs x_1 and x_2 constant (i.e. proportional scaling along the dashed line).

The isocost¹⁵ line represents the different input combinations corresponding to the efficient cost c^* ¹⁶. The difference in cost resulting from choosing input bundle $\tilde{x}$ over x^* is due to a suboptimal allocation between x_1 and x_2 . Even though $\tilde{x}$ and x^* are both on the efficient frontier (i.e. technically efficient), x^* represents the least expensive input-mix. The extent of this suboptimal allocation of inputs can be assessed by *allocative efficiency*.

$$AE = \frac{wx^*}{w\tilde{x}} \quad (2.11)$$

By using the additional information provided by input prices we can compute a firm's cost efficiency and decompose it into two non-additive and mutually exclusive components, technical and allocative efficiency¹⁷.

$$CE = TE \times AE = \frac{wx^*}{wx} = \frac{w\tilde{x}}{wx} \times \frac{wx^*}{w\tilde{x}} \quad (2.12)$$

“To be cost-efficient, the firm must be able to choose the right mix of inputs and use them in a technically efficient manner. It must use the right resources, and it must use them in the right way.” (Bogetoft and Otto (2010))

In order to become cost efficient (i.e. move from x to x^*) two different kinds of improvement are required. First, the firm needs to catch-up its technical inefficiency by moving towards the isoquant. Assuming it does not change its input-mix, this is represented by the firm moving from point x to point $\tilde{x}$. This is done through a more efficient use of inputs allowing the firm to maintain its outputs while proportionally reducing its inputs. Then, in order to catch-up its allocative inefficiency, it will need to change its input-mix—while staying technically efficient—to move along the isoquant from $\tilde{x}$ to x^* . This requires the firm to change its input-mix towards the optimal one by reallocating its resources between the different inputs—here, x_1 and x_2 . It is easier for a firm to be

¹⁵Set of production plans for a constant level of cost.

¹⁶This is why the optimal input combination x^* lies where the isoquant—or technical efficient frontier—is tangent to the isocost.

¹⁷Kopp and Diewert (1982) proposed a method for decomposing the deviations from the cost frontier (i.e. cost efficiency) into the measures of technical and allocative efficiency developed by Farrell (1957).

efficient in a model only considering the level of inputs (i.e. technical efficiency). To be cost efficient not only has the firm to be on a technically efficient point of the technology set ($\tilde{x}$), but it also has to allocate its inputs properly in order to have the least costly input mix (x^*). Hence, a firm will be cost efficient if and only if it is technically and allocatively efficient.

TABLE 2.4: Technical vs. allocative efficiency

Efficiency measure	Requirement	Catch-up strategy
Technical efficiency	Use of optimal processes	Proportional reduction of inputs
Allocative efficiency	Use of optimal input-mix	Reallocation of inputs

It is worth noting that this approach computes allocative efficiency as a “residual” that explains the offset between cost and technical efficiency. Technical efficiency is calculated first and allocative efficiency is defined as “what is left”. Allocative efficiency as we defined it is therefore measured presuming technical efficiency. The rationale behind this approach is that the rate of substitution between the inputs can only be properly defined at the efficient frontier. Bogetoft et al. (2006) proposed an alternative definition of allocative efficiency that does not assume that technical efficiency has first been eliminated. This alternative efficiency assessment is to be considered when reallocation is easier to promote than technical improvements.

2.2.2.3 Dynamic efficiency

In order to assess the change in a firm’s performance and productivity we need to develop dynamic efficiency measures. One of the most famous dynamic measure used in benchmarking is the *Malmquist index*.

Initially used in consumer theory by Malmquist (1953), the Malmquist index was first introduced by Caves et al. (1982)¹⁸ in the benchmarking research. This bilateral index measures the improvement of a firm’s efficiency between two periods t and $t + j$ ¹⁹. Let $E(a, b)$ be the Farrell—or technical—efficiency²⁰ of the performance in period a against the technology set T of period b . The Malmquist index $MI(t, t + j)$ is computed as the geometric mean of the productivity improvement relative to the technology set of period t and the improvement relative to the set of period $t + j$.

$$MI(t, t + j) = \sqrt{\frac{E(t + j, t)}{E(t, t)} \frac{E(t + j, t + j)}{E(t, t + j)}} \quad (2.13)$$

¹⁸Nishimizu and Page (1982) were the first to compute the Malmquist index in an empirical context.

¹⁹For a lag of j periods.

²⁰Developed in Section 2.2.2.1.

Färe et al. (1994a) proposed a decomposition of the Malmquist index. It was demonstrated that an improvement in a firm's performance can be explained by two factors, an individual improvement at the firm level and a technological progress affecting the whole industry equally.

They decomposed the Malmquist index into two non-additive and mutually exclusive components. The first one is the *efficiency change* EC and represents the *catch-up* demonstrated by a firm relative to its efficient frontier. An efficiency change index above unity (i.e. $EC > 1$) means that the firm got closer to the efficient frontier and therefore improved its technical efficiency. The second component is *technological change* TC and represents the change in efficient frontier—or *frontier shift*. A frontier shift above unity (i.e. $TC > 1$) indicates technological progress, which means that the new technology allows firms to produce more with less.

$$MI(t, t+j) = \underbrace{\frac{E(t+j, t+j)}{E(t, t)}}_{EC(t, t+j)} \times \underbrace{\sqrt{\frac{E(t+j, t)}{E(t+j, t+j)} \frac{E(t, t)}{E(t, t+j)}}}_{TC(t, t+j)} \quad (2.14)$$

“It is not sufficient for a firm to improve compared to itself. The firm must also improve relative to others, and they have also benefited from general technological progress. Thus, the only way to improve is to catch-up to the [...] frontier.” (Bogetoft and Otto (2010))

The Malmquist index has two drawbacks. First—as showed by Førsund (2002)—the index is not circular which means that we might have

$$MI(t, t+2) \neq MI(t, t+1) \times MI(t+1, t+2) \quad (2.15)$$

Even though the whole index is not circular the efficiency change EC component is (Daskovska et al. (2010)). The second drawback is that the index is computed without input prices and therefore only represents the technical efficiency (i.e. Farrell efficiency). Maniadakis and Thanassoulis (2004) addressed this issue by proposing a Cost-Malmquist index CM that takes into account input prices.

$$CM(t, t+j) = \Delta OE(t, t+j) \times \Delta CT(t, t+j) \quad (2.16)$$

where ΔOE is the firm-specific overall efficiency change (corresponds to EC in (2.14)) and ΔCT the cost-technical change of the industry (corresponds to TC). Note that the overall cost efficiency change ΔOE can be further decomposed into technical efficiency change ΔTE and allocative efficiency change ΔAE (see (2.12)).

$$\Delta OE(t, t+j) = \Delta TE(t, t+j) \times \Delta AE(t, t+j) \quad (2.17)$$

2.2.3 Frontier models

Now that we have defined different efficiency measures we need a model to compute the technology set T presented in Section 2.2.1. The most common approach in modern benchmarking is to use *frontier models* that only model the frontier of the technology (i.e. the set of best practices).

Following Bogetoft and Otto (2010) frontier models can be classified into four categories according to a simple taxonomy. We can distinguish between parametric and non-parametric models. The parametric approach defines a set of parameters a priori while the non-parametric one does not. This makes the latter approach not as restricted since it does not assume as much as the former. We also distinguish between deterministic and stochastic models. Deterministic models assume that any variation in the observations holds meaningful information that should be used in computing the technology. Stochastic models—on the other hand—assume that observations can be affected by random noise and aims at removing its impact on efficiency estimates.

TABLE 2.5: Frontier models (Bogetoft and Otto (2010))

	Deterministic	Stochastic
Parametric	Corrected Ordinary Least Squares (COLS)	Stochastic Frontier Analysis (SFA)
Non-parametric	Data Envelopment Analysis (DEA)	Stochastic Data Envelopment Analysis (SDEA)

Table 2.5 summarises the four main frontier models used in the benchmarking literature. Among them two models have been given much more attention, especially in the regulatory context: *Data Envelopment Analysis* (DEA) and *Stochastic Frontier Analysis* (SFA). Both of these approaches have their advantages.

“An important property of a benchmarking approach is its ability to reflect and respect the characteristics of the industry. [...] Another important property of a benchmarking approach is its ability to cope with noisy data.” (Bogetoft and Otto (2010))

Being a non-parametric deterministic model DEA is the most flexible model as it will adapt to the data without assuming anything about its structure and consider that any variation carries relevant information on the industry. Despite being less flexible in the mean structure of the data because of its a priori assumptions, SFA is better at dealing with noisy data as it will try to mitigate the effect of random noise on its estimation of the technology.

We decide to use DEA as our frontier model. One should know that the main drawback of DEA is that it requires a lot of data points in order to be accurate. Since the method is non-parametric the shape of the frontier will be computed based on the observed production plans. Fortunately, when assessing the efficiency of network operators, regulatory authorities have access to a wide panel of firms granting them with numerous data points to make their estimation of the technology as accurate as can be.

2.2.4 Data Envelopment Analysis

The *CCR DEA model* proposed by Charnes et al. (1978) is “a linear programming method that constructs a non-parametric production frontier by fitting a piece-wise linear surface over the data points” (Coelli (2003)). As explained in Section 2.2.3, the model does not make any assumption on the technology—or frontier—structure and exclusively use data points to model it.

2.2.4.1 Models for technical and cost efficiency

Assume n firms each using a vector of inputs x to produce a vector of outputs y under the technology T . By observing the production plans (x_n, y_n) the model can compute the technical efficiency TE_i developed in Section 2.2.2.1 for a firm i operating at production plan (x_i, y_i) . The DEA model calculates the technical efficiency $E(x_i, y_i)$ as follows.

$$TE_i = E(x_i, y_i) = \min_{\theta_i} \{ \theta_i | (\theta_i x_i, y_i) \in T(r) \} \quad (2.18)$$

where $T(r)$ is the technology set under the returns to scale assumption r^{21} . The minimisation model is

$$E(x, y) = \min_{\theta, \lambda} \{ \theta \} \quad (2.19)$$

$$\text{s.t. } y \leq \sum_{n=1}^N \lambda_n y_n \quad (2.20)$$

$$\theta x \geq \sum_{n=1}^N \lambda_n x_n \quad (2.21)$$

$$\lambda_n \in \Lambda(r) \quad \forall n \in N \quad (2.22)$$

As explained in Section 2.2.2.2, if we have access to a price vector w for the input vector x we can weight those inputs to estimate firm i 's input-based cost efficiency. Following

²¹See Section 2.2.1.

Agrell et al. (2005) the DEA-based cost norm can be defined as

$$C(y_i, w_i) = \min_{x_i} \{w_i x_i | (x_i, y_i) \in T(r)\} \quad (2.23)$$

By observing production plans (x_n, y_n) with input prices w_n the DEA model defines the cost of the best practice producing y with at prices w as

$$C(y, w) = \min_{x, \lambda} \{wx\} \quad (2.24)$$

$$\text{s.t. } y \leq \sum_{n=1}^N \lambda_n y_n \quad (2.25)$$

$$x \geq \sum_{n=1}^N \lambda_n x_n \quad (2.26)$$

$$\lambda_n \in \Lambda(r) \quad \forall n \in N \quad (2.27)$$

This function gives the optimal cost a firm producing outputs y at input prices w could be incurring. The cost efficiency of firm i can then be computed as the ratio of the optimal minimal cost $C(y_i, w_i)$ of producing the outputs y_i —according to the best practice—on the cost $w_i x_i$ actually incurred by firm i .

$$CE_i = \frac{C(y_i, w_i)}{w_i x_i} \quad (2.28)$$

As we showed in (2.12), after having computed technical and cost efficiency we can easily find allocative efficiency.

$$AE_i = \frac{CE_i}{TE_i} = \frac{C(y_i, w_i)}{w_i x_i E(x_i, y_i)} \quad (2.29)$$

$$CE_i = TE_i \times AE_i = \underbrace{E(x_i, y_i)}_{\text{technical efficiency}} \times \underbrace{\frac{C(y_i, w_i)}{w_i x_i E(x_i, y_i)}}_{\text{allocative efficiency}} \quad (2.30)$$

2.2.4.2 Non-discretionary variables

When benchmarking public utilities against one another the regulator should consider the fact that they may not be operating under the same conditions. As Coelli (2003) pointed out “*the least efficient firms are likely to try to convince the regulator that they have good reasons for being less efficient*”, and they are not necessarily wrong.

“*The applicability of DEA [...] depends crucially on the underlying production process. It is essential that the model be consistent with known properties of the production process.*” (Ruggiero (1998))

Production is assumed to be determined by both discretionary²² inputs (e.g. operating expenditure, capital) and non-discretionary environmental factors of production (e.g. market density) (Bradford et al. (1969)). Compared to firms from denser and more competitive industries, natural monopolies such as DSOs are few and spread out, each taking care of its own region. They thus exhibit varying exogenous preconditions called *non-discretionary*, *non-controllable*, or *z-variables*. Empirical evidence suggests that such variables have a significant impact on the productivity of utilities (Ruggiero (1998)).

Several extensions of the standard DEA model have been proposed to account for non-controllable inputs. The rationale is to break the link between the efficiency and fixed factors by assessing the efficiency of the discretionary inputs while controlling for the non-discretionary ones. Banker and Morey (1986) proposed an extended DEA model to control for such fixed factors by adding a new set of constraints to the model developed in (2.19). Assuming a vector of non-discretionary inputs z the model can be rewritten as follows.

$$E(x, y, z) = \min_{\theta, \lambda} \{\theta\} \quad (2.31)$$

$$\text{s.t. } y \leq \sum_{n=1}^N \lambda_n y_n \quad (2.32)$$

$$\theta x \geq \sum_{n=1}^N \lambda_n x_n \quad (2.33)$$

$$z \geq \sum_{n=1}^N \lambda_n z_n \quad (2.34)$$

$$\lambda_n \in \Lambda(r) \quad \forall n \in N \quad (2.35)$$

where the set of constraints (2.34) is similar to set (2.33) except for the fact that it is not assuming a possible reduction of inputs by θ .

The vector of non-discretionary inputs z is included to assure comparable technologies but is considered fixed and therefore not minimised by the DEA model. Non-controllable inputs must be computed in such a way that a lower value means more “efficiency-friendly” conditions. (Agrell et al. (2005))

*“This modification purportedly controls for fixed factors of production by requiring a convex combination of the referent production possibilities to have an environment no better than the DMU under analysis.”*²³ (Ruggiero (1998))

²²Controllable.

²³Nonetheless—because the convexity and returns to scale assumptions presented in Section 2.2.1 are also defined for non-discretionary inputs—the technology set might be erroneously restricted, thus leading to inaccurate efficiency measurement. (Ruggiero (1996))

This approach is sometimes defined as *sub-vector efficiency* since it estimates the potential reduction of only a sub-vector (i.e. z) of the vector of all inputs (i.e. $x \cup z$) (Färe et al. (1994b)). The set of z -variables is considered as an additional set of inputs.

In order to simplify the DEA program used in computing sub-vector efficiency one can also consider non-controllable inputs z as negative outputs (Bogetoft and Otto (2010)). The set of constraints (2.34) can thus be rewritten as follows.

$$-z \leq \sum_{n=1}^N \lambda_n (-z_n) \quad (2.36)$$

The model for cost efficiency developed in (2.24) can also be extended to consider sub-vector efficiency by adding the set of constraints (2.34) to the model. We can thus developed cost, technical, and allocative efficiency measurements taking into account the idiosyncratic environments and conditions public utilities are operating under.

$$CE_i = \frac{C(y_i, w_i, z_i)}{w_i x_i} = \underbrace{\frac{E(x_i, y_i, z_i)}{w_i x_i}}_{\text{technical efficiency}} \times \underbrace{\frac{C(y_i, w_i, z_i)}{w_i x_i E(x_i, y_i, z_i)}}_{\text{allocative efficiency}} \quad (2.37)$$

2.2.5 Benchmarking in regulation

“One of the more prominent applications of state-of-the-art benchmarking is in the regulation of natural monopolies in general and electricity and gas networks, in particular. Benchmarking studies applied to inform such regulation has considerable economic impact on firms and consumers alike.” (Agrell and Bogetoft (2013))

Distribution operators are few and operating under various exogenous conditions. By using a benchmarking approach in regulation, competition is simulated among the regulated entities. For lack of market competition benchmarking forces operators to compete against one another by comparing them. Regulatory authorities are relying more and more on benchmarking methods when establishing their incentive regulation schemes. (Haney and Pollitt (2009))

The Argentinean land transport regulator made very clear that promoting efficiency among network utilities should be a priority in regulation (Campos-Méndez et al. (2001)). Its three principal obligations were the following:

- Secure that the decisions of the regulated operators take into account the interest of consumers.
- Make sure that the sector is competitive and that consumers are supplied at a fair price.

- Provide the operators with the proper incentives to grow more efficient in their operations and investments.

The use of benchmarking in regulation has many advantages. One of its most interesting aspect is that regulators do not need much experience in the regulated industry to compute meaningful efficiency measures. Thanks to benchmarking, regulatory costs are greatly reduced. Another valuable feature is that benchmarking “*externalises the terms of service*” (Lowry and Getachew (2009)), which means that the regulatory requirements are not based on a firm’s own cost structure but on external targets set by the best practices of the industry. Those targets are established according to a relevant set of comparators since benchmarking takes into account discrepancies in operating conditions.

Nonetheless—despite its attractiveness for productivity assessment—using benchmarking in the regulatory context has several limitations. A recurring critic made in the regulatory literature is that a poorly defined benchmarking model can expose operators to non-negligible risks. Even though its primary objective is to protect social welfare and consumers, regulation should not jeopardise the stability of the regulated entities. Empirical evidence shows that the computed efficiency measures vary greatly depending on the benchmarking methods that is being used (Jamasb and Pollitt (2003)). In Section 2.2.3 we briefly stated the pros and cons of the different frontier models, with a focus on DEA and SFA. Because of their different approaches, the efficiency assessed by the regulator can be significantly different whether he is using DEA or SFA—or any other benchmarking model. The model specification and chosen set of variables has also a great impact on efficiency measurement.

“Efficiency scores can be sensitive to model specification and choice of input, output, and environmental variables.” (Jamasb and Pollitt (2000))

Shuttleworth (2005) argued that the regulatory risk due to benchmarking makes it an inappropriate tool for regulation. Since different productivity assessments can be obtained depending on the chosen approach, relying on benchmarking when setting regulatory schemes poses a great risk. To mitigate the biases of the different available measures, some authors and regulators proposed to consolidate the results from various preferred approaches and model specifications in order to compute a robust and trustworthy efficiency estimate (Jamasb and Pollitt (2000)). Bauer et al. (1998) defined consistency conditions to assess if benchmarking results can be used in the regulatory context. For instance, one may require the efficiency estimates resulting from the set of selected models to have similar means and standard deviations before considering using them in regulation. The gap between a firm realised cost and the cost efficient frontier

can be in part due to factors that were not taken into account in the analysis and might thus not be entirely explained by inefficient production (Makholm (2007)).

Keeping those limitations in mind regulatory authorities should not, however, miss out on the advantages of benchmarking. The aforementioned limitations have only been presented to provide a critical view of benchmarking in the regulatory context.

2.3 Frontier-based revenue cap

In Section 2.1.2 we showed that most of the European regulators are in favour of high-powered fixed-price regimes. The most widely used high-powered mechanism is revenue cap regulation.

Sometimes called *CPI-X regulation*, this scheme redefines the formula from (2.1) to set a cap on the regulated firm's revenue instead of its price. The formula caps the revenue each firm will be allowed to earn throughout a pre-determined regulatory period of T years—typically 3 to 5 years. The revenue cap is inflation adjusted and stipulates annual productivity improvement. Following Agrell and Bogetoft (2013), the scheme can be defined using the following generic function.

$$R_i(t) = R_i(0)(CPI - X)^t, \quad t = 1, \dots, T \quad (2.38)$$

where $R_i(t)$ is the allowed revenue for firm i at year t , $R_i(0)$ its base allowed revenue at the outset of the regulatory period, and $(CPI - X)$ a yearly adjustment for price developments (i.e. adjustment for inflation) and efficiency requirements. The consumer price index (CPI) is typically used as a measure of inflation. Efficiency requirements are embodied by the productivity offset factor X .

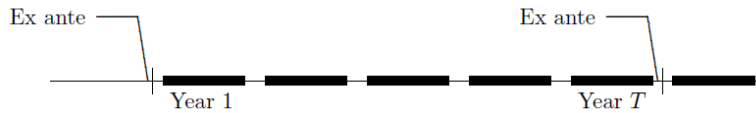


FIGURE 2.4: Ex ante regulation (Agrell and Bogetoft (2013))

We already stated in Section 2.1.1 that fixed-price regimes such as the revenue cap regulation are established ex ante (i.e. before the regulatory period). As shown in Figure 2.4, the revenue cap components (i.e. base revenue $R_i(0)$, price adjustment CPI , and productivity requirement X) are reviewed every T years at the outset of the next regulatory period. Because the cap is fixed for the duration of the regulation, if firm i manages to outperform the expected productivity improvement X it will be allowed to

retain the profit resulting from those efficiency gains. The regulator should be careful when setting the revenue cap as it should not compromise the financial integrity of the firms under its regulation. Because the regulated entities should be able to cover their real costs to avoid bankruptcy, regulators must be extremely careful.

As explained in Agrell et al. (2005) and Bernstein and Sappington (2000), a poorly defined revenue cap can lead to two practical issues:

- If the revenue cap is set too high informational rents will be too great and firms will earn excessive profit at the expense of consumers. Because firms will get away with poor cost efficiency, social welfare will be lowered.
- If the revenue cap is set too low the financial viability of the operators is put at risk as they might not be able to break even.

The risks and rewards of under- or outperforming the efficiency expectations of the regulator can have damaging repercussions if the cap is not set correctly.

2.3.1 X-factor

“The X-factor should reflect the extent to which the regulated industry has historically achieved higher productivity growth and faced lower input price inflation than other industries in the economy.” (Bernstein and Sappington (2000))

According to the risks stated in Section 2.3, the X-factor should represent a reasonable and surmountable challenge for the regulated entities.²⁴

Because the allowed revenue is fixed and decreasing at a yearly rate of X%—leaving aside inflation—the only way for the operators to break even is through the minimisation of their actual costs. Any efficiency gains resulting from an improvement beyond those required by the model will be retained by the firms, providing them with a major incentive for incumbent inefficiency reduction (i.e. catch-up to best practice).

While using the consumer price index (CPI) as the price adjustment is quite straightforward, setting X is not as trivial. Beesley and Littlechild (1989) stated that the X-factor is *“a number specified by the government”*. Bernstein and Sappington (1999) argued that—when no major economical disruptions (e.g. structural changes) are expected in the regulated industry—historical data on previously observed productivity growth rates offers a reasonable estimate to assist the regulator in setting the X-factor. Most regulators choose to set a firm-specific X-factor as firms might not be operating under the

²⁴An increase in X will result in a decrease in the cap R_i (i.e. $\partial R_i / \partial X < 0$).

same efficiency. Companies being less efficient would therefore be required to improve faster than others operating closer to the efficient frontier.

The revenue cap from (2.38) can be redefined as follows (Bogetoft and Otto (2010)).

$$R_i(t) = C_i(0)(1 - \chi - \chi_i)^t, \quad t = 1, \dots, T \quad (2.39)$$

where the cost level of the outset of the regulation $C_i(0)$ is used as the base revenue $R_i(0)$ and X has been decomposed into a general and individual requirement, χ and χ_i . Just like the X-factor those two components are estimated through benchmarking analyses.

The general expected productivity improvement χ is most commonly set based on the technological change—or frontier shift—component of the Malmquist index (see (2.14)) in the years preceding the regulation. For example, if a frontier shift of α has been observed in the industry for the past years it means that—on average—all the firms from the industry demonstrated an improvement of $(\alpha - 1)\%$ (or regression if $\alpha < 1$). Hence, the regulator could set the requirement χ to $\max\{(\alpha - 1); 0\}\%$ ²⁵.

The individual requirement χ_i is set based on the efficiency of firm i —most commonly the Farrell efficiency E_i developed in Section 2.2.2.1—at the outset of the regulation. While χ can be associated with the efficient frontier shift, χ_i represents the extent to which firm i needs to catch-up to the frontier each year. The rule of thumb is that the lower efficiency E_i is, the higher should χ_i be.

“There is [...] a considerable element of negotiation in the rules that are applied [when setting the individual requirement].” (Bogetoft and Otto (2010))

As we will see in Section 2.4 the number of years operators are granted to catch-up to efficient varies greatly depending on the regulatory authority. Just like for the estimation of the general requirement χ , the choice of χ_i should be based on an analysis of past demonstrated catch-up rates (Bogetoft and Otto (2010)). Nonetheless, by setting χ_i based on the past performance of firm i the regulator might encourage it to strategically hide its true productivity improvement potential by fear of being penalised for it in subsequent periods. This “*ratchet effect*” arises from “*the well-known tendency of planners to use current performance as a criterion in determining future goals*” (Weitzman (1980)).

²⁵If the industry demonstrated technological regression (i.e. $\alpha < 1$) the regulator would then set the requirement χ to zero.

“Practitioners of the CPI-X regulation [...] stress that performance measures used to set the X-factor for a firm must not be derived solely from the firm’s past performance, because this will negate the incentives involved.” (Coelli (2003))

As we will see in Section 2.4, the approach used by most regulators is to decide on the number of years every operator should be granted to catch-up to efficiency. All the firms are therefore given the same amount of time to completely remove their incumbent inefficiency.

Agrell et al. (2005) proposed the introduction of a parameter δ representing the fraction of its initial cost efficiency deficit a DSO is required to eliminate each year. Following the insights of this paper the model defined in (2.39) can be rewritten as

$$R_i(t) = R_i(0)(1 - \chi - \delta(1 - CE_i(0)))^t, \quad t = 1, \dots, T \quad (2.40)$$

where $CE_i(0)$ is the cost efficiency of DSO i at the outset of the regulatory period—computed through benchmarking—and $\delta(1 - CE_i(0))$ the yearly inefficient cost reduction imposed by the regulator (i.e. χ_i). The parameter δ has to be set such that the total required efficiency improvement of the whole regulatory period of T years does not exceed firm i ’s incumbent inefficiency (i.e. $(1 - CE_i(0))^{26}$).

$$1 - (1 - \delta(1 - CE_i(0)))^T \leq 1 - CE_i(0) \quad (2.41)$$

2.4 Practical cases

Some countries require DSOs to catch-up to best practice in only a year’s time (e.g. Denmark) while others are allowing the firms up to two regulatory periods of 5 years to remove their incumbent cost inefficiency (e.g. Germany) (Agrell and Bogetoft (2013)). By looking at the whole industry and its average historical catch-up rate, the regulators usually decide on the number of years they will give the firms to eliminate their inefficiency.

Schweinsberg et al. (2011) analysed the cost benchmarking approaches of different European regulators in the energy sector of electricity distribution. We selected five of them.

²⁶In this scheme DSO i is not eliminating $\delta\%$ of its initial cost inefficiency per year but reducing its cost by $\delta(1 - CE_i(0))\%$ each year.

2.4.1 Austria - E-Control

Austrian electricity DSOs are operating under incentive regulation since 2006. A hybrid price cap regime was established for the 2006-2009 period. The formula was the following.

$$C_t = C_{t-1} \times [(1 - X) \times (1 + \Delta NPI_t)] \times (1 + k \times \Delta M_t) \quad (2.42)$$

where C_t is the cost expected in year t , X the efficiency requirement, and ΔNPI_t and $k\Delta M_t$ accounts for inflation and change in output quantity, respectively. The X-factor is firm-specific.

$$X = 1 - (1 - X_{gen}) \times \sqrt[N]{ES} \quad (2.43)$$

where X_{gen} corresponds to the general requirement χ , ES is the efficiency score (i.e. Farrell efficiency) at the outset of the regulatory period, and N represents how many periods of T years DSOs have to catch-up their inefficiency (i.e. $1 - ES$). The decision of parameter X_{gen} was based on the frontier shift of various countries.

Concerning the efficiency score ES , it was computed as the weighted mean of the scores given by three different benchmarking models.

$$ES = 0.4 \times ES_{DEA(I)} + 0.2 \times ES_{DEA(II)} + 0.4 \times ES_{MOLS} \quad (2.44)$$

where DEA(I) and DEA(II) are DEA models with different model specifications and sets of variables, and MOLS²⁷ is a *Modified Ordinary Least Squares* model with the same specification and variables as DEA(I). As we explained in Section 2.2.5, the rationale behind this approach is to mitigate the individual biases of the three different benchmarking approaches.

The regulator decided to put a lower bound of 74.76% on ES so that DSOs with lower scores will be asked to catch-up as if they had a score of 74.76%.

2.4.2 Germany - Bundesnetzagentur

In 2009 the German regulator established a new revenue cap regulation. The formula was the following.

$$R_t = C_{ex,t} + [C_{en,0} + (1 - d_t) \times C_{i,0}] \times (i_t - x_{gen,t}) \times ef_t \quad (2.45)$$

²⁷Like Corrected Ordinary Least Squares (COLS), Modified Ordinary Least Squares (MOLS) is a parametric deterministic frontier model (see Section 2.2.3).

where R_t is the allowed revenue, $C_{ex,t}$ the non-controllable costs²⁸, $C_{en,0}$ the efficient controllable costs, $C_{i,0}$ the identified inefficient costs, and $(i_t - x_{gen,t}) \times ef_t$ an update factor accounting for inflation (i_t), frontier shift ($x_{gen,t}$), and investment incentives (ef_t).

The assessed inefficient costs $C_{i,0}$ had to be linearly removed through factor d_t . The regulated entities were allowed two regulatory periods of T years to catch-up to best practice (i.e. $d_t(T) = t/2T$).

$$C_{i,t} = (1 - d_t)C_{i,0} = \left(1 - \frac{t}{2T}\right) C_{i,0} \quad (2.46)$$

The inefficient costs $C_{i,t}$ were identified through a complex benchmarking methodology that used two DEA models and two SFA models.

$$ES = \max(ES_{DEA(I)}, ES_{DEA(II)}, ES_{SFA(I)}, ES_{SFA(II)}, 0.6) \quad (2.47)$$

According to (2.47) the German regulator selected to benchmarking model that computed the most favourable efficiency score with a lower bound at 60%.

2.4.3 Norway - Norwegian Water Resources and Energy Directorate

In 2002 the Norwegian Water Resources and Energy Directorate (NVE) reformed its regulatory scheme as a revenue cap with catch-up requirement. The expected individual efficiency improvement was measured via DEA. Two different DEA models were computed and the firms were assigned the most favourable score by the regulator. An upper bound of 3% was set on the yearly requirement. The improvements target EFK was computed as follows.

$$EFK = 1 - (1 - (1 - KE) \times 0.3824)^{1/4} \quad (2.48)$$

where KE is the TOTEX efficiency. The NVE decided that operators should recover 38.24% of their inefficiency in the next four years of regulation.

2.4.4 Spain - Comisión Nacional de Energía

Since Spain has established a new regulatory framework for electricity DSOs in 2009 the regulator has been using the following revenue cap formula.

$$R_t = R_0 \times (1 + A_t) + Y_{t-1} + Q_{t-1} + L_{t-1} \quad (2.49)$$

²⁸Updated yearly.

where the starting revenue R_0 is updated every year with the index

$$A_t = 0.2 \times (i_{CPI} - x) + 0.8 \times (i_{IPI} - y) \quad (2.50)$$

where i_{CPI} and i_{IPI} are the inflation rates for consumer and industrial prices, while x and y are their corresponding efficiency factors. The Spanish Ministry of Economics set those two factors to 0.8% and 0.4% per annum without even consulting the commission.

2.4.5 Sweden - Energy Market Inspectorate

Starting in 2012 the Energy Market Inspectorate (EI) established a new regulatory scheme designed as a revenue cap regulation.

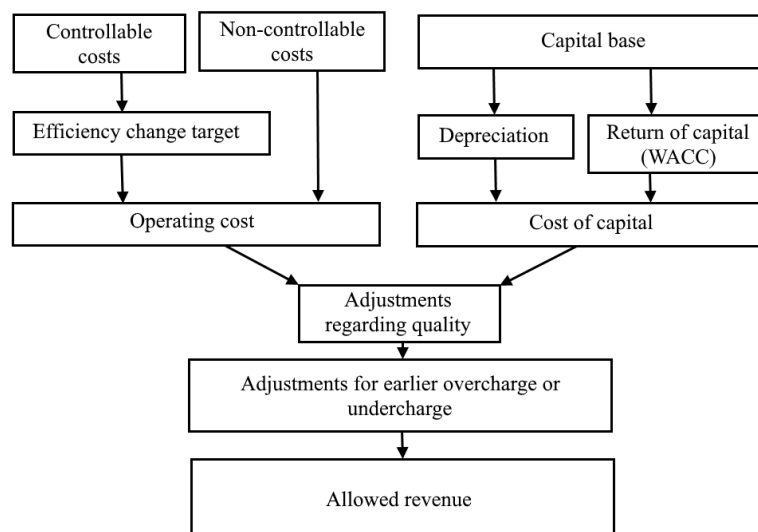


FIGURE 2.5: Elements of the EI regulation (Schweinsberg et al. (2011))

The regulatory period lasted four years and the revenue caps were calculated following the framework illustrated in Figure 2.5. As we can see the regulator determined an efficiency change target on the controllable costs only (e.g. labor and other operating expenditures). The efficiency target for the controllable operating costs was set to 1% per annum. No individual benchmarking was carried out to determine this required catch-up which means that all DSO were asked the same efficiency improvement without regard for their differences in terms of efficiency or catch-up potential. Nonetheless, this efficiency target was not set arbitrarily but according to the result of a productivity analysis of the sector through DEA and SFA.

2.5 Summary and review

In the need for incentive regulation of network companies, regulators often rely of various benchmarking tools to established their fixed-prices regimes. As we have seen in Section 2.4, it is clear that benchmarking measures like those presented in Section 2.2 are used extensively in the regulatory context. Nevertheless, regulators should be careful when establishing their models as the choice of the benchmarking approach (i.e. parametric versus non-parametric, deterministic versus stochastic) and model specification has a great impact on the efficiency score estimates (see Section 2.2.5).

When setting the firm-specific X-factor—or efficiency requirement—of their schemes, regulators usually process as follows.

1. Definition of the benchmarking model(s) used to compute the efficiency scores for the set of firms.
2. Definition of the number of years T inefficient firms will be granted to catch-up to efficiency.
3. Calculation of the individual yearly requirement needed to reach the efficient frontier after T years.

The limitation of this approach is that the allowed period to catch-up lacks foundation and is not firm-specific. Leaving aside the various parameters and factors regulators have been adding to their revenue cap formulas, most of the reviewed regulations rely on computing the χ_i requirement as a function of firm i 's incumbent efficiency θ_i and the number of years T that the whole industry has to eliminate inefficiency. While the efficiency θ_i is computed through extensive benchmarking analyses, the catch-up period T is usually set based on past productivity growth, but lacks research.

“The period over which companies must achieve target performance identified by benchmarking is unclear.” (Lowry and Getachew (2009))

“The key to effective regulation is found in the access to information” (Agrell et al. (2005)). The X-factor and catch-up period should thus be given a deep and thorough analysis in terms of individual cost efficiency improvement potential. In order to be able to assess this improvement potential, regulators should fully use the different benchmarking measures that have been developed.

Recall that our research question was *“How can regulatory authorities make use of the available benchmarking tools and measures to set the required firm-specific catch-up speed in the fixed-price regulation of network utilities?”*

Looking at the current state-of-the-art it is clear that benchmarking is extensively used to compute the cost efficiency scores of the regulated firms. The incumbent cost inefficiency and allowed period to catch-up T are then used to compute the requirement χ_i . Nonetheless, scant attention has been given to the decomposition of cost efficiency into technical and allocative efficiency. Even though those measures are often computed and analysed in the benchmarking literature, the decomposition of cost efficiency has yet to be used in the computation of the X-factor.

In Chapter 3 we propose an approach addressing how regulators can use this decomposition to better assess the source of an operator's cost inefficiency. Following this decomposition we propose a model to define a firm-specific catch-up period T_i based on the assessment of a DSO's sources of cost inefficiency.

Chapter 3

Model

In this chapter we propose an approach and model for computing the allowed number of years to catch-up cost inefficiency at the firm level. To answer the need for a firm-specific catch-up period we propose an assessment of the decomposition of cost efficiency into technical and allocative efficiency. We present in more details the definition of a firm's efficiency profile and its implications in terms of catch-up speed. We propose a complexity factor computed from the efficiency profile and illustrating the complexity of a firm's catch-up strategies. From this complexity factor a linear model is proposed to compute the firm-specific catch-up period. The strengths and weaknesses of the model are then assessed.

It is important to specify that variable returns to scale¹ is assumed in our DEA models.

3.1 Decomposition of cost efficiency

“The availability and access to information is a key issue in the regulatory game and regulators can use benchmarking to undermine the informational asymmetry.” (Agrell and Bogetoft (2013))

When setting the revenue cap and—more specifically—its catch-up requirement χ_i regulators should extensively use the information and benchmarking tools at their disposal. The more information a regulator possesses the more accurate he can pretend his regulatory scheme to be. While pure technical efficiency analyses are the most widely used in regulation, by knowing and adding the input prices to its model a regulator is able to assess where each firm is standing compared to the best practice with greater accuracy. Doing so allows him to conduct cost efficiency analyses and decompose this cost

¹See Section 2.2.1.

efficiency into technical and allocative efficiency, providing him with a more detailed benchmarking assessment.

When considering multiple inputs and outputs the *total-factor productivity*—or cost efficiency—can differ between firms for four reasons: pure technical efficiency, scale efficiency, input-mix allocative efficiency, and output-mix allocative efficiency² (Coelli (2003)). As we have seen in Section 2.2.2.2, we can get a more detailed assessment of the sources of a firm’s cost inefficiency by decomposing it. By assessing this decomposition we get a better understanding on why a firm is not operating at optimal cost. Choosing which efficiency measures to use is to be done carefully and according to a *controllability* principle (Bogetoft and Otto (2010)).

3.1.1 Technical efficiency

In Section 2.2.2.1 we focused on *input-based* technical efficiency. Before assessing the technical efficiency of a production plan, one should determine whether the inputs or outputs are controllable by the firm. In the case of network utilities, the level of outputs is exogenously determined by the market size, profile, and density. Because DSOs are not in control of their outputs (i.e. demand is determined by the energy market) regulators should always assess input-based instead of output-based technical efficiency.

Technical input efficiency can be defined in the light of the input-based *Pareto efficiency*: “A DMU³ is Pareto-efficient if it is not possible to lower anyone of its input levels without increasing at least another one of its input levels and/or without lowering at least one of its output levels.” (Thanassoulis (2013))

As we explained, input-based technical efficiency represents—for a given input-mix⁴—the extent to which inputs can be proportionally reduced while still producing the same outputs—according to the assumed technology set (Koopmans (1951)).

In the context of network regulation, technical efficiency⁵ can be interpreted as an estimate of *managerial performance*. Because it represents how efficiently an operator is using its input-mix, catching-up technical inefficiency would mean eliminating managerial sub-performance through better procedures and processes.

²In Section 2.2.2 we only developed pure technical efficiency and input-mix allocative efficiency. This is because—when regulating network companies—scale and output-mix allocative efficiency are generally assumed not controllable.

³Decision-Making Unit.

⁴“The input-mix of a DMU is reflected in the ratio its input levels are to each other.” (Thanassoulis (2013))

⁵From now on we will refer to “input-based technical efficiency” as “technical efficiency”.

Let $\phi_i(t)$ be the unmeasurable *managerial improvement effort* of firm i in period t . We assume there is an unknown function $f(\cdot)$ such that

$$TEC_i(t, t+1) \leq f(\phi_i(t)|TE_i(t)) \quad (3.1)$$

where $TEC_i(t, t+1)$ is the technical efficiency change between periods t and $t+1$ ⁶ and $TE_i(t)$ is the technical efficiency in period t .

What (3.1) means is that—for an observed initial technical efficiency score $TE_i(t)$ —the maximal possible technical efficiency catch-up is determined by the improvement effort $\phi_i(t)$ management demonstrated throughout period t —through function $f(\cdot)$ ⁷.

Many regulators only consider technical efficiency when establishing the individual improvement that their regulatory schemes will require from the firms. The rationale is that—in the short run—technical inefficiency is the most controllable type of inefficiency. Nonetheless, the principal limitation of technical efficiency in the regulatory context is that—because it is independent from the relative prices of resources used in distributing energy—it provides a biased assessment of true production performance.

3.1.2 Allocative efficiency

By considering the values of the operators' resources when comparing them, we get a more meaningful assessment of efficiency. As we have seen in Section 2.2.2.2, it is harder for an operator to be efficient when considering input prices. A firm may be operating on the technical efficient frontier given its current input-mix (i.e. technically efficient) but that input-mix may be sub-optimal and therefore imply inefficient costs compared to other firms operating with a more efficient combination of inputs. Given the prices it is facing, a firm's input-based allocative inefficiency⁸ reflects “how far” this firm is from the least costly mix of inputs.

“Input-mix allocative efficiency is a firm's ability to select the correct mix of input quantities so as to ensure that the input price ratios equal the ratios of the corresponding marginal products, that is, the additional output obtained from an additional unit of input.” (Coelli (2003))

Depending on the decision-making units being benchmarked, input prices might not always be available. It is for instance the case for the benchmarking of schools where it would be irrelevant to associate a price to the socio-economic background of a student

⁶Corresponds to component $\Delta TE(t, t+i)$ in (2.17).

⁷We assume that function $f(\cdot)$ is increasing in $\phi_i(t)$ (i.e. $\partial f(\phi_i(t)|TE_i(t))/\partial \phi_i(t) > 0$) and decreasing in $TE_i(t)$ (i.e. $\partial f(\phi_i(t)|TE_i(t))/\partial TE_i(t) < 0$)

⁸From now on we will refer to “input-based allocative efficiency” as “allocative efficiency”.

(Thanassoulis (2013)). In the case of network companies regulators have usually access to a wide range of variables enabling them to compute input quantities, prices, and costs.

“In the regulated case, allocative inefficiency is most likely the outcome of inefficient mixes of capital and operating expenditures.” (Commission et al. (2013))

For example, let us assume a firm assessed as allocatively inefficient in a benchmarking model considering labour and capital as inputs. If labour is marginally more expensive than capital, then the firm could improve its cost efficiency by reducing personnel (i.e. labour) and investing instead in some kind of automated system (i.e. capital) that will fulfil the same need. Through this capitalisation, the firm would thus reduce labour and increase capital, changing its input-mix for a more efficient one.

As stated in Coelli (2003)—despite the additional information it provides on a firm’s X-inefficiency—when setting the individual requirement χ_i a regulator may want to remove allocative efficiency from its analysis for different reasons. One of them is that the network capital of a DSO (e.g. network length, substations) highly depends on the density of the area it is operating in. Other non-controllable factors might also influence the allocative efficiency catch-up potential of an industry. The access to capital markets is also quite limited for distributors, preventing them from easily reallocating their inputs. Compared to other resources, capital assets are generally long-lived and can not be reallocated that easily. For a DSO it is quite difficult to trade those assets since they are firm-specific and embodied in its operations.

“Limited access to capital markets and national agreements with unions unrelated to the operator’s specialised employment needs are two common examples of sources of allocative inefficiency for which the operator should not necessarily be blamed. Understanding that this is the case is a matter of fairness.” (Coelli (2003))

On the one hand—recalling the controllability requirement stated by Bogetoft and Otto (2010) for a fair benchmarking assessment and regulation—some regulators are not considering allocative efficiency and thus perform pure technical analyses for setting the catch-up requirement χ_i . On the other hand, addressing allocative inefficiency provides them with more precision in terms of efficiency assessment. Thanks to the currently used benchmarking methods (e.g. DEA, SFA) regulators can greatly improve their schemes by observing the decomposition of cost efficiency that those methods allow.

“At the end of the day, regulators are interested in the degree to which firms could be reasonably expected to reduce costs, in real terms. Thus they should be interested in allocative efficiency.” (Coelli (2003))

In a network regulation considering long-term inputs such as network capital in its benchmarking, catching-up allocative inefficiency would imply investment or disinvestment in some of those long-term resources. Unlike technical inefficiency, allocative inefficiency is not internally controllable as it requires purchasing and reselling resources.

Let $\psi_i(t)$ represents the unobservable *financial constraint* of firm i in period t . We assume there is an unknown function $g(\cdot)$ such that

$$AEC_i(t, t+1) \leq g(\psi_i(t)|AE_i(t)) \quad (3.2)$$

where $AEC_i(t, t+1)$ is the allocative efficiency change between periods t and $t+1$ ⁹ and $AE_i(t)$ is the allocative efficiency in period t . The extent to which a firm can catch-up its allocative inefficiency is constrained by its financial constraint ψ_i ¹⁰.

3.1.3 Assessment of efficiency profile

Because technical and allocative inefficiency are not eliminated through the same actions and strategies regulators should address the decomposition of cost efficiency when setting the cost efficiency catch-up requirement of their fixed-price scheme.

Definition 1. The efficiency profile η_i of firm i represents the distribution of its cost efficiency score CE_i into its technical and allocative components, TE_i and AE_i .

$$\eta_i = \begin{pmatrix} CE_i \\ TE_i \\ AE_i \end{pmatrix} \quad (3.3)$$

The efficiency profile η_i allows us to assess the relative “impact” of technical and allocative inefficiency on cost inefficiency¹¹. When setting the catch-up requirement this assessment is crucial since eliminating technical and allocative inefficiency do not require the same strategies and might not be reducible at the same pace.

Example 1. Let there be three DSOs A, B, and C. While they all have an identical cost efficiency of 66.5%, a more detailed benchmarking analysis provides the results given in Table 3.1.

⁹Corresponds to component $\Delta AE(t, t+i)$ in (2.17).

¹⁰Like for function $f(\cdot)$ we assume that function $g(\cdot)$ is increasing in $\psi_i(t)$ (i.e. $\partial g(\psi_i(t)|AE_i(t))/\partial \psi_i(t) > 0$) and decreasing in $AE_i(t)$ (i.e. $\partial g(\psi_i(t)|AE_i(t))/\partial AE_i(t) < 0$).

¹¹Efficiency profile η_i is merely a vector consisting of the cost, technical, and allocative efficiency scores.

TABLE 3.1: Example - Efficiency profile η_i

Measure	Unit	DSO A	DSO B	DSO C
Cost efficiency	%	66.5	66.5	66.5
Technical efficiency	%	70.0	95.0	81.5
Allocative efficiency	%	95.0	70.0	81.5

Even though those three network operators have an identical cost inefficiency of 33.5%, the sources of that inefficiency—and their efficiency profile η_i —greatly differ. We can see that the inefficient costs of DSO A are primarily due to the firm being technically inefficient—30% of technical inefficiency against 5% of allocative inefficiency. The main reason behind DSO A’s inefficiency is managerial sub-performance. On the other hand, DSO B is cost inefficient mainly because of his allocative inefficiency—30% of allocative inefficiency and 5% of technical inefficiency. Unlike DSO A, DSO B is incurring excess costs because of his costly and inefficient allocation of inputs. DSO C’s efficiency profile lies in-between those of A and B with a cost inefficiency equally explained by sub-optimal managerial effort and input-mix—18.5% each.

Because of their different efficiency profile η_i , the strategies and actions required from those three DSOs to remove cost inefficiency are different. Let us suppose that the regulator does not distinguish between those firms and set the χ_i such as they have to catch-up to the frontier at the same speed—as it is done in network regulation now. In order to meet the improvement requirement the main strategy DSO A needs to implement is improved managerial effort for a more efficient use of its resources (i.e. proportional reduction of inputs). For instance, this can be achieved by improving staff performance or eliminating non-technical energy losses through a more optimal metering system. On the other hand—in order to catch-up—DSO B primarily needs to change its input-mix. That can imply reselling part of its network capital or capitalising part of its operating expenditures¹², which is—according to the insights of the literature—not the easiest thing to do in a sector like energy distribution.

From Example 1 we can see that the efficiency profile η_i of a DSO should be assessed and taken into account when setting the catch-up requirement. Hence, we propose the following hypothesis.

Hypothesis 1. The catch-up speed—or cost efficiency change—of a distribution system operator i depends on its efficiency profile η_i .

¹²Assuming the benchmarking model considers TOTEX (i.e. operating and capital expenditures).

3.2 Complexity factor ξ_i

If Hypothesis 1 holds, we can define a complexity factor ξ_i that reflects the impact that the efficiency profile η_i is assumed to have on cost efficiency change.

$$\xi_i = h(\eta_i) = h(CE_i, TE_i, AE_i) \quad (3.4)$$

where $h(\cdot)$ is a function defined by the regulator to translate the efficiency profile η_i of firm i in its complexity factor ξ_i .

The factor ξ_i represents the relative difficulty in catching-up cost inefficiency resulting from η_i . Hence, we need a function $h(\cdot)$ such that, as ξ_i increases, the cost efficiency change CEC_i is expected to decrease¹³.

We define two measures to assess “how much” of firm i ’s cost inefficiency is technical and allocative. Recall from Section 2.12 that TE_i and AE_i are two non-additive and mutually exclusive components.

$$CE_i = TE_i \times AE_i \quad (3.5)$$

By taking the logarithm of (3.5) we obtain

$$\log(CE_i) = \log(TE_i) + \log(AE_i) \quad (3.6)$$

Now that we have converted the product to an addition we can compute the following measures.

$$\zeta_{TE_i} = \frac{\log(TE_i)}{\log(CE_i)} \in [0; 1] \quad \forall CE_i \in]0; 1[\quad (3.7)$$

$$\zeta_{AE_i} = \frac{\log(AE_i)}{\log(CE_i)} \in [0; 1] \quad \forall CE_i \in]0; 1[\quad (3.8)$$

where ζ_{TE_i} and ζ_{AE_i} are the proportion of firm i ’s cost inefficiency attributable to technical and allocative inefficiency, respectively. On the one hand, if inefficient firm i is technically efficiency we will have $\zeta_{TE_i} = 0$ and $\zeta_{AE_i} = 1$. On the other hand, if firm i is allocatively efficiency then $\zeta_{TE_i} = 1$ and $\zeta_{AE_i} = 0$. Those are the two extreme situations and most inefficient firms will be operating in an intermediary situation. We always have $\zeta_{TE_i} + \zeta_{AE_i} = 1$.

One should be careful when using those measures as they are only defined for cost inefficient firms (i.e. $CE_i < 1$). Measures ζ_{TE_i} and ζ_{AE_i} are not defined for a cost efficient

¹³More complexity results in more difficult catch-up.

firm i (i.e. $CE_i = TE_i = AE_i = 1$) because it would result in dividing zero by zero—since we know that $\log(1) = 0$ ¹⁴.

Example 2. Let us consider again the DSOs from Example 1. By computing the measures developed in (3.7) and (3.8) we obtain the following.

TABLE 3.2: Example - Efficiency profile η_i and proportions ζ_{TE_i} and ζ_{AE_i}

Measure	Unit	DSO A	DSO B	DSO C
Cost efficiency	%	66.5	66.5	66.5
Technical efficiency	%	70.0	95.0	81.5
Allocative efficiency	%	95.0	70.0	81.5
ζ_{TE_i}	%	87.4	12.6	50.0
ζ_{AE_i}	%	12.6	87.4	50.0

We can see from the above example that 87.4% of DSO A's excess costs (i.e. cost inefficiency) are due to managerial sub-performance (i.e. technical inefficiency) and 12.6% are due to a sub-optimal mix of inputs (i.e. allocative inefficiency)—the opposite is true for DSO B. DSO C's cost inefficiency can be equally attributed to technical and allocative inefficiency.

In Section 3.1 we interpreted the implications of technical and allocative inefficiency in the context of network regulation. From the insights of the regulatory literature and the model specifications of European regulators, it seems that the allocative efficiency of DSOs is not as controllable as their technical efficiency¹⁵. We thus assume the following.

Hypothesis 2. For distribution system operators, improving cost efficiency through a reallocation of inputs (i.e. reduction of allocative inefficiency) is more complex and less controllable than improving cost efficiency through a more efficient use of those inputs (i.e. reduction of technical inefficiency).

The rationale behind Hypothesis 2 is that managerial efficiency is easier to improve than input allocation efficiency. From Hypothesis 2 we can state the following.

Hypothesis 3. Controlling for cost efficiency, the lower a distribution system operator's allocative efficiency is (i.e. the higher its allocative inefficiency) the lower its subsequent catch-up speed will be (i.e. the harder it is to improve cost efficiency).

¹⁴It also true that $\lim_{x \rightarrow 0^+} \log(x) = -\infty$, but this is not an issue as the efficiency scores of benchmarking have non-zero values.

¹⁵“The regulator may specify that it is only interested in TE comparisons across firms (it may perhaps believe that scale and AE are not controllable in the short run).” (Coelli (2003))

If Hypothesis 3 holds¹⁶ we can propose the following function $h(\cdot)$ that combines measures ζ_{TE_i} and ζ_{AE_i} .

$$\xi_i = h(\eta_i) = \zeta_{AE_i} - \zeta_{TE_i} \quad (3.9)$$

$$= \frac{\log(AE_i) - \log(TE_i)}{\log(CE_i)} \in [-1; 1] \quad \forall CE_i \in]0; 1[\quad (3.10)$$

This complexity factor ξ_i ranges from -1 to 1. At one extreme, a value of -1 indicates that firm i is allocatively efficient and that catching-up to the frontier would mean improving its use of inputs (i.e. reducing technical inefficiency). At the other extreme, a value of 1 indicates that the firm is technically efficient and that it would have to reallocate its inputs in order to reduce its costs (i.e. reducing allocative inefficiency). Between those two extremes a value of 0 indicates that firm i 's cost inefficiency is due as much to technical as it is to allocative inefficiency. Just like ζ_{TE_i} and ζ_{AE_i} , ξ_i is only defined if firm i is cost inefficient (i.e. $CE_i < 1$).

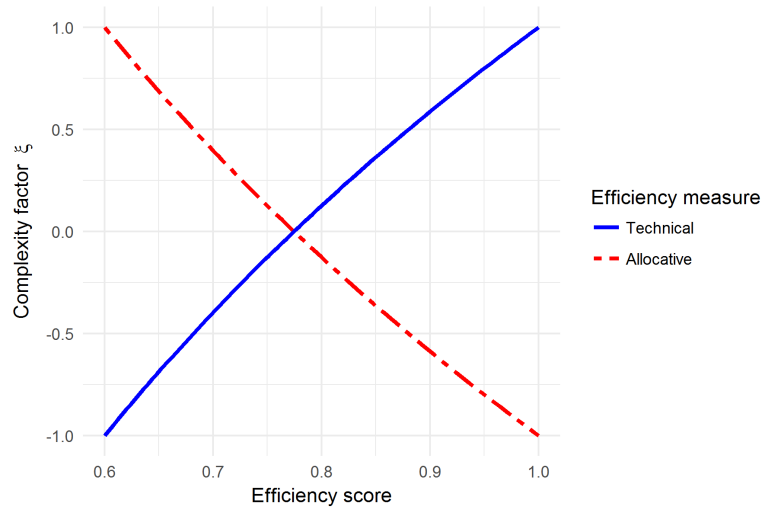


FIGURE 3.1: Complexity ξ vs. efficiency profile η (cost efficiency of 60%)

Figure 3.1 illustrates the relationship between the complexity factor ξ and the efficiency profile η for a firm at a cost efficiency of 60%. The complexity ξ is positively correlated with technical efficiency and negatively correlated with allocative efficiency. The factor ξ has a value of zero when both technical and allocative efficiency have a value of $\sqrt{CE}$ (i.e. $\sqrt{0.6} = 0.77$)—since cost efficiency is equally distributed between technical and allocative efficiency.

The function $h(\cdot)$ defined in (3.9) assumes that allocative inefficiency is more complex than technical efficiency (i.e. assumes that Hypothesis 3 holds). This function should

¹⁶Hypothesis 3 will be empirically demonstrated in Section 5.3.1.

be defined according to the industry under regulation. The regulator should define $h(\cdot)$ such that a higher ξ_i (i.e. higher catch-up complexity) is negatively correlated with the cost efficiency change observed in the regulated industry. Other functions $h(\cdot)$ can be developed to translate a firm's efficiency profile η_i into its complexity ξ_i .

3.3 Firm-specific catch-up period T_i

As we have seen in Section 2.5—when setting the catch-up factor χ_i —most fixed-price regimes are relying on computing the cost efficiency of each DSO and granting each of them the same number of regulatory periods—or years—to reach the frontier.

More generally regulators are computing the χ_i requirement as a function of the incumbent efficiency of firm i (i.e. $CE_i(0)$) and the number of years T given to eliminate it. Let us assume $C_i(0)$ and C_{opt} are, respectively, the realised cost of DSO i and the optimal cost of the best practice of the industry at the outset of the regulation.

Assuming no technological development during the regulatory period (i.e. C_{opt} is constant throughout the regulatory period and χ can be left out), from (2.39) we compute the following equation.

$$C_{opt} = C_i(0)(1 - \chi_i)^T \quad (3.11)$$

Since we know that the cost efficiency of firm i is the ratio of the minimal cost and its realised cost (i.e. $CE_i(0) = C_{opt}/C_i(0)$), we have

$$CE_i(0) = (1 - \chi_i)^T \quad (3.12)$$

$$\chi_i = 1 - CE_i(0)^{1/T} \quad (3.13)$$

As explained in Section 2.5, the primary limitation of this approach lies in parameter T which is not firm-specific. We propose to define a firm-specific catch-up period T_i defined as

$$T_i = \Gamma(\xi_i) \quad (3.14)$$

where $\Gamma(\cdot)$ is a function defined by the regulator to compute T_i from the complexity factor ξ_i . We propose the following linear function.

$$T_i = T_0 + \alpha\xi_i \in [T_0 - \alpha; T_0 + \alpha] \quad (3.15)$$

where T_0 is the *default catch-up period* and α a *sensitivity parameter* for complexity.

The parameter T_0 is the catch-up period granted to a firm which cost efficiency is perfectly distributed between technical and allocative efficiency (i.e. $\xi_i = 0$). The parameter

α reflects the relative difficulty of catching-up allocative inefficiency compared to technical inefficiency in the regulated industry. This second parameter will determine of how many years T_i will increase or decrease according to complexity. In order to mitigate arbitrariness in the choice of T_0 and α , their values should rely as much as possible on empirical data from the regulated industry. Like for function $h(\cdot)$ the regulator can define another function $\Gamma(\cdot)$ ¹⁷ to compute T_i from ξ_i .

Now that we have developed a firm-specific T_i , we can use (3.13) and (3.15) to define the individual improvement requirement χ_i .

$$\chi_i = 1 - CE_i(0) \exp(1/(T_0 + \alpha\xi_i)) \quad (3.16)$$

Even though the catch-up speed is now firm-specific the regulatory period of T years will still be the same for all firms. Therefore if $T_i < T$ the requirement χ_i should be zero after T_i years and for the remainder $(T - T_i)$ years of regulation. This is to prevent the revenue cap from expecting further cost reduction from firms that have already reached efficiency.

To prevent regulators from arbitrarily setting parameters T_0 and α we propose guidelines to base their value on empirical data.

3.3.1 Default catch-up period T_0

Definition 2. The default catch-up period T_0 is the number of years granted to a firm for eliminating cost inefficiency if this firm's cost efficiency is equally distributed between its technical and allocative components.

Two approaches are possible for setting parameter T_0 . The first one is to determine its value based on previous empirical studies from similar industries operating under similar conditions. Regulators can chose studies done on the same industry from the same country but at a different point in time (e.g. several years or decades ago). Doing so would make the estimate of T_0 relevant in terms of time-invariant exogenous conditions specific to the geographic area (i.e. the country). Nevertheless, by determining its estimate of T_0 based on an older study the regulator is neglecting time-variant conditions that might have influenced the efficiency change at that time and that have changed ever since. For instance, if different regulatory schemes have been implemented throughout the years those might have had various level of incentive power in their regulation, therefore influencing the catch-up of the set of regulated entities. Different studies may

¹⁷Instead of being linear $\Gamma(\cdot)$ could for instance be quadratic. $\Gamma(\cdot)$ must be continuous, strictly positive, and monotonously increasing in ξ_i .

also have been using different benchmarking methods, model specifications, or sets of variables in their analyses therefore influencing efficiency estimates¹⁸. Those are but a fraction of the factors that may cause discrepancies in the results of those studies.

The second approach relies on using more recent historical observations of efficiency change for the same industry. The regulator would then be using the average catch-up that the set of firms under its regulation demonstrated in the previous years as a proxy for average improvement potential. The main advantage of this second approach is that the set of operators will most likely be identical to the one about to be regulated—leaving aside potential mergers and bankruptcies. This means that the conditions under which the industry will be operating in the coming years will most likely be close to those it has been operating under in the previous years—since the time difference is moderate. The benchmarking tools will most likely be consistent as the regulator can use the same benchmarking approach he developed to compute current efficiency scores. There is nevertheless one drawback: the ratchet effect. As mentioned in Section 2.3, if the firms are “aware” that their current performance will determine their future requirements they will most likely hide their true potential by fear of being punished in a subsequent regulation. Even though this is true at the individual level—or firm level—we can assume that by averaging the past performances across the whole industry the ratchet effect will be mitigated enough so that it will not be an issue for the incentive power of the regulation¹⁹.

Regardless of the chosen approach, the parameter T_0 will be computed as follows. Suppose the regulator found an average yearly cost efficiency change (i.e. catch-up) of $\overline{CEC}(t, t+1)$ with an average initial efficiency score of $\overline{CE}(0)$. We want to find out how many years T a firm with an incumbent efficiency of $\overline{CE}(0)$ will need to reach cost efficiency if its yearly catch-up is $\overline{CEC}(t, t+1)$. Since the catch-up rate is a circular index²⁰ we have

$$\overline{CE}(0) \times \overline{CEC}(t, t+1)^T = 1 \quad (3.17)$$

After rearranging (3.17) to isolate T , we found

$$T = \frac{-\ln(\overline{CE}(0))}{\ln(\overline{CEC}(t, t+1))} \quad (3.18)$$

¹⁸See Section 2.2.5.

¹⁹This assumption is reasonable as long as the set of operators is not too small. If it is not large enough there is a risk of collusion.

²⁰See Section 2.2.2.3.

Example 3. Assume an initial mean efficiency of 78.5% (i.e. $\overline{CE}(0) = 0.785$) and yearly catch-up of 1.021 (i.e. $\overline{CEC}(t, t+1) = 1.021$).

$$T = \frac{-\ln(0.785)}{\ln(1.021)} = 11.65 \quad (3.19)$$

A theoretical firm demonstrating those mean values should reach efficiency in 11.65 years if it keeps the same catch-up speed.

Using empirical efficiency changes and scores, the regulator can therefore use the above equation to empirically set parameter T_0 ²¹.

3.3.2 Sensitivity parameter α

Definition 3. The sensitivity parameter α is the number of additional years a firm at maximal complexity (i.e. $\xi_i = 1$) is granted compared to a firm at zero complexity (i.e. $\xi_i = 0$).

The parameter α can also be defined as the number of additional years a firm at zero complexity is given compared to one at minimal complexity (i.e. $\xi_i = -1$). In other words, it reflects the extent to which the regulator will differentiate between firms according to their complexity when setting the catch-up period T_i . Parameter α should therefore reflect “how complicated” the regulator assumes reallocating inputs is compared to reducing them. Because the catch-up period must be greater than zero (i.e. $T_i > 0$), α has to be less than T_0 (i.e. $\alpha < T_0$).

Compared to T_0 , computing α from historical data turns out to be much more complicated. Developing a model to set this parameter based on empirical data is beyond the scope of this thesis as more extensive econometric analyses on the relationship between cost efficiency change and allocative efficiency are required. One thing that is certain is that α and T_0 should be set accordingly.

Example 4. Considering once again Example 1 and using the formulas developed in (3.9), (3.15), and (3.13), we can compute the complexity ξ_i , catch-up period T_i , and catch-up requirement χ_i for DSOs A, B, and C. We assume $T_0 = 11.65$ as computed in (3.19) and a sensitivity α of 3.

$$T_i = T_0 + \alpha\xi_i = 11.65 + 3\xi_i \in [8.65; 14.65] \quad (3.20)$$

²¹Since the regulator can choose the source of the empirical data used in computing T_0 there is still a form of arbitrariness in its computation.

$$\chi_i = 1 - CE_i \exp(1/(11.65 + 3\xi_i)) \quad (3.21)$$

TABLE 3.3: Example - Complexity ξ_i , catch-up period T_i , and catch-up requirement χ_i

Measure	Unit	DSO A	DSO B	DSO C
Cost efficiency	%	66.5	66.5	66.5
Technical efficiency	%	70.0	95.0	81.5
Allocative efficiency	%	95.0	70.0	81.5
Complexity ξ_i	N/A	-0.74	0.74	0.00
Catch-up period T_i	years	9.43	13.87	11.65
Catch-up requirement χ_i	%	4.23	2.89	3.44

Because of their different efficiency profiles η_i , DSO B is granted 4.44 additional years to reach the frontier compared to DSO A. With a complexity factor of zero, DSO C has the default catch-up period T_0 to become cost efficient.

3.4 Discussion

The model we developed through this chapter offers a detailed approach to cost inefficiency assessment to make the computation of the catch-up requirement χ_i more accurate. By using the benchmarking measures of technical and allocative efficiency the regulator is provided with a more detailed analysis of a firm's inefficiency. With the use of the regulated firms' efficiency profiles η_i and complexity factors ξ_i , the regulator can differentiate between firms when setting the firm-specific catch-up period T_i and requirement χ_i . By understanding why a firm is cost inefficient and which strategies are available for it to improve, regulatory authorities can make the improvement requirement of fixed-price regulations fairer in terms of controllability. By estimating the period T_0 based on empirical data and using α to differentiate between inefficient operators the computation of the required catch-up χ_i is given economical sense and foundation.

Nonetheless, this approach has several limitations that should not be overlooked. First, it implicitly assumes that a firm is not in control of its efficiency profile η_i . Because the catch-up requirement χ_i is determined by the complexity factor ξ_i which is in turn determined by the distribution of the firm's cost efficiency into its technical and allocative components—or efficiency profile η_i —there may be a risk of strategic behaviour that would lower the incentive power of the regulation. An operator might make poor input allocation decisions on purpose in order to get further away from the optimal input-mix. Doing so would then increase its complexity score ξ_i and catch-up period T_i ,

thus reducing its required catch-up speed. But this situation is only possible if the firm manages to lower its allocative efficiency without decreasing its technical efficiency. This situation seems highly unlikely since making poor expenditure and investment decisions will most likely lower both technical and allocative efficiencies, keeping the complexity score rather stable²².

Second, the model fails to completely solve the issue of arbitrariness. Because we want the computation of χ_i to rely on empirical observations as much as possible, we proposed in Sections 3.3.1 and 3.3.2 guidelines to set parameters T_0 and α according to empirical data. Nevertheless, the regulator has a great control on the approach and set of data used in estimating those parameters.

Third, we assumed that technical inefficiency is easier to remove relative to allocative inefficiency. This assumption essentially originates from current regulatory approaches relying exclusively on pure technical analyses and the theoretical definition of technical and allocative measures along with their practical implications. Even though this assumption empirically holds for network utilities such as distribution system operators²³, it may not hold in other industries in need of regulation. Regulators should thus define the complexity factor ξ_i (i.e. function $h(\cdot)$) according to the profile of the regulated industry.

Fourth, the benchmarking model specification and its set of inputs has a significant impact on the calculated allocative efficiency scores. As the number of inputs considered in the benchmarking model increases, allocative efficiency decreases thus impacting the efficiency profile η_i . Therefore functions $h(\cdot)$ and $\Gamma(\cdot)$ should be defined according to the model specification and set of inputs.

²²Or at least out of the control of the regulated firm.

²³See Section 5.3.1.

Chapter 4

Data

In this chapter we are briefly introducing the set of data that is used in our analysis. The set of variables used in our DEA model is presented and its descriptive statistics are provided.

4.1 Dataset

The dataset consists of detailed information on the costs and assets of Swedish distribution system operators for the period 2000-2006. This panel has been consolidated by the regulator STEM.

Due to mergers that took place between 2000 and 2006—or data that is missing altogether—the panel is unbalanced. Since we need to analyse the efficiency changes of each individual DSO we need to have the exact same set of firms from 2000 through 2006 (i.e. a balanced panel). We retroactively created the merged entities prior to their actual merger event by consolidating their data in the panel and removed operators with missing data. The balanced panel consists of 143 DSOs.

Starting in 2003 the Swedish regulator established a new regulatory scheme called the “Network Performance Assessment Model” (NPAM). Unlike many regulatory tools implemented in Europe this scheme did not rely on a benchmarking assessment of efficiency. The regulation was *ex post* and used data on consumers and revenues. Due to its limited information value, bad documentation, lack of investment incentives, and complex application the NPAM failed as an incentive regulation. Even though its approach was generally superior to most econometric models in use at the time, many operators undertook administrative actions and the regulator backed down. NPAM was replaced in 2012 by an interim cost-recovery regime. Because of its lack of incentive power, the

NPAM experience caused the industry to demonstrate poor productivity change during this period. (Agrell and Grifell-Tatjé (2016))

4.2 Model variables

Like we explained in Section 2.2.5, properly choosing the set of variables used in benchmarking is crucial as it will have a significant impact on the computed efficiency scores. After assessing the most widely used variables in network benchmarking studies¹ and the data available we chose the following inputs and outputs.

4.2.1 Outputs

Unlike firms operating in a competitive market, DSOs are not in control of the level of output they generate. We consider two cost drivers.

Output 1. The **peak load** is the highest energy requirement of the distribution system during a given period of time—a year in our case. Fluctuations in peak demand of electricity varies on a daily basis as well as throughout the year.

Output 2. The **number of connections** is the number of exit points the distribution grid is providing with electricity. It primarily depends on the size of the market and its density.

4.2.2 Inputs

Three controllable inputs are considered in our model.

Input 1. The **energy losses** is the difference between the energy transmitted through the grid and the energy actually delivered to its exit points—losses typically amount to 4-5% of the transmitted energy. As explained in Navani et al. (2012) those energy losses can either be technical or non-technical. Technical losses are caused by the physical properties of the distribution grid. Part of the power will inevitably dissipate in some of the system's components such as the lines, cables, transformers, or substations. Even though those losses are occurring naturally and cannot be completely eliminated, they can nevertheless be observed and controlled. Non-technical losses—on the other hand—are the result of external actions. Some common examples are energy theft and metering inaccuracies. Although those non-technical losses are harder to assess than the technical ones, they can be prevented.

¹See Jamasb and Pollitt (2000) and Jamasb and Pollitt (2003) for more details on the inputs and outputs most commonly used in the benchmarking of network utilities.

Input 2. The **weighted network length** represents the level of inputs associated with operating expenditure. The operating expenditure—or OPEX—includes the costs incurred for operating and maintaining the distribution network. Along with other costs, OPEX is composed of the costs of metering and personnel.

$$\text{OPEX} = \text{Cost of metering} + \text{Cost of personnel costs} + \text{Other costs} \quad (4.1)$$

The weighted network length (WNL) is computed as the weighted sum of the lengths of the grid's low voltage lines (LVL), high voltage lines (HVL), low voltage cables (LVC), and high voltage cables (HVC).

$$\text{WNL}_i = \beta_1^* \text{LVL}_i + \beta_2^* \text{HVL}_i + \beta_3^* \text{LVC}_i + \beta_4^* \text{HVC}_i, \quad i = 1, \dots, n \quad (4.2)$$

where weights β_1^* , β_2^* , β_3^* , and β_4^* are the fitted coefficients of the following multivariate regression.

$$\text{OPEX}_i = \beta_1 \text{LVL}_i + \beta_2 \text{HVL}_i + \beta_3 \text{LVC}_i + \beta_4 \text{HVC}_i, \quad i = 1, \dots, n \quad (4.3)$$

TABLE 4.1: Weighted network length OLS results

Variable	Coefficient β_i^*	t -statistic	p-value
LVL	27.751	3.140	0.00173
HVL	0.612	0.099	0.92140
LVC	-12.521	-2.307	0.02122
HVC	140.814	10.493	< 2.2e−16

Input 3. The **network capital** represents the distribution assets and equipment of the operator. The cost associated with network capital is the capital expenditure—or CAPEX—and is here calculated as the sum of the costs of finance and depreciation.

$$\text{CAPEX} = \text{Cost of finance} + \text{Cost of depreciation} \quad (4.4)$$

4.2.3 Non-discretionary variables

As presented in Section 2.2.4.2, z-variables are included in the DEA model to account for the exogenous conditions faced by DSOs. This allows for a fairer comparison and more relevant productivity assessment. We propose two z-variables to account for the profile of an operator's market.

Z-variable 1. The **line proportion** is the proportion of lines in an operator's distribution network—which consists of overhead lines and underground cables. Lines are installed overhead on transmission towers and are easier and cheaper to install. For those reasons they are primarily used in rural areas where the distance that needs to be covered is greater and infrastructure less present. Cables are built underground and more expensive and tedious to maintain. They leave more room for infrastructure, do not decrease property value, and result in reduced electromagnetic fields (EMF) emissions. Those are mainly used in urban areas. Assuming that overhead lines are exclusively used in rural setting while underground cables are exclusively used in urban setting, we can use the distribution of lines and cables length in a DSO's total network length to assess the extent to which the firm is providing a rural or urban market. We propose the following variable.

$$\text{Line proportion (\%)} = \frac{\text{Line length (km)}}{\text{Network length (km)}} \quad (4.5)$$

where the line length (LL) is the sum of low (LVL) and high voltage line length (HVL) (i.e. $LL = LVL + HVL$) and the network length (NL) is the sum of line and cable length at low and high voltage (i.e. $NL = LVL + HVL + LVC + HVC$). The line proportion provides us with a proxy to assess the extent to which a firm is operating in a rural environment. A smaller value indicates that the environment is primarily urban while a greater value indicates a rural setting.

Z-variable 2. The **high voltage proportion** is the proportion of high voltage lines and cables in an operator's distribution network. We can distinguish between circuit for high voltage (HV) electricity—used to provide industrial customers—and for low voltage (LV) electricity—used to provide households and other smaller customers². Assuming LV electricity is exclusively distributed to households and HV electricity is exclusively distributed to industrial consumers, we can use the distribution of an operator's grid between HV and LV circuit length to assess the profile of its market.

$$\text{High voltage proportion (\%)} = \frac{\text{High voltage network length (km)}}{\text{Network length (km)}} \quad (4.6)$$

where the high voltage network length (HVNL) is the sum of high voltage line (HVL) and cable length (HVC) (i.e. $HVNL = HVL + HVC$). The high voltage proportion allows us to assess the extent to which an operator is providing industrial consumers. A smaller value indicates that the users are mainly households while a greater value indicates they are mainly industrial companies.

²High voltage typically ranges from 1 to 35 kV. Low voltage is below 1 kv.

TABLE 4.2: Descriptive statistics of the model variables

Category	Unit	Variable definition	Mean	Median	SD
Outputs y	MWh	Peak load	155.40	56.00	576.56
	c	Connections	35,028.83	12,356.00	102,354.50
Inputs x	MWh	Energy losses	30,252.88	11,002.00	81,897.23
	km	Weighted network length	62,631.25	20,580.69	198,823.90
Input prices w	kSEK	Network capital	422,243.20	96,288.00	1,444,795.00
	kSEK/MWh	Cost per energy losses	0.262	0.256	0.118
	kSEK/km	Cost per connection-line meter	1.595	1.310	0.940
	%	Cost of capital	8.91	8.33	4.24
Costs wx	kSEK	Cost of energy losses	7,387.75	2,685.00	20,361.00
	kSEK	Operating expenditure (OPEX)	76,869.62	30,716.00	221,169.90
	kSEK	Capital expenditure (CAPEX)	28,587.57	7,898.00	97,736.35
Z-variables	%	Line proportion	32.74	32.02	21.02
	%	High voltage proportion	33.65	33.23	8.51

Notes: c = connection, km = 1000 m, kSEK = 1000 Swedish crowns (SEK), SD = standard deviation.

Chapter 5

Analysis

In this chapter we apply the approach and model proposed in Chapter 3 to the data presented in Chapter 4. The significance of the proposed non-controllable variables is empirically demonstrated. The benchmarking exercise is performed to compute the measures of cost, technical, and allocative efficiency scores and changes. The efficiency profile η of the different operators is assessed and analysed. Hypothesis 3 is empirically demonstrated through a fixed-effects multivariate regression analysis before applying our model by computing the complexity factor ξ . Finally we use the efficiency scores from 2006 to simulate the catch-up period T_i and requirement χ_i of a fictional revenue cap starting in 2007.

5.1 Significance of non-discretionary variables

In order to assess the significance of the z-variables proposed in Section 4.2.3, we computed the operators' efficiency scores via a DEA model not including them. Table 5.1 provides the average descriptive statistics for the 143 DSOs through period 2000-2005.

TABLE 5.1: Efficiency scores without z-variables (2000-2005)

Measure	Unit	Min	Mean	Median	Max	SD
Cost efficiency	%	18.75	57.52	57.24	100.00	18.35
Technical efficiency	%	27.16	75.08	75.05	100.00	17.96
Allocative efficiency	%	28.80	76.64	78.76	100.00	15.35

Notes: SD = standard deviation.

To assess the significance of our z -variables we performed two of the following univariate linear regression—one for each z -variable.

$$CE_i = \beta_0 + \beta_1 Z_i + \varepsilon_i, \quad i = 1, \dots, n \quad (5.1)$$

where CE_i is the cost efficiency, Z_i the z -variable which significance is tested—either “line proportion” or “high voltage proportion”, β_0 the intercept, β_1 the regression coefficient of Z_i , and ε_i an idiosyncratic error term.

TABLE 5.2: Analysis of the z -variables significance

Variable Z_i	Intercept β_0	Coefficient β_1	Coefficient t -statistic
Line proportion	0.732	−0.471	−19.04***
High voltage proportion	0.865	−0.863	−12.91***

Notes: *** p -value $< 2.2\text{e}^{-16}$.

Figures A.1 and A.2 (see Appendix A) display the plots of cost efficient against our two non-controllable variables along with the computed linear regression. When performing the Student’s t -test for coefficient β_1 significance both demonstrate a p -value below 5%, therefore proving there is a significant linear relationship between our two z -variables and cost efficiency. Both z -variables are negatively correlated with cost efficiency. Recall from Section 2.2.4.2 that the non-controllable inputs must be negatively correlated with efficiency as the DEA model will try to reduce them.

5.2 Benchmarking results

Now that we have proven the significance of our z -variables we can include them in our DEA model and recompute efficiency.

5.2.1 Efficiency scores

Table 5.3 gives once again the average descriptive statistics for the period 2000-2005, this time including non-controllable variables.

TABLE 5.3: Efficiency scores with z-variables (2000-2005)

Measure	Unit	Min	Mean	Median	Max	SD	CV
Cost efficiency	%	18.75	60.50	58.53	100.00	20.26	0.334
Technical efficiency	%	27.16	77.75	79.06	100.00	18.75	0.241
Allocative efficiency	%	28.80	77.71	80.24	100.00	16.10	0.207

Notes: SD = standard deviation, CV = coefficient of variation.

Those descriptive statistics summarise the efficiency measures for 143 operators through six years (i.e. from 2000 to 2005), thus amounting to 858 observations. The estimated cost efficiency ranges between 18.75 and 100%, demonstrating a substantial variation in efficiency among the DSOs and across the years. The mean cost efficiency is 60.50% which means that—on average between 2000 and 2005—operators could have saved 39.50% of their total costs by performing like the best practices (i.e. the frontier).

The mean technical efficiency is quite close to the mean allocative efficiency, indicating that cost inefficiency seems due to technical efficiency as much as it is due to allocative inefficiency. Nonetheless, technical efficiency demonstrates a higher coefficient of variation (i.e. 24.1%) than allocative efficiency (i.e. 20.7%). This indicates that allocative scores display a higher dispersion than technical scores. Figure 5.1 shows the density plots of the cost, technical, and allocative efficiency scores. Those density plots help us visualise how efficiency scores are distributed for each of the three measures. Unlike allocative scores, technical scores seem much more concentrated towards higher efficiency (i.e. 90-100%). Table 5.4 and Figure 5.2 display the mean yearly cost efficiency score and its decomposition for the period 2000-2005.

TABLE 5.4: Mean efficiency scores by year

Measure	Unit	2000	2001	2002	2003	2004	2005
Cost efficiency	%	53.48	53.63	63.74	63.76	65.22	63.15
Technical efficiency	%	75.39	77.51	78.48	76.92	77.61	80.58
Allocative efficiency	%	70.84	69.05	80.92	83.25	84.32	77.87

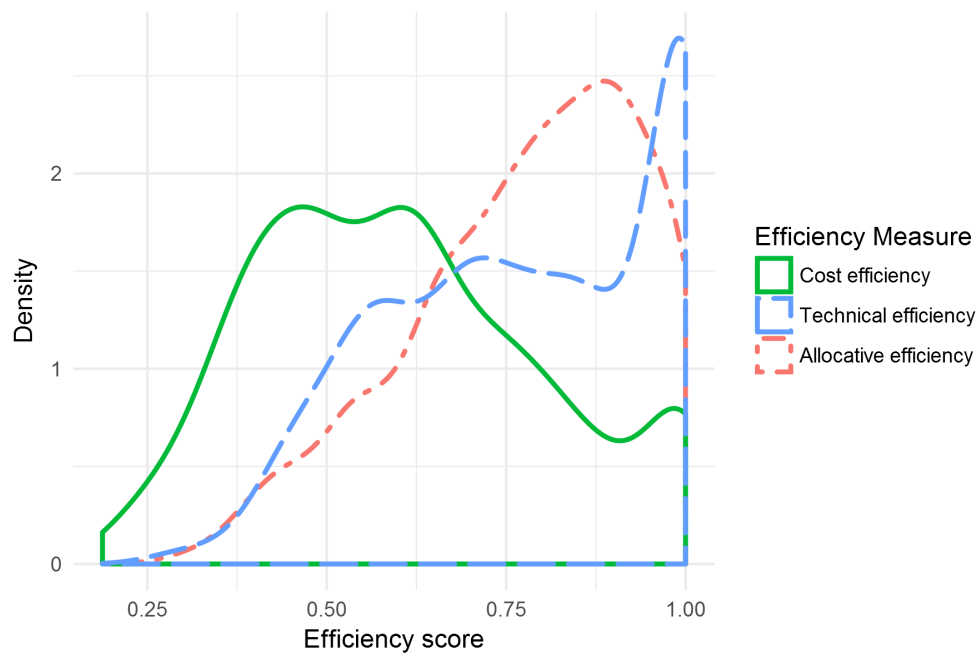


FIGURE 5.1: Density of efficiency scores (2000-2005)

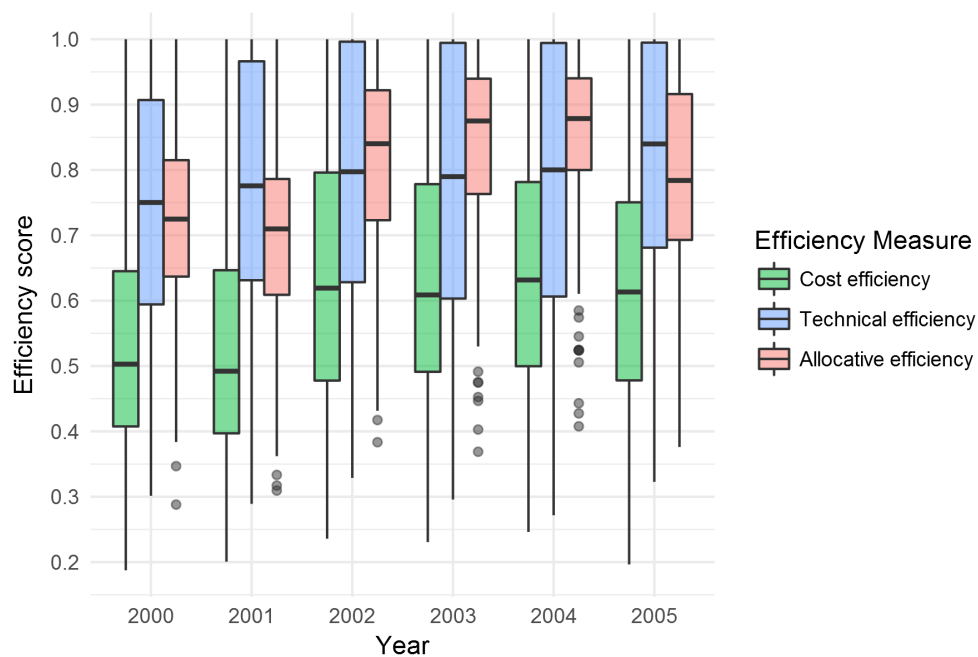


FIGURE 5.2: Efficiency scores by year (2000-2005)

Figure 5.3 shows us that—as expected—the cost efficiency scores computed with the DEA model considering z -variables are higher. As we explain in Section 2.2.4.2, this is due to the fact that the frontier model addresses non-discretionary factors to allow for a fairer and more comprehensive benchmarking assessment.

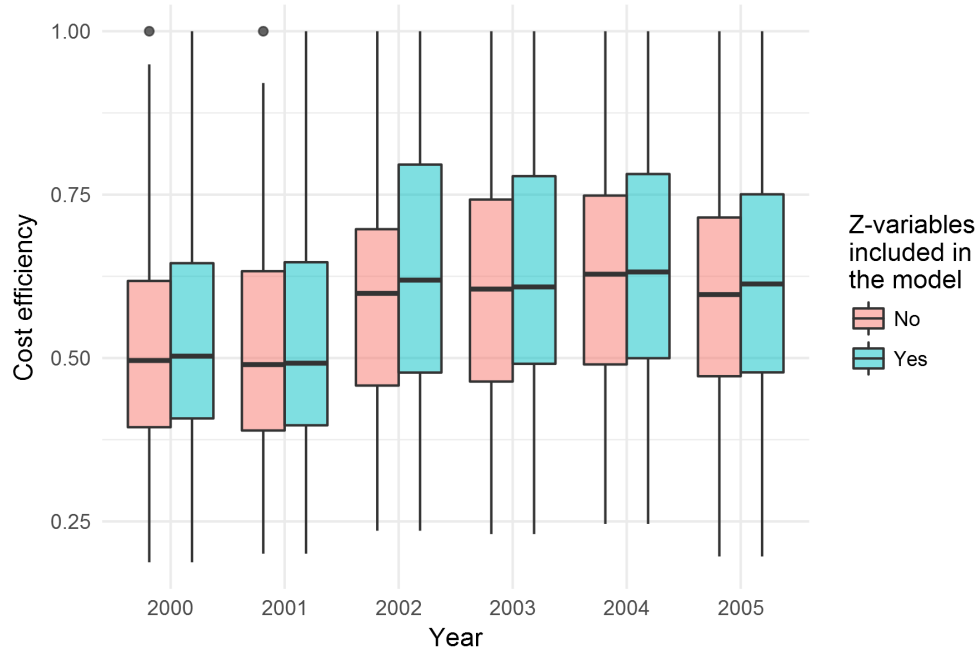


FIGURE 5.3: Cost efficiency by year without and with z-variables

5.2.2 Efficiency profile η

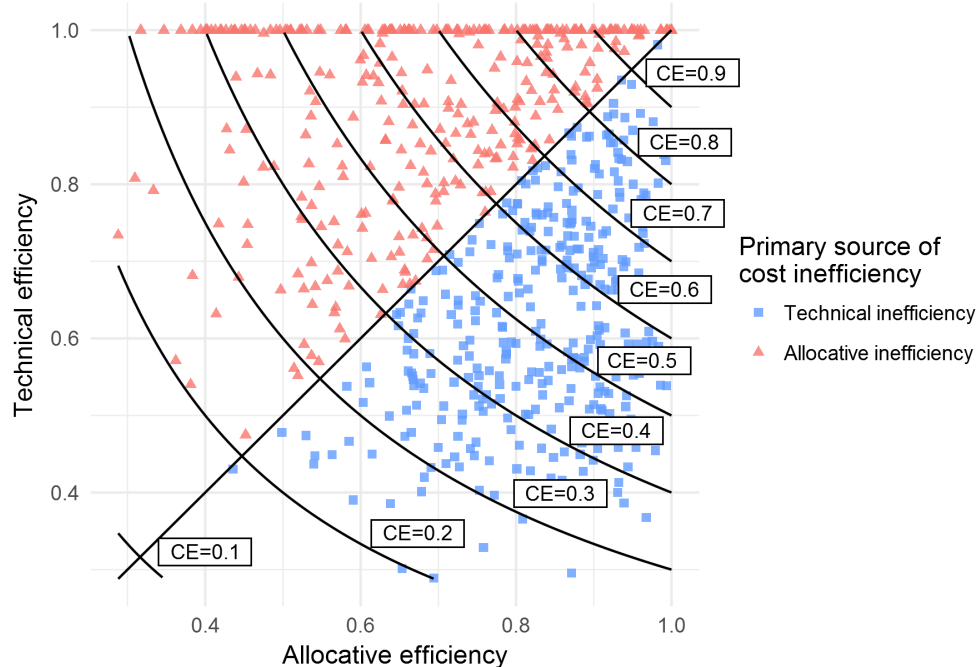
Let us now take a closer look at the distribution of cost efficiency between its technical and allocative components for our sample.

Figure 5.4 displays the plot of technical efficiency against allocative efficiency. The “*iso-cost-efficiency*” curves represent the possible efficiency profiles η —or technical-allocative combinations—for different scores of cost efficiency. By looking at the efficiency profile we can assess whether the primary source of excess cost is technical or allocative inefficiency. Out of the 858 observed efficiency profiles, the primary source of inefficiency was technical for 407 of them (i.e. 47.43%) and allocative for 451 of them (i.e. 52.56%).

Example 5. Let us consider operators “SEVAB Nät AB” and “Ystad Energi AB”. In 2004 those DSOs demonstrated the following efficiency profiles.

TABLE 5.5: Example - Efficiency profile η of operators SEVAB Nät AB and Ystad Energi AB (2004)

Measure	Unit	SEVAB Nät AB	Ystad Energi AB
Cost efficiency	%	53.49	52.44
Technical efficiency	%	53.83	100.00
Allocative efficiency	%	99.36	52.44
ζ_{TE_i}	%	98.98	0.00
ζ_{AE_i}	%	1.02	100.00

FIGURE 5.4: Efficiency profile η

Only considering cost efficiency, SEVAB Nät AB and Ystad Energi AB are nearly identical with a small difference of 1.05% (i.e. both their efficiency profiles are near the iso-cost-efficiency curve for $CE=0.53$). Nonetheless, an assessment of their efficiency profiles reveals that the source of their excess cost—and thus the strategies needed to eliminate it—are substantially different. After computing measures ζ_{TE} and ζ_{AE} we see that the cost inefficiency of SEVAB Nät AB is due at 98.98% to technical inefficiency while the cost inefficiency of Ystad Energi AB is exclusively due to allocative inefficiency. While SEVAB Nät AB needs to demonstrate improved managerial effort to catch-up to the frontier, Ystad Energi AB needs to implement resources reallocation strategies.

5.2.3 Efficiency changes

Table 5.6 summarises the descriptive statistics for efficiency changes between 2000 and 2005.

TABLE 5.6: Efficiency changes (2000-2005)

Measure	Min	Mean	Median	Max	SD	CV
Cost efficiency	0.653	1.045	1.000	2.724	0.168	0.160
Technical efficiency	0.524	1.024	1.000	2.481	0.153	0.150
Allocative efficiency	0.531	1.032	1.000	1.839	0.171	0.165

Notes: SD = standard deviation, CV = coefficient of variation.

The cost efficiency change ranges from 0.653 (i.e. cost efficiency reduced by 34.7%) to 2.724 (i.e. cost efficiency improved by 172.4%), demonstrating a high dispersion. During the 2000-2005 period operators showed—on average—a cost efficiency catch-up of 1.045, corresponding to an improvement in cost efficiency of 4.5%. Allocative efficiency change has a higher dispersion than technical efficiency change with coefficients of variation of 16.5% and 15.0%, respectively.

Table 5.7 shows the mean yearly cost efficiency change and its decomposition.

TABLE 5.7: Mean efficiency change by year

Measure	2000/2001	2001/2002	2002/2003	2003/2004	2004/2005
Cost efficiency	1.007	1.213	1.008	1.029	0.968
Technical efficiency	1.036	1.024	0.985	1.017	1.060
Allocative efficiency	0.979	1.194	1.041	1.023	0.924

Except for the change between 2004 and 2005, operators generally demonstrated a moderate but positive cost efficiency catch-up. The catch-up from 2001 to 2002 was particularly higher with a value of 1.213. Figure 5.5 displays the histogram for cost efficiency changes for the period.

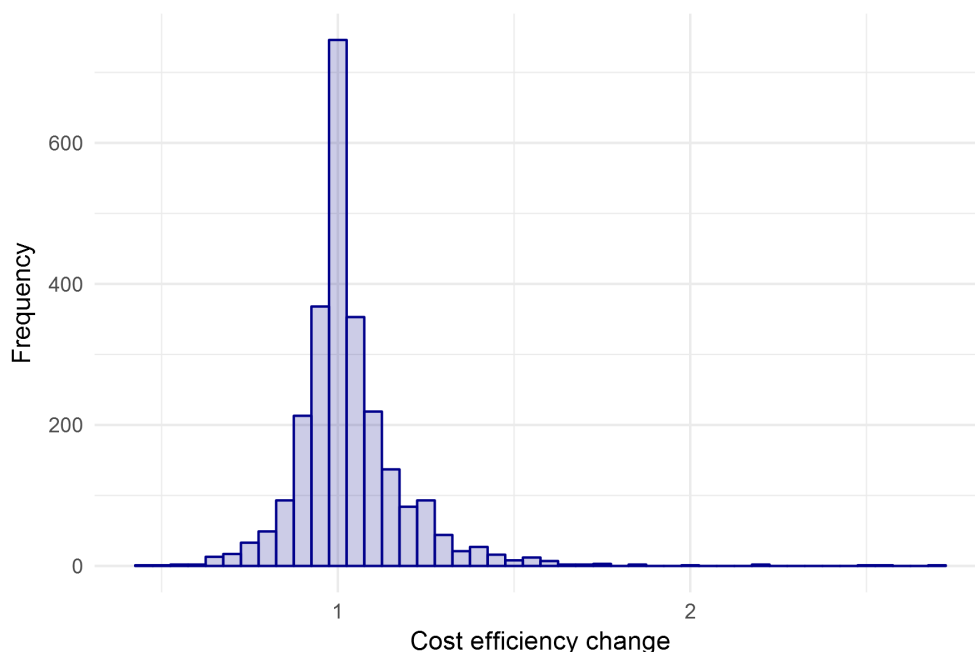


FIGURE 5.5: Histogram for cost efficiency change

5.3 Model application

Now that we have computed the efficiency measures for our set of operators with the benchmarking tools presented in Section 2.2, we can apply the model proposed in Chapter 3. We are empirically testing Hypothesis 3 before computing and analysing the complexity factor ξ .

5.3.1 Hypothesis

Recall that Hypothesis 3 proposed in Section 3.2 assumed that:

“Controlling for cost efficiency, the lower a distribution system operator’s allocative efficiency is (i.e. the higher its allocative inefficiency) the lower its subsequent catch-up speed will be (i.e. the harder it is to improve cost efficiency).”

To see if our hypothesis empirically holds we need to test the relationship between a firm’s allocative efficiency score and its subsequent cost efficiency change, while controlling for the effect of cost efficiency. Recall from Section 2.2.2.3 that the cost efficiency change of firm i (i.e. $CEC_i(t, t+j)^1$) is an index number measuring the change in its cost efficiency between periods t (i.e. $CE_i(t)$) and $t+j$ (i.e. $CE_i(t+j)$). We are here considering a lag of one period (i.e. $j = 1$). Decomposing CE_i we have the following.

$$CEC_i(t, t+1) = \frac{CE_i(t+1)}{CE_i(t)} = \frac{TE_i(t+1) \times AE_i(t+1)}{TE_i(t) \times AE_i(t)} \quad (5.2)$$

Hence, by definition, the base cost efficiency score $CE_i(t)$ —along with its components—is negatively correlated with the catch-up $CEC_i(t, t+1)$. Since $AE_i(t)$ is positively correlated with $CE_i(t)$, in order to assess the relationship between variables $AE_i(t)$ and $CEC_i(t, t+1)$ we must control for variable $CE_i(t)$. We propose the following multivariate linear regression.

$$\log(CEC_{i,t}) = \beta_0 + \beta_1 \log(AE_{i,t}) + \beta_2 \log(CE_{i,t}) + \varepsilon_{i,t} \quad (5.3)$$

where $CEC_{i,t}$ is the cost efficiency change of firm i between periods t and $t+1$, $AE_{i,t}$ and $CE_{i,t}$ its allocative and cost efficiency at period t , β_0 the intercept, β_1 and β_2 the regression coefficients, and $\varepsilon_{i,t}$ an idiosyncratic error term. We are taking the logarithm of the independent and dependent variables in order to have a linear relationship².

¹Where j is the lag.

² $\log(CEC_{i,t}) = \log(CE_{i,t+1}) - \log(CE_{i,t})$ has a linear relationship with $\log(CE_{i,t})$.

Because the control variable $\log(CE_{i,t})$ is included in the model, the fitted coefficient $\hat{\beta}_1$ is the partial—or *ceteris paribus*³—effect of $\log(AE_{i,t})$ on $\log(CEC_{i,t})$ when $\log(CE_{i,t})$ is “held fixed”. The model in (5.3) allows use to compute the predicted change in cost efficiency change given a change in allocative efficiency—controlling for cost efficiency. If the estimated change in $\log(CE_{i,t})$ is held constant (i.e. $\Delta\log(\widehat{CE}_{i,t}) = 0$) we have the following.

$$\Delta\log(\widehat{CEC}_{i,t}) = \hat{\beta}_1\Delta\log(\widehat{AE}_{i,t}) + \hat{\beta}_2\Delta\log(\widehat{CE}_{i,t}) \quad (5.4)$$

$$\Delta\log(\widehat{CEC}_{i,t}) = \hat{\beta}_1\Delta\log(\widehat{AE}_{i,t}) \quad (5.5)$$

$$\% \Delta \widehat{CEC}_{i,t} = \hat{\beta}_1 \% \Delta \widehat{AE}_{i,t} \quad (5.6)$$

What (5.6) means is that—assuming no change in $CE_{i,t}$ —a change of 1% in variable $AE_{i,t}$ is estimated to change variable $CEC_{i,t}$ by $\hat{\beta}_1\%$.

Because we are dealing with panel data we should control for unobservable variables that may have influenced cost efficiency changes. Those variables can be of two types. First, there may be unmeasurable variables that change across DSOs such as business practices or environmental factors. Those account for *individual heterogeneity* (i.e. firm-specific unobservable factors). Second, there may also be unmeasurable variables that change over time, but not across DSOs. Those account for *time heterogeneity*. We rewrite (5.3) as

$$\log(CEC_{i,t}) = \beta_0 + \beta_1 \log(AE_{i,t}) + \beta_2 \log(CE_{i,t}) + \mu_i + \nu_t + \varepsilon_{i,t} \quad (5.7)$$

where μ_i and ν_t are the individual and time effects, respectively. To test for the significance of those effects on our regression model we can compute the F -statistic comparing regression models (5.3) and (5.7). The F -test concludes that there are significant *two-ways* effects⁴. We can assess whether those effects are fixed or random with the Hausman test (Hausman (1978)). The test rejected the null hypothesis⁵ in favour of fixed effects. We can thus conclude that it is relevant to use the *two-ways fixed-effects regression model* presented in (5.7) for our analysis.

Like every linear regression, model (5.7) relies on several assumptions. We need to run some diagnostics on our regression model to confirm that its results are robust enough to be trusted. We need to test for the following assumptions (Wooldridge (2015)):

- The model is linear in its parameters.
- The error term $\varepsilon_{i,t}$ of the model is normally distributed.

³“All other things being equal.”

⁴p-value < 2.2e−16 in favour of the alternative hypothesis assuming significant individual and time effects.

⁵p-value < 2.2e−16

- There is no perfect collinearity among the independent variables.
- There is no serial correlation.
- The variance of the error term $\varepsilon_{i,t}$ is constant (i.e. homoscedasticity).

We already know that the model is linear since we used the logarithms of the variables instead of their true value. To test whether the residuals are normally distributed we first take a look at the density and Q-Q plots of the standardised residuals of model (5.7) (see Figures A.3 and A.4 in Appendix A). The standardised residuals do not seem to follow the Normal distribution. The Shapiro-Wilk normality test confirms that intuition by rejecting the null hypothesis of normality⁶. Nonetheless—following the conclusion of Lumley et al. (2002)⁷—our data sample is sufficiently large for us to perform our regression even though the distribution of its standardised residuals is not normal.

In order to assess whether the collinearity of our independent variables poses an issue to the interpretation of the regression coefficients we compute the *variance inflation factor*—or VIF. The VIF represents the increase in the variance of estimated coefficients β_1 and β_2 due to collinearity. A rule of thumb often used in the literature is to assume significant collinearity if the square root of the VIF is above 2. There is no reason for concern if the VIF is below 10 (Bowerman and O’connell (1990)). The VIF has a value of 1.75 for both our independent variables. We can thus conclude that there is no significant collinearity among them.

When analysing panel data the error term should be serially uncorrelated. Wooldridge (2010) proposed a test for serial correlation in the errors of fixed-effects panel models. This test did not reject the null hypothesis of serially uncorrelated residuals⁸ for our model.

The last assumption we need to test is homoscedasticity. Looking at the plot of the standardised residuals against the values fitted by the model (see Figure A.5 in Appendix A) the variance of residuals seems quite constant except for some data points around 0.3 and 0.8. Heteroscedasticity does not bias the estimation of the regression coefficients but causes the OLS standard errors to be biased, preventing us from relying on the t -test to see if the coefficients are significant (Wooldridge (2015)). When performing the Breusch-Pagan test for heteroscedasticity we rejected the null hypothesis of homoscedasticity⁹, concluding that the variance of the error term $\varepsilon_{i,t}$ is not constant (Breusch and Pagan (1979)). White (1980) proposed a heteroscedasticity-consistent covariance matrix

⁶p-value < 2.2e-16

⁷“The t -test and least-squares linear regression do not require any assumption of Normal distribution in sufficiently large samples. Previous simulations studies show that “sufficiently large” is often under 100.” (Lumley et al. (2002))

⁸p-value = 0.6472

⁹p-value = 0.003147

estimator to compute heteroscedasticity-robust t -tests. With this estimator we can test the significance of coefficients β_i even if the homoscedasticity assumption is violated.

TABLE 5.8: Summary of the regression model from (5.7)

Variable	Estimate	SD	t -statistic	p-value
$\log(AE_{i,t})$	0.082	0.041	1.998	0.046*
$\log(CE_{i,t})$	-0.783	0.071	-11.026	$< 2.2\text{e-}16^*$

Notes: *significant (< 0.05), adjusted $R^2 = 29.15\%$, SD = standard deviation.

Since $\log(AE_{i,t})$ has a significant positive effect¹⁰ on $\log(CEC_{i,t})$ (i.e. $\hat{\beta}_1 > 0$) we can confirm that—controlling for its cost efficiency score and for out set of operators—a firm’s allocative efficiency score is positively correlated with its subsequent cost efficiency change. Hence, the lower an DSO’s allocative efficiency is—or the higher its allocative inefficiency—the lower its subsequent cost efficiency change will be. We can conclude that we have empirically demonstrated Hypothesis 3¹¹.

5.3.2 Complexity factor ξ

We now compute the complexity factor ξ_i proposed in (3.9).

$$\xi_i = \frac{\log(AE_i) - \log(TE_i)}{\log(CE_i)} \quad (5.8)$$

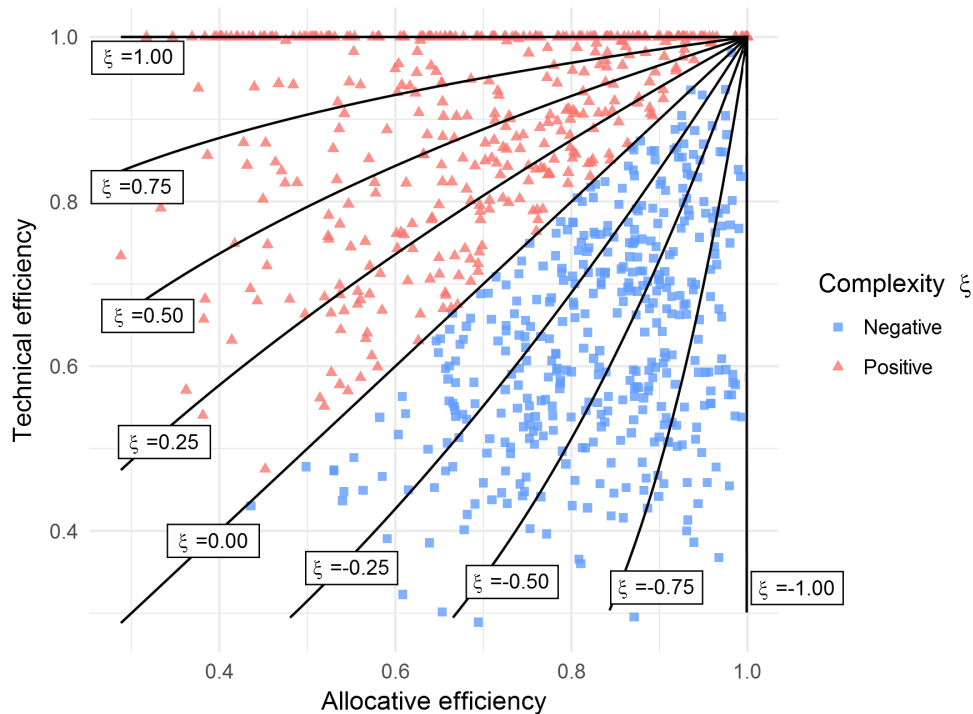
Figure 5.6 plots the observed efficiency profiles along with the “*iso-complexity*” curves representing the possible efficiency profiles η for different level of complexity ξ . By definition, efficiency profiles which primary source of cost inefficiency is allocative have a positive complexity factor while those which primary source is technical have a negative complexity factor.

Table 5.9 summarises the descriptive statistics of variables ζ_{TE_i} , ζ_{AE_i} , and ξ_i for observations with cost efficiency below 99%¹².

¹⁰Even though the effect is small (i.e. $\log(CEC_{i,t})$ increases of 0.082% when $\log(AE_{i,t})$ increases of 1%) the t -statistic shows that this effect is significantly different from zero.

¹¹One should remember that correlation does not imply causation. We can not conclude that we have proven Hypothesis 3 but only that we found a significant relationship with the available data.

¹²As explained in Section 3.2 the complexity factor ξ_i is only defined for cost inefficient firms. We thus only consider operators with at least 1% of cost inefficiency (i.e. 794 observations out of 858).

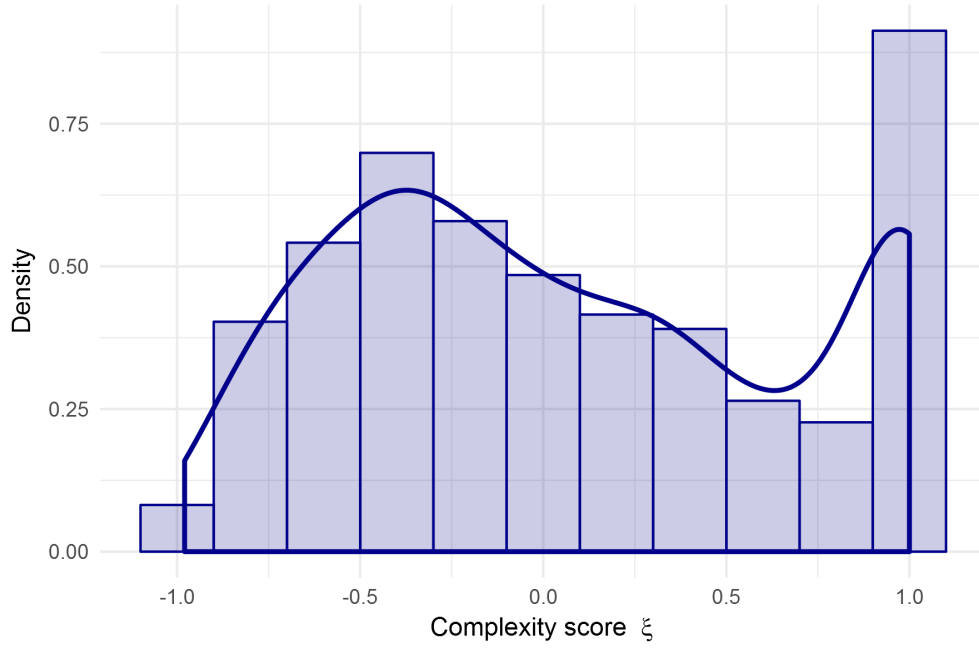
FIGURE 5.6: Efficiency profile η and complexity ξ TABLE 5.9: Proportions ζ_{TE} and ζ_{AE} , and complexity ξ

Measure	Unit	Min	Mean	Median	Max	SD
ζ_{TE_i}	%	0.00	46.31	51.00	98.97	30.55
ζ_{AE_i}	%	1.02	53.69	49.00	100.00	30.55
ξ_i	N/A	-0.979	0.073	-0.020	1.000	0.611

Notes: Descriptive statistics for the 794 observations which cost efficiency is below 99%. SD = standard deviation.

While cost inefficiency can be exclusively explained by allocative inefficiency, it seems that—in our sample—no operator is exclusively technically inefficient. Proportion ζ_{TE_i} has a maximal value of 98.97% indicating that technical inefficiency can be—at most—responsible for 98.97% of an operator's cost inefficiency. For this reason the minimal observed complexity factor ξ_i is -0.979. The mean complexity is close to zero (i.e. $\bar{\xi} = 0.073$) indicating that the cost inefficiency of the average operator is equally due to its technical and allocative components.

The density plot of the complexity factor (see Figure 5.7) exhibits a rather Gaussian shape up to $\xi = 1$ where density greatly increases. There is a high frequency of operators with a complexity at unity as they reach technical efficiency before reaching allocative efficiency.

FIGURE 5.7: Density plot and histogram of complexity ξ

Example 6. Let us consider once again operators SEVAB Nät AB and Ystad Energi AB from Example 5. While they are both operating around 53% cost efficiency, SEVAB Nät AB has a complexity factor of -0.979 and Ystad Energi AB a complexity of 1.000.

5.3.3 Simulation

We have demonstrated the relevance of the complexity factor ξ and analysed its statistics and distribution for our sample of observations. We are now simulating its application in the regulatory context by computing the firm-specific catch-up period T_i developed in Section 3.3 for a fictional revenue cap regulation starting in 2007. Since we are only interested in inefficient firms we are only considering operators with cost efficiency below 99%¹³.

¹³127 operators out of 143.

TABLE 5.10: Results (efficiency scores, ζ_{TE_i} , ζ_{AE_i} , and ξ_i) (2006)

Measure	Unit	Min	Mean	Median	Max	SD
Cost efficiency	%	22.87	60.03	61.44	98.18	16.71
Technical efficiency	%	32.75	77.25	76.44	100.00	16.88
Allocative efficiency	%	41.92	77.69	78.60	98.18	13.05
ζ_{TE_i}	%	0.00	46.64	50.01	92.81	28.86
ζ_{AE_i}	%	7.18	53.36	49.99	100.00	28.86
ξ_i	N/A	-0.856	0.067	0.000	1.000	0.577

Notes: Descriptive statistics for the 127 observations which cost efficiency is below 99%. SD = standard deviation.

The average complexity factor of operators is close to zero (i.e. $\bar{\xi} = 0.067$). Cost inefficiency is—on average—equally distributed between its technical and allocative components. 63 out of the 127 inefficient operators have to implement reallocation strategies as their primary mean of catching-up to the frontier. The main strategy that the remainder 64 DSOs should established is improved managerial performance for a proportional reduction of inputs.

To put it differently, the primary source of inefficient costs of half of the regulated operators has to be reduced through the reallocation of energy losses, weighted network length, and network capital. This is done by at least two of the following:

- Increase or reduction of energy losses. Depending on the relative price of energy losses the operator should allow more energy losses (if cheaper) or implement solutions to prevent them (if more expensive).
- Increase or reduction of weighted network length. In other words, it represents the inputs associated with operating expenditures such as labour, metering systems, etc. For example the operator might have to let go of part of its personnel or hire more people.
- Investment or disinvestment in network capital. This means purchasing or reselling distribution assets and equipment which—as stated in Section 3.1.2—is difficult to do for DSOs.

After computing complexity ξ_i we need to set parameters T_0 and α . Recall from Section 3.3.1 that we can set the default catch-up period T_0 based on empirical data as follows.

$$T_0 = \frac{-\ln(\overline{CE}(0))}{\ln(\overline{CEC}(t, t+1))} \quad (5.9)$$

where $\overline{CE}(0)$ is the average cost efficiency and $\overline{CEC}(t, t+1)$ the average yearly cost efficiency change. As explained in Section 4.1, the failed NPAM regulation was implemented in 2003. Since post-2003 average efficiency improvement is poor we decide to use the efficiency measurements from period 2000-2003 to estimate T_0 . From Table 5.4 we know that the average cost efficiency score was 53.48 and 63.74% in 2000 and 2003, respectively. Therefore the average yearly catch-up in this period was

$$\overline{CEC}(t, t+1) = \sqrt[3]{\overline{CEC}(2000, 2003)} = \sqrt[3]{\frac{\overline{CE}(2003)}{\overline{CE}(2000)}} = \sqrt[3]{\frac{0.6374}{0.5348}} = 1.060 \quad (5.10)$$

Using (5.9) we compute the default catch-up period T_0 .

$$T_0 = \frac{-\ln(0.5348)}{\ln(1.060)} = 10.69 \quad (5.11)$$

A firm with a complexity of zero will be granted 10.69 years to catch-up to the frontier. The advantage of computing T_0 as such is that it addresses the issue of arbitrariness by exclusively relying on efficiency measurements computed with the same DEA model and for the same set of operators.

As stated in Section 3.3.2, computing parameter α based on empirical data is beyond the scope of this thesis as it requires more extensive econometric analyses on the relationship between the complexity ξ and the catch-up speed. Since we need to give α a value to simulate the catch-up requirement we will arbitrarily set it to 2.5. This means that a DSO which cost inefficiency is exclusively due to allocative inefficiency will be granted 5 additional years (i.e. $2\alpha = 5$) compared to a DSO which cost inefficiency is exclusively due to technical inefficiency.

$$T_i = T_0 + \alpha\xi_i = 10.69 + 2.5\xi_i \in [8.19; 13.19] \quad (5.12)$$

$$\chi_i = 1 - CE_i(2006) \exp(1/(10.69 + 2.5\xi_i)) \quad (5.13)$$

TABLE 5.11: Catch-up (T_i , χ_i)

Measure	Unit	Min	Mean	Median	Max	SD
Catch-up period T_i	years	8.54	10.85	10.68	13.19	1.44
Catch-up requirement χ_i	%	0.01	5.09	4.72	14.51	2.85

Notes: Descriptive statistics for the 127 observations which cost efficiency is below 99%. SD = standard deviation.

If—starting in 2007—a fictional revenue cap was implemented with the catch-up requirement defined in (5.13) the average operator would be granted 10.85 years to catch-up

to efficiency. The DSO with the smallest catch-up period of 8.54 years is “Nynäshamn Energi AB” which demonstrated a cost efficiency of 74.30% decomposed in 75.91 and 97.88% for technical and allocative efficiency, respectively. Among the firms which have been granted the longest period of 13.19 years, “Herrljunga Elektriska AB” is the operator demonstrating the lowest cost efficiency of 41.92% exclusively explained by allocative inefficiency. Figure 5.8 plots the cost efficiency scores against their catch-up requirement χ_i along with the curve corresponding to a complexity factor of zero.

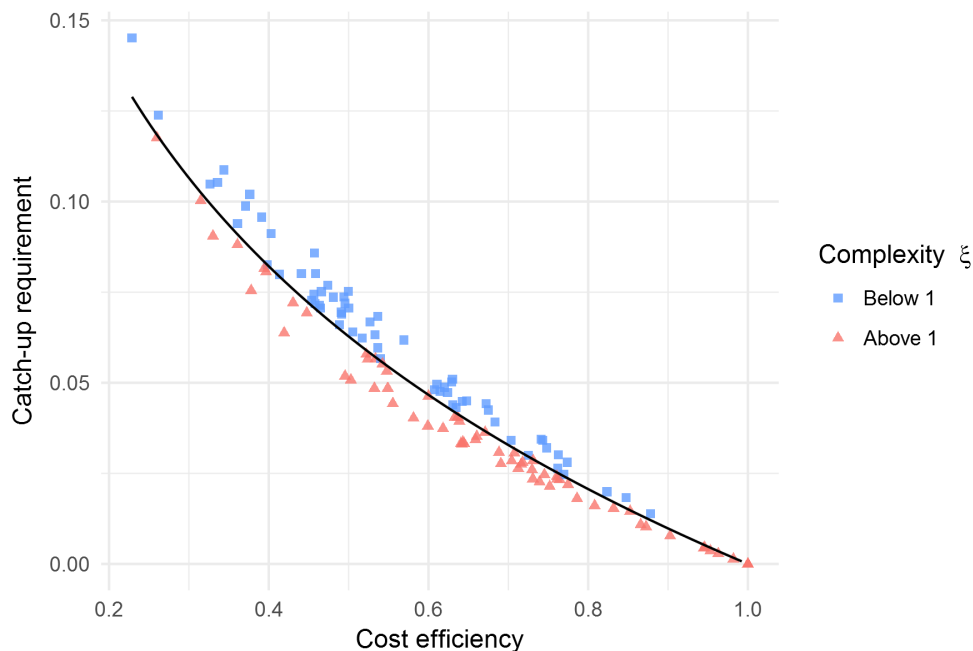


FIGURE 5.8: Catch-up requirement χ_i vs. Cost efficiency

5.4 Summary of the results

In this chapter we empirically demonstrated Hypothesis 3 through regression analysis. We can conclude that—for the same cost efficiency score—an operator with higher allocative inefficiency demonstrates a slower catch-up than one with lower allocative inefficiency. By applying the proposed complexity factor to our data we found that the DSOs demonstrate various efficiency profiles. The average operator has its cost inefficiency equally explained by poor managerial effort and allocation of inputs. The proposed model for firm-specific catch-up period has been applied to the 2006 data with an average catch-up period of 10.85 years.

Chapter 6

Conclusion

When determining the catch-up speed that their fixed-price regulation will require from inefficient network utilities, regulators should make use of the benchmarking measures of technical and allocative efficiency that the decomposition of cost efficiency provides. Through this decomposition regulators can assess with greater precision the source of a distribution system operator's excess costs. Due to the different improvement strategies technical and allocative inefficiency imply, the efficiency profile of a cost inefficient network utility has a significant impact on its catch-up speed and must therefore be assessed.

We reviewed the current state-of-the-art on the benchmarking tools and measures used in network regulation and—more specifically—the computation of the X-factor in fixed-price contracts. The lack of economical foundation for the catch-up speed that incentive regulation requires from distribution system operators has been assessed. Benchmarking measurements are extensively used in the estimation of an operator's incumbent cost inefficiency but the period over which network operators must eliminate that inefficiency lacks research.

To address this gap in the literature we proposed a model to compute a firm-specific catch-up period based on the historical efficiency improvements of the industry and the efficiency profiles of its operators. Through a review of the literature on regulation and regression analysis we showed that operators which primary source of cost inefficiency is allocative demonstrate slower catch-up due to the more complex strategies they have to implement in order to improve. From this assessment we proposed a complexity factor computed from the efficiency profile that represents the extent to which the cost inefficiency of an operator is due to managerial sub-performance or sub-optimal input-mix. This complexity factor allows us to differentiate between firms with identical cost

efficiency based on the complexity of their efficiency profile and to grant them a catch-up period accordingly.

The primary contribution of this thesis to the literature on network regulation comes from the analysis of the source of a firm's cost inefficiency that enables us to better assess its controllability. When setting the catch-up requirement the regulator must consider the implications that this required cost reduction will have on the regulated utilities. By addressing the source of cost inefficiency—whether it is technical, allocative, or both—regulators are making the most out of the available efficiency measures in order to make their regulatory scheme fairer and more relevant in terms of catch-up speed.

Our methodology displays some limitations that should be addressed. The conclusions we came to result from analyses only considering a single benchmarking model. Decisions on the preferred frontier model, its specification and set of variables have a significant impact on the computed efficiency measures, and therefore on the assessment of an operator's efficiency profile. More specifically, the implications of allocative inefficiency on the catch-up speed highly depend on the set of inputs considered in the benchmarking model. The complexity of a reallocation of inputs is clearly determined by the aforementioned inputs.

Even though we only considered the sector of electricity distribution the insights provided in this thesis can and should be applied to other sectors. Incentive regulation is present in many other industries such as transport (e.g. airports) or healthcare (e.g. hospitals). The conclusions we proposed concerning the implications of technical and allocative efficiency on the catch-up potential of regulated firms are only relevant for distribution system operators. Further research is required to apply our approach to other industries in need of regulation.

Appendix A

Additional figures

A.1 Non-discretionary variables

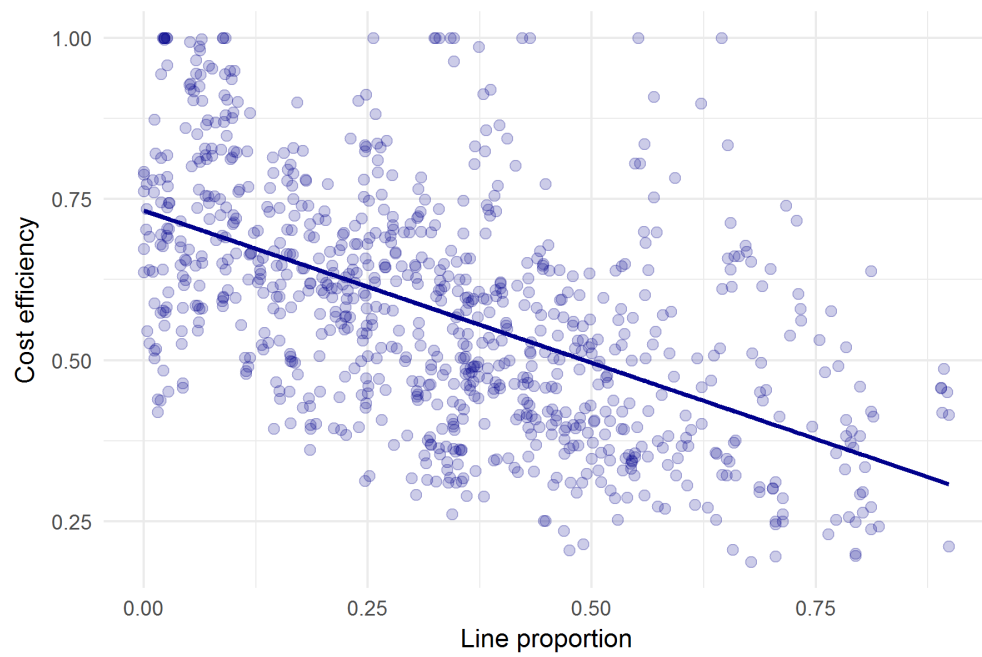


FIGURE A.1: Cost efficiency vs. Line proportion

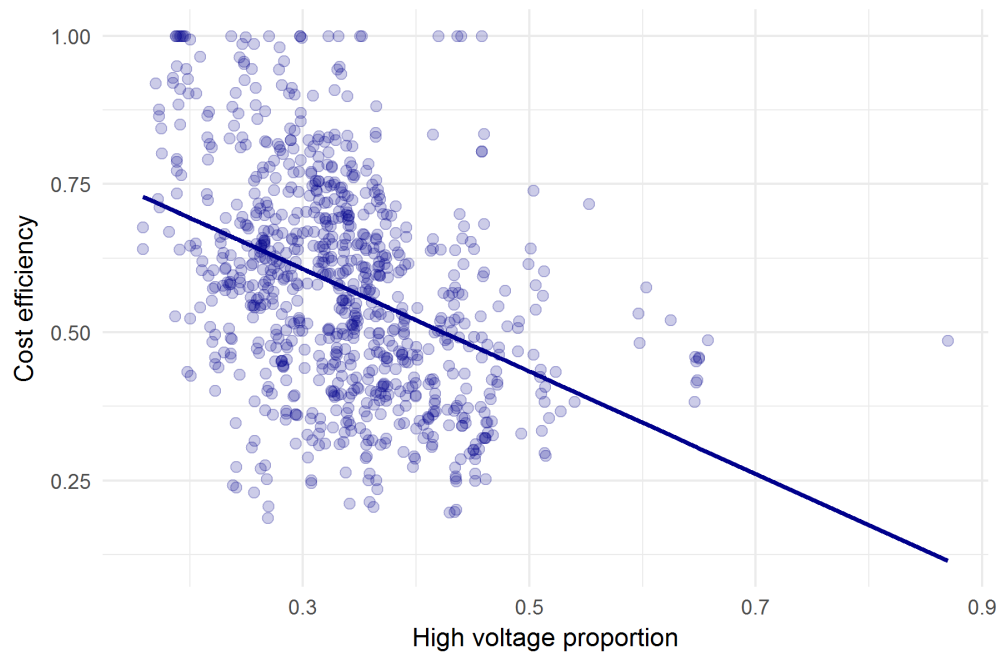


FIGURE A.2: Cost efficiency vs. High voltage proportion

A.2 Regression diagnostics

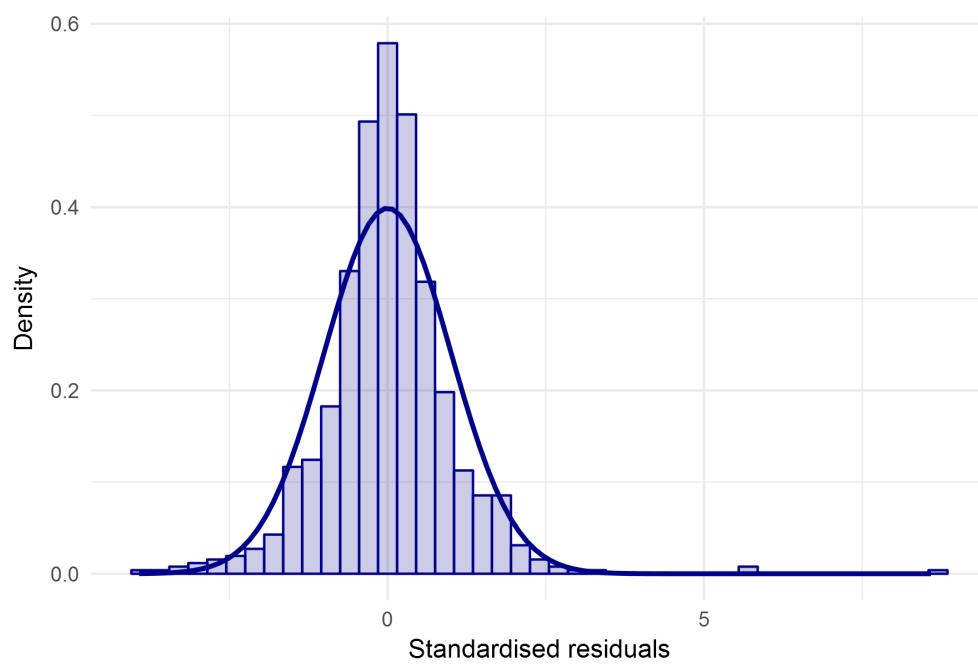


FIGURE A.3: Density plot of the standardised residuals of model (5.7)

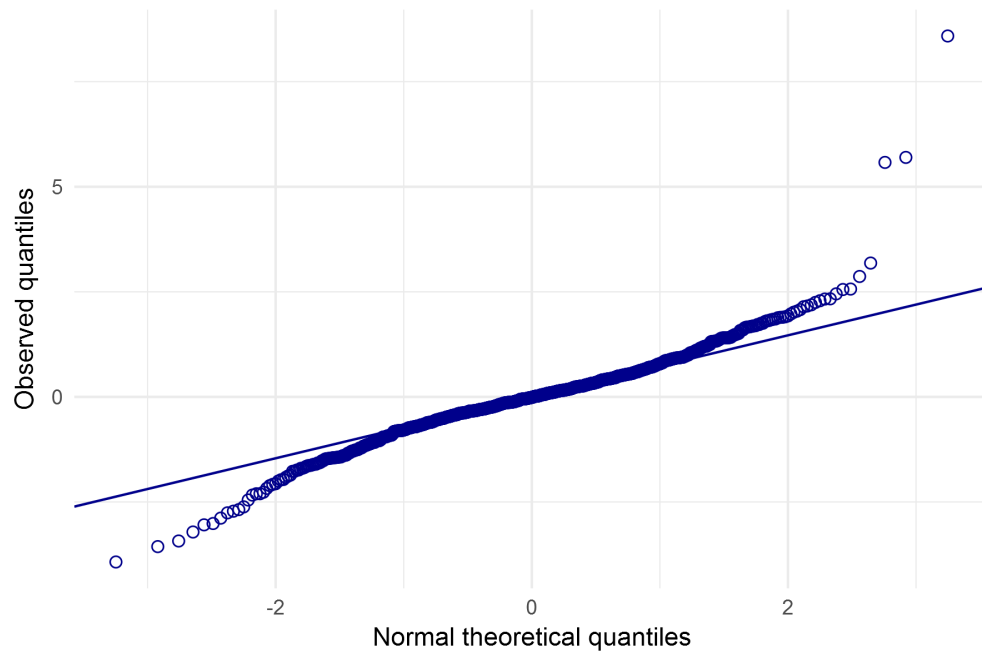


FIGURE A.4: Q-Q plot for the standardised residuals of model (5.7)

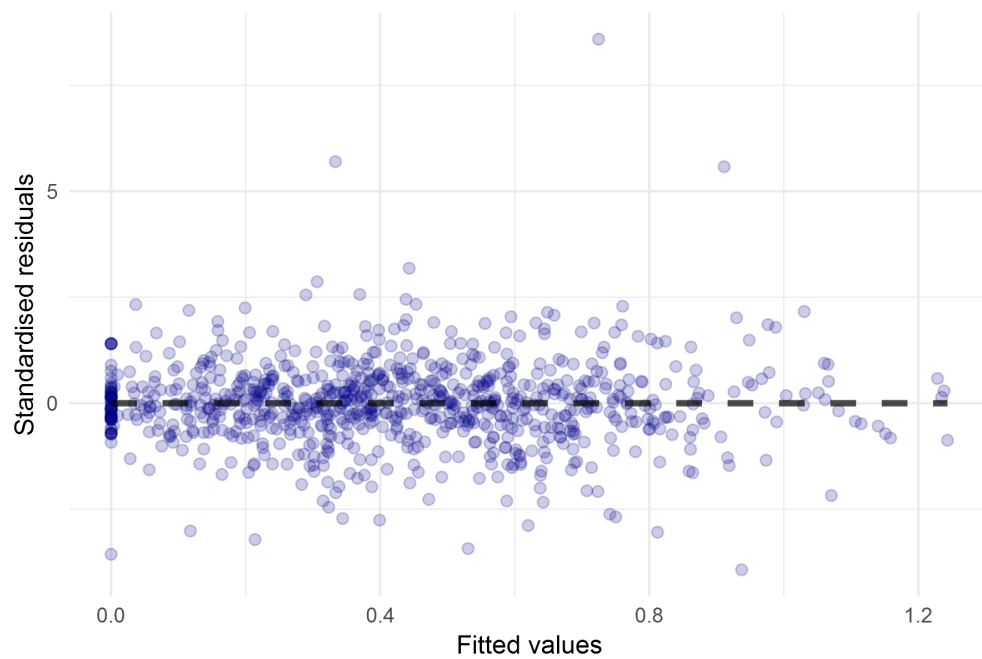


FIGURE A.5: Standardised residuals vs. Fitted values of model (5.7)

Appendix B

CD/USB content

Name	Extension	Description
Raw dataset	.csv	Unprocessed dataset from the regulator STEM. (Section 4.1)
Merging	.R	R script to retroactively create the merged operators. (Section 4.1)
Set up dataset	.R	R script to clean and prepare the dataset for analysis. (Section 4.1)
DEA without z-variables	.R	R script to compute the efficiency measures without considering non-discretionary variables. (Section 5.1).
Test for z-variables significance	.R	R script to test the significance of the proposed non-discretionary variables. (Section 5.1)
DEA with z-variables	.R	R script to compute the efficiency measures while considering non-discretionary variables. (Section 5.2)
Test for hypothesis	.R	R script to demonstrate Hypothesis 3. (Section 5.3.1)
Simulate catch-up requirement	.R	R script to simulate the catch-up requirement of the fictional revenue cap regulation. (Section 5.3.3)

Bibliography

- Agrell, Per J (2015) “Incentive Regulation of Networks,” *Reflets et perspectives de la vie économique*, Vol. 54, pp. 103–132.
- Agrell, Per J and Peter Bogetoft (2013) “Benchmarking and regulation,” CORE Discussion Papers 2013008, Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE), Belgium.
- Agrell, Per J, Peter Bogetoft, and Jørgen Tind (2005) “DEA and dynamic yardstick competition in Scandinavian electricity distribution,” *Journal of Productivity Analysis*, Vol. 23, pp. 173–201.
- Agrell, Per J and Emili Grifell-Tatjé (2016) “A dynamic model for firm-response to non-credible incentive regulation regimes,” *Energy Policy*, Vol. 90, pp. 287–299.
- Armstrong, Mark and David EM Sappington (2007) “Recent developments in the theory of regulation,” *Handbook of industrial organization*, Vol. 3, pp. 1557–1700.
- Banker, Rajiv D and Richard C Morey (1986) “Efficiency analysis for exogenously fixed inputs and outputs,” *Operations research*, Vol. 34, pp. 513–521.
- Bauer, Paul W, Allen N Berger, Gary D Ferrier, and David B Humphrey (1998) “Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods,” *Journal of economics and business*, Vol. 50, pp. 85–114.
- Beesley, Michael E and Stephen C Littlechild (1989) “The regulation of privatized monopolies in the United Kingdom,” *The RAND Journal of Economics*, Vol. 20, pp. 454–472.
- Bernstein, Jeffrey I and David EM Sappington (1999) “Setting the X factor in price-cap regulation plans,” *Journal of Regulatory Economics*, Vol. 16, pp. 5–26.
- (2000) “How to determine the X in RPI-X regulation: a user’s guide,” *Telecommunications Policy*, Vol. 24, pp. 63–68.

- Bogetoft, Peter, Rolf Färe, and Børge Obel (2006) “Allocative efficiency of technically inefficient production units,” *European Journal of Operational Research*, Vol. 168, pp. 450–462.
- Bogetoft, Peter and Lars Otto (2010) *Benchmarking with DEA, SFA, and R*, Vol. 157: Springer Science & Business Media.
- Bowerman, Bruce L and Richard T O’connell (1990) *Linear statistical models: An applied approach*: Brooks/Cole.
- Bradford, David F, Richard A Malt, and Wallace E Oates (1969) “The rising cost of local public services: some evidence and reflections,” *National Tax Journal*, Vol. 22, pp. 185–202.
- Breusch, Trevor S and Adrian R Pagan (1979) “A simple test for heteroscedasticity and random coefficient variation,” *Econometrica: Journal of the Econometric Society*, Vol. 47, pp. 1287–1294.
- Campos-Méndez, Javier, Antonio Estache, and Lourdes Trujillo (2001) “Processes, information, and accounting gaps in the regulation of Argentina’s private railways,” Policy Research Working Paper Series 2636, The World Bank.
- Carrington, Roger, Tim Coelli, and Eric Groom (2002) “International benchmarking for monopoly price regulation: The case of Australian gas distribution,” *Journal of Regulatory Economics*, Vol. 21, pp. 191–216.
- Caves, Douglas W, Laurits R Christensen, and W Erwin Diewert (1982) “The economic theory of index numbers and the measurement of input, output, and productivity,” *Econometrica: Journal of the Econometric Society*, Vol. 50, pp. 1393–1414.
- Charnes, Abraham, William W Cooper, and Edwardo Rhodes (1978) “Measuring the efficiency of decision making units,” *European journal of operational research*, Vol. 2, pp. 429–444.
- Coelli, Tim (2003) *A primer on efficiency measurement for utilities and transport regulators*, Vol. 953: World Bank Publications.
- Commission, Productivity et al. (2013) “Electricity network regulatory frameworks,” *Inquiry Report*, Vol. 1, p. 62.
- Daraio, Cinzia and Léopold Simar (2007) *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*: Springer Science & Business Media.

- Daskovska, Alexandra, Léopold Simar, and Sébastien Van Bellegem (2010) "Forecasting the Malmquist productivity index," *Journal of Productivity Analysis*, Vol. 33, pp. 97–107.
- Debreu, Gerard (1951) "The coefficient of resource utilization," *Econometrica: Journal of the Econometric Society*, Vol. 19, pp. 273–292.
- Färe, Rolf, Shawna Grosskopf, Björn Lindgren, and Pontus Roos (1994a) "Productivity developments in Swedish hospitals: a Malmquist output index approach," in *Data envelopment analysis: Theory, methodology, and applications*: Springer, pp. 253–272.
- Färe, Rolf, Shawna Grosskopf, and CA Knox Lovell (1994b) *Production frontiers*: Cambridge University Press.
- Farrell, Michael James (1957) "The measurement of productive efficiency," *Journal of the Royal Statistical Society. Series A (General)*, Vol. 120, pp. 253–290.
- Førsund, Finn R. (2002) "On the circularity of the Malmquist productivity index," ICER Working Papers 29-2002, ICER - International Centre for Economic Research.
- Hackett, Steven C (2010) *Environmental and natural resources economics: Theory, policy, and the sustainable society*: ME Sharpe.
- Haney, Aoife Brophy and Michael G Pollitt (2009) "Efficiency analysis of energy networks: An international survey of regulators," *Energy policy*, Vol. 37, pp. 5814–5830.
- Hausman, Jerry A (1978) "Specification tests in econometrics," *Econometrica: Journal of the Econometric Society*, Vol. 46, pp. 1251–1271.
- Hölmstrom, Bengt (1979) "Moral hazard and observability," *The Bell journal of economics*, Vol. 10, pp. 74–91.
- Jamasb, Tooraj and Michael Pollitt (2000) "Benchmarking and regulation: international electricity experience," *Utilities policy*, Vol. 9, pp. 107–130.
- (2003) "International benchmarking and regulation: an application to European electricity distribution utilities," *Energy policy*, Vol. 31, pp. 1609–1622.
- Joskow, Paul L (2008) "Incentive regulation and its application to electricity networks," *Review of Network Economics*, Vol. 7, pp. 1–14.
- Jullien, Bruno (2000) "Participation constraints in adverse selection models," *Journal of Economic Theory*, Vol. 93, pp. 1–47.
- Koopmans, Tjalling C (1951) *Activity analysis of production and allocation*, No. 13: Wiley New York.

- Kopp, Raymond J and W Erwin Diewert (1982) "The decomposition of frontier cost function deviations into measures of technical and allocative efficiency," *Journal of Econometrics*, Vol. 19, pp. 319–331.
- Laffont, Jean-Jacques and Jean Tirole (1993) *A theory of incentives in procurement and regulation*: MIT press.
- Leibenstein, Harvey (1966) "Allocative efficiency vs. "X-efficiency"," *The American Economic Review*, Vol. 56, pp. 392–415.
- Liston, Catherine (1993) "Price-cap versus rate-of-return regulation," *Journal of Regulatory Economics*, Vol. 5, pp. 25–48.
- Littlechild, Stephen C (1983) *Regulation of British Telecommunications profitability: report to the Secretary of State, February 1983*: Department of Industry.
- Lowry, Mark Newton and Lullit Getachew (2009) "Statistical benchmarking in utility regulation: Role, standards and methods," *Energy Policy*, Vol. 37, pp. 1323–1330.
- Lumley, Thomas, Paula Diehr, Scott Emerson, and Lu Chen (2002) "The importance of the normality assumption in large public health data sets," *Annual review of public health*, Vol. 23, pp. 151–169.
- Makholm, Jeff D (2007) "Elusive efficiency and the X-factor in incentive regulation: The Törnqvist v. DEA/Malmquist dispute," *The Line in the Sand: The Shifting Boundaries Between Markets and Regulation in Network Industries. National Economic Research Associates, White Plains: New York*, pp. 95–115.
- Malmquist, Sten (1953) "Index numbers and indifference surfaces," *Trabajos de Estadística y de Investigacion Operativa*, Vol. 4, pp. 209–242.
- Maniadakis, Nikolaos and Emmanuel Thanassoulis (2004) "A cost Malmquist productivity index," *European Journal of Operational Research*, Vol. 154, pp. 396–409.
- Navani, JP, NK Sharma, and Sonal Sapra (2012) "Technical and non-technical losses in power system and its economic consequence in Indian economy," *International Journal of Electronics and Computer Science Engineering*, Vol. 1, pp. 757–761.
- Nishimizu, Mieko and John M Page (1982) "Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia, 1965–78," *The Economic Journal*, Vol. 92, pp. 920–936.
- Ruggiero, John (1996) "On the measurement of technical efficiency in the public sector," *European Journal of Operational Research*, Vol. 90, pp. 553–565.

- (1998) “Non-discretionary inputs in data envelopment analysis,” *European Journal of Operational Research*, Vol. 111, pp. 461–469.
- Schweinsberg, Andrea, Marcus Stronzik, and Matthias Wissner (2011) “Cost Benchmarking in Energy Regulation in European Countries—Study for the Australian Energy Regulator (WIK-Consult).”
- Shepherd, Ronald William (2015) *Theory of cost and production functions*: Princeton University Press.
- Shuttleworth, Graham (2005) “Benchmarking of electricity networks: Practical problems with its use for regulation,” *Utilities Policy*, Vol. 13, pp. 310–317.
- Thanassoulis, Emmanuel (2013) *Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software*: Springer Science & Business Media.
- Uri, Noel D (2001) “Technical efficiency, allocative efficiency, and the implementation of a price cap plan in telecommunications in the United States,” *Journal of Applied Economics*, Vol. 4, pp. 163–186.
- Weitzman, Martin L (1980) “The "ratchet principle" and performance incentives,” *The Bell Journal of Economics*, Vol. 11, pp. 302–308.
- White, Halbert (1980) “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica: Journal of the Econometric Society*, Vol. 48, pp. 817–838.
- Wooldridge, Jeffrey M (2010) *Econometric analysis of cross section and panel data*: MIT press.
- (2015) *Introductory econometrics: A modern approach*: Nelson Education.

