This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Finding the right yardstick: regulation of electricity networks under heterogeneous environments

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1016/j.ejor.2017.07.059>

PUBLISHER

© Elsevier

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This paper was accepted for publication in the journal European Journal of Operational Research and the definitive published version is available at <https://doi.org/10.1016/j.ejor.2017.07.059>

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Bjorndal, Endre, Mette Bjorndal, Astrid Cullmann, and Maria Nieswand. 2017. "Finding the Right Yardstick: Regulation of Electricity Networks Under Heterogeneous Environments". Loughborough University.
<https://hdl.handle.net/2134/26264>.

Finding the Right Yardstick: Regulation under Heterogeneous Environments

Endre Bjørndal^{a,*}, Mette Bjørndal^a, Astrid Cullmann^b, Maria Nieswand^b

^a*Department of Business and Management Science, Norwegian School of Economics,
Helleveien 30, 5045 Bergen, Norway*

^b*Department of Firms and Markets, DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany*

Abstract

Revenue cap regulation is often combined with systematic benchmarking to reveal the managerial inefficiencies when regulating natural monopolies. One example is the European energy sector, where benchmarking is based on actual cost data, which are influenced by managerial inefficiency as well as operational heterogeneity. This paper demonstrates how a conditional nonparametric method, which allows comparison of firms operating under heterogeneous technologies, can be used to estimate managerial inefficiency. A dataset of 123 distribution firms in Norway is used to show aggregate and firm-specific effects of conditioning. We compare the models presently used by the Norwegian regulator to a suggested conditional model, thus providing a viable alternative to distinguish between managerial inefficiency and operational heterogeneity. The use of the conditional model in this regulatory context leads, first, to a decrease in aggregate efficient costs and, second, to a reallocation effect that affects the relative profitability of firms and relative customer prices, thus providing a fairer basis for setting revenue caps.

JEL-Classification: L94, C44, L51

Keywords: Data Envelopment Analysis, Yardstick Regulation, Electricity Distribution

*Corresponding author

Email addresses: `endre.bjorndal@nhh.no` (Endre Bjørndal),
`mette.bjorndal@nhh.no` (Mette Bjørndal), `acullmann@diw.de` (Astrid Cullmann),
`mnieswand@diw.de` (Maria Nieswand)

1. Introduction

In general, revenue-cap schemes tend to constrain monopolistic firm behavior by “capping” the revenues that regulated firms are allowed to earn. Similar to price cap regulation, revenue caps provide incentives for cost reductions by efficiency improvement since firms are allowed to keep any additional profits from cost savings.

From the perspective of maximizing welfare, the regulator aims to implement a revenue-cap scheme in which the regulated firms set prices equal to their average costs, i.e. the efficient costs C_i^* (the so-called second-best solution in regulation economics, Laffont and Tirole (1993); Armstrong and Sappington (2007); Shleifer (1985)). The challenge, then, is to determine C_i^* due to information asymmetry.¹ To determine C_i^* and reveal managerial inefficiency, regulators may combine revenue-cap schemes with systematic benchmarking techniques (Agrell and Bogetoft, 2013).

By comparing the performance of the firms via cost functions, benchmarking provides information on the unknown technology, the cost structure and efficient costs, and thus is crucial to determining revenue caps.² In other words, the more precisely C_i^* is approximated, the closer the revenue caps are to the second-best solution (Laffont and Tirole, 1993).

The precision of efficient cost approximation is related to external factors (Shleifer, 1985) describing the operational environment of regulated firms. External factors cause heterogeneity in terms of technology, and hence, efficient costs. Consequently, the levels of efficient cost differ between firms according to their external conditions.

In this paper, we discuss how regulators may fail to implement efficient cost levels and revenue caps by not properly accounting for external firm characteristics. We consider unconditional nonparametric benchmarking applica-

¹The regulator does not know the efficient cost structures of the regulated firms, and firms have no incentives to reveal this information.

²Generally, benchmarking can be interpreted as creating a hypothetical competition among natural monopolies. Thereby, a mechanism of yardstick competition in the spirit of Shleifer (1985) is introduced, since the revenue of a particular electricity distribution firm is based on the costs of all other distribution firms.

tions used in regulation mainly based on data envelopment analysis (DEA), which may set revenue caps too high and compensate firms for external disadvantages *and* managerial inefficiency. In particular, we emphasize the importance of selecting the appropriate reference set for comparing firms to determine efficient costs, C_i^* .

We suggest that conditioning a firm’s production process on its operational environment provides more reliable peers. The Norwegian Water Resources and Energy Directorate (NVE), the Norwegian regulator, has extensive experience with revenue cap regulation since 1997. We use a dataset of 123 Norwegian electricity distribution firms to compare a proposed conditional DEA benchmarking model to unconditional models, such as the one used by the Norwegian regulator. The conditional model leads to a different peer selection, and thus different compensation for external factors. We demonstrate the effect of conditioning on the aggregate efficient costs, firms’ profits, and consumers’ prices for the distribution service. For the Norwegian case, the use of the conditional approach leads to a reduction in aggregate efficient cost, although the main effect is a reallocation of revenue between firms, leading to changes in relative profitability and customer prices in the industry.

Even though we use Norway as a case study, our empirical application is relevant for the European regulators using nonparametric benchmarking methods to determine revenue caps.³ Our analysis is also relevant for other industry analyses in which nonparametric benchmarking is applied to com-

³Bogetoft and Otto (2011) and Agrell and Bogetoft (2013) give a summary of different benchmarking methods used across European countries. In Austria, Belgium, Germany, and the Netherlands the nonparametric DEA is used to determine revenue caps. Finland used until 2007 DEA to assign efficiency improvements targets; and from 2008 to 2011 the average of DEA and SFA. From 2012 onwards, the Finnish regulator applies a semiparametric method (StoNED) to derive efficiency estimates (Kuosmanen and Kortelainen, 2012). With respect to the treatment of external factors in regulatory practice, we note that e.g. in Germany the only external variable taken into account is size of the distribution area. It is directly included in the LP model as an additional output. The influence of other external factors is tested in second stage regressions. Finland has implemented innovative semiparametric models to account for external variables (see e.g. Kuosmanen, 2012; Johnson and Kuosmanen, 2011).

pare decision-making units in different environments.

The remainder of this paper is organized as follows. Section 2 explains the implementation of yardstick regulation via benchmarking and the challenges of separating managerial inefficiency and operational heterogeneity. Section 3 shows that the conditional framework is appropriate when heterogeneous firms are benchmarked against each other. Section 4 presents our empirical strategy and our data. Section 5 summarizes the empirical findings. Section 6 shows the effects of conditioning for regulation. Section 7 concludes.

2. Revenue cap regulation and benchmarking

2.1. Revenue cap regulation

A revenue cap (Joskow, 2007) can be implemented based on a combination of observed and efficient costs, formalized as

$$R_i = \rho \cdot C_i^* + (1 - \rho) \cdot C_i, \quad (1)$$

where the capped revenue R_i for firm i is determined by its observed cost C_i and efficient cost C_i^* . The parameter $\rho \in [0, 1]$ determines the weights of the efficient and observed costs, and thus the strength of the regulatory system.

Obviously, the observed firm-specific costs available to the regulator are influenced by firms' managerial inefficiency and different operational environments. A major challenge for the regulator is to determine the firm-specific revenue caps, R_i , such that they account for the external effects, meaning that firms will only be compensated for disadvantages, not managerial inefficiency.

2.2. Benchmarking and managerial inefficiency

In regulatory practice, the efficient cost, C_i^* , is often determined by means of an *unconditional* nonparametric benchmarking method, where data envelopment analysis (DEA), developed by Charnes et al. (1978), is the most popular variant.⁴ DEA estimates the unknown technology or production set

⁴Frontier models are widely applied in performance measurement. The theoretical foundations derive from Koopmans (1957), Debreu (1959) and Farrell (1957). In regulatory

Ψ from a given sample of observed vectors of inputs $x_i \in \mathbb{R}_+^p$ and outputs $y_i \in \mathbb{R}_+^q$ used by the firms $i = 1, \dots, n$, and where p and q represent the number of input and output factors, respectively. The boundary of Ψ is the frontier and represents the estimated unknown technology (see Appendix A for the standard DEA model). Comparing each individual firm to this frontier determines the firm's specific managerial efficiency. Solving a linear program assuming constant returns to scale (CRS) according to Charnes et al. (1978) provides efficiency scores $\hat{\theta} \in \mathbb{R}^n$ such that $0 \leq \hat{\theta}_i \leq 1$ for $i = 1, \dots, n$.⁵ The efficiency estimate for firm i , i.e., $\hat{\theta}_i$, is a measure relative to the frontier, which is determined by fully efficient observations $j \in \{1, \dots, n\}$ with $\hat{\theta}_j = 1$, i.e., the peers.

In the unconditional case, where DEA does not consider external factors, all observations belong to the reference set of the particular observation of interest and can potentially serve as its peers, irrespective of their operational environments. However, this involves the risk of evaluating an observation to an infeasible frontier when external factors are disadvantageous or beneficial, respectively, to its production process.

practice, nonparametric approaches have outperformed parametric methods due to their easy implementation and interpretation, which is often why regulators favor their use. See Bogetoft and Otto (2011) for a comparison and critical evaluation of the approaches in regulatory practice.

⁵In regulatory practice the choice of the returns to scale assumption is made by the regulator. Thus, this assumption could serve specific regulatory objectives and can affect the whole structure of the sector in the long run. Regulators often assume CRS as proposed by Charnes et al. (1978), because it implies that any observed production plan can be arbitrarily scaled up and down, and that all firms are able to reach the optimal scale and operate at an optimal scale. The optimal firm size is under the CRS assumption within management's control. This creates incentives for mergers e.g. and spin-offs to take place for firms to achieve the most productive scale size (Amundsveen and Kvile, 2015; Kuosmanen, 2012). A VRS or FDH assumption instead would make the efficiency measures conditional on scale without creating incentives for mergers or divestitures. Moreover, with VRS or FDH, there is a chance that the regulation of large firms becomes weak (Bjørndal et al., 2010). CRS is the strictest regulatory objective with respect to scale adaption, and the Norwegian Regulator assumes CRS for the regulatory model. In order to be able to compare our model results to the regulatory outcome we assume CRS as well.

2.3. Managerial inefficiency versus operational heterogeneity

It is worthwhile to emphasize that the external factors can have different channels through which they influence firms' performance. Bădin et al. (2012) point out that external factors may influence the technology itself represented by the boundary of the production set Ψ , thus causing a frontier shift, a change in the distribution of inefficiency between firms, or both. The regulator has to compensate for the resulting cost disadvantages only from exogenous frontier shifts due to different environments, and not for managerial inefficiency.

We will assume that the external impact can be measured by a set of vectors $z_i \in \mathbb{R}^r$ for $i = 1, \dots, n$, where r is the number of external factors. In practice, regulators use second-stage regressions (Bjørndal et al., 2010; Agrell et al., 2014) to control for the heterogeneity of operational environments on the frontier. The z -variables are regressed on the DEA estimates to determine the impact of the operational environments on efficiency. The efficiency scores are then adjusted to compensate for the impact of z -variables. One concern, however, is that second-stage regressions are correct only when z -variables fulfill the separability condition (Bădin et al., 2012; Simar and Wilson, 2007).⁶ However, when there is separability, there is no frontier shift due to environmental differences, and a compensation for external factors is irrelevant, since they do not affect efficient costs.⁷

In addition, if z -variables cause a frontier shift, and thus are not separable, which is usually the case, DEA efficiency scores are not economically

⁶Separability means that the z -variables do not influence the attainable set, and thus, the frontier.

⁷We note that alternative approaches were developed to account for the effect of external factors that do not require the separability condition. Banker and Morey (1986b) and Banker and Morey (1986a) derived a one-stage-model where continuous or categorical external factors are included directly via an additional restriction in the formulation of the linear program. Their approach has often been used because of its simple implementation. Ruggiero (1996, 1998) also extend the original DEA linear program to allow for external factors and compare different approaches using simulated data. Newer semi-nonparametric efficiency analysis models (StoNEZD) are proposed by Johnson and Kuosmanen (2011, 2012) which extend the standard StoNED estimator Kuosmanen and Kortelainen (2012) to account for external factors. This approach is currently applied in the regulation of the Finnish electricity distribution sector (Kuosmanen, 2012).

sound, since they are based on a frontier that production units are unable to reach (Bădin et al., 2012). Using second-stage regressions ignores the potential impact of z -variables on the frontier and their potential impact on the distribution of inefficiency. Therefore, compensation based on the two-stage approach is likely to capture multiple effects and lead to under- or overcompensation of the individual regulated firms, which results in a too high or too low C_i^* , and therefore a too high or too low R_i .

3. Conditional benchmarking

Conditional nonparametric benchmarking approaches, which do not require the separability condition, can account for the multiple effects of the operational environment on a firm's performance⁸. This is accomplished by restricting the selection of reference sets, i.e., the group of firms used to compare against the firm of interest. By means of kernel estimation, the reference sets are restricted with respect to z -variables, prior to measuring actual performance such that firms are only compared to others with similar environments.⁹ Compensation for the resulting cost disadvantages is only based on the frontier shift, thus leading to an adequate implementation of the revenue cap scheme.

3.1. Constructing a conditional efficiency estimator

The conditional DEA estimator is based on an attainable production set, $\hat{\Psi}^z$, given as the support of a conditional distribution function where the production process is conditional to a particular level of z (Daraio and Simar, 2007b; Bădin et al., 2010).¹⁰ Estimation of the distribution function

⁸To incorporate external factors into the performance evaluation, conditional efficiency estimation was proposed initially by Cazals et al. (2002) in the order- m framework and further developed by Daraio and Simar (2005, 2007a,b). The approach aims to compare only units that operate under similar operational environments, i.e. the selection of the reference group for a particular observation is conditional on their z -variables.

⁹Thus, conditioning the performance evaluation on z -variables assumes that the frontier is feasible for the firm to reach, and that the efficiency scores are meaningful (Bădin et al., 2012).

¹⁰The statistical properties of this estimator are derived in Kneip et al. (2008) and its consistency is established in Jeong et al. (2010).

requires a smoothing technique, and we therefore perform kernel estimation with an Epanechnikov kernel $K(\cdot)$ as in, e.g., Daraio and Simar (2005) and Daraio and Simar (2007b).¹¹ The kernel K_h is defined as

$$K_h = K((z_i - z_k)/h), \quad (2)$$

where z_i and z_k are vectors of z -variables for a unit i and its reference unit k , respectively, and h is the vector of selected bandwidths. For each of the external factor variables we compute a bandwidth based on least squares cross validation (Hall et al., 2004; Li and Racine, 2007, 2008). Note that the bandwidth selection procedure relies on estimating the conditional probability distribution function of y , given a particular level of z .¹² Hall et al. (2004) emphasize that their proposed method assigns large smoothing parameters to components of z that are irrelevant for estimating the density of y . Therefore, the sizes of the selected bandwidths themselves already contain information about the impact of particular z -variables on the production output y . We then use the obtained bandwidths to estimate the kernel probabilities with the function in Equation 2. Firms closely located to firm i in terms of z thereby receive higher probabilities to be selected into the reference set of the observation of interest, whereas small (or even zero) kernel probabilities are assigned to firms with very different operational environments than firm i .

As shown in Daraio and Simar (2007b), the conditional DEA efficiency estimate for firm i under the assumption of CRS is given by

$$\hat{\theta}_i^c = \min\{\theta \mid \theta x_i \geq \sum_{j|z_i-h \leq z_j \leq z_i+h}^n \lambda_j x_j, \ y_i \leq \sum_{j|z_i-h \leq z_j \leq z_i+h}^n \lambda_j y_j, \quad (3)$$

and $\lambda_j \geq 0$ for $j = 1, \dots, n\}$,

where the vector $h \in \mathbb{R}^r$ represents bandwidths of appropriate size. For

¹¹The authors suggest using kernels with compact support in the framework of conditional boundary estimation.

¹²See Hall et al. (2004) for a detailed presentation of the method.

each observation, the bandwidths determine the range of z in which other observations are considered being similar. Hence, we consider only the observations within this range as potential peers for the unit of interest and select them into the respective reference group. That is, we restrict the reference set of firm i to firms with positive kernel probabilities.

3.2. Bias-correction

DEA efficiency scores are based on finite samples of observations and construct a best-practice frontier, i.e. by construction they are upward-biased Simar and Wilson (1998). We correct for the bias in $\hat{\theta}$ and $\hat{\theta}^c$ by applying the m -bootstrap first proposed by Kneip et al. (2008), and extended by Simar and Wilson (2011).¹³ Unlike the naive bootstrap, this approach allows consistent bias-correction by drawing bootstrap subsamples of size $m = n^\kappa$ from the given sample of size n with $\kappa \in (0, 1)$. Each of the $b = 1, \dots, B$ bootstrap samples (replications) provides a random subsample of size m which we use to compute the bootstrapped vector of efficiency scores denoted as $\hat{\theta}_{m,b} \in \mathbb{R}^n$. The bias for the unconditional DEA and the conditional DEA is defined as

$$\hat{bias}_B(\hat{\theta}) = \left(\frac{m}{n}\right)^{2/(p+q)} \cdot \left(B^{-1} \sum_{b=1}^B \hat{\theta}_{m,b} - \hat{\theta}\right), \quad (4)$$

$$\hat{bias}_B(\hat{\theta}^c) = \left(\frac{m}{n}\right)^{2/(p+q+r)} \cdot \left(B^{-1} \sum_{b=1}^B \hat{\theta}_{m,b}^c - \hat{\theta}^c\right), \quad (5)$$

where the factors $(m/n)^{2/(p+q)}$ and $(m/n)^{2/(p+q+r)}$ control for the estimators' convergence rates¹⁴ and the effects of the different sample sizes in both the true world and bootstrap world (Simar and Wilson, 2008). Then, we obtain bias-corrected efficiency scores $\tilde{\theta}$ and $\tilde{\theta}^c$, by subtracting the bias from $\hat{\theta}$ and

¹³For the conditional case, we use the kernel probabilities to construct the bootstrap samples. This is consistent with the idea of conditioning the production process and is supposed to give even more precise insights about the effect of z -variables on the production process.

¹⁴Park et al. (2010) show that the DEA estimator under CRS assumption converges at rate $n^{-2/(p+q)}$. Accordingly, for the conditional DEA, we take the number of z -variables into account.

$\hat{\theta}^c$, respectively:

$$\tilde{\theta} = \hat{\theta} - bias_B(\hat{\theta}) \quad (6)$$

$$\tilde{\theta}^c = \hat{\theta}^c - bias_B(\hat{\theta}^c). \quad (7)$$

The described subsampling procedure further provides the information required for estimating confidence intervals. According to Simar and Wilson (2011), the confidence interval estimates $\hat{CI}_{\hat{\theta}}$ and $\hat{CI}_{\hat{\theta}^c}$ of nominal size α are given by

$$\hat{CI}_{\hat{\theta}} = \left[\hat{\theta}(1 + n^{-2/(p+q)} \cdot \psi_{\alpha/2}), \hat{\theta}(1 + n^{-2/(p+q)} \cdot \psi_{1-\alpha/2}) \right], \quad (8)$$

$$\hat{CI}_{\hat{\theta}^c} = \left[\hat{\theta}^c(1 + n^{-2/(p+q+r)} \cdot \psi_{\alpha/2}), \hat{\theta}^c(1 + n^{-2/(p+q+r)} \cdot \psi_{1-\alpha/2}) \right], \quad (9)$$

with $\psi_{\alpha/2}$ and $\psi_{1-\alpha/2}$ being the percentiles of the empirical distribution of $m^{2/(p+q)} \cdot \left(\frac{\hat{\theta}}{\hat{\theta}_b} - 1 \right)$ and $m^{2/(p+q+r)} \cdot \left(\frac{\hat{\theta}^c}{\hat{\theta}_b^c} - 1 \right)$, respectively. The confidence intervals are used to investigate the accuracy of the point estimates.

4. Empirical strategy and data

4.1. Empirical strategy

Our objective is to demonstrate that *conditioning* a firm's production process on its operational environment gives us a better yardstick for determining efficient cost. Therefore, we compare the outcomes of the *unconditional* model used in regulation with our proposed *conditional* model to show the effects of conditioning on efficiency estimates, revenue caps, firms' profits, and prices consumers have to pay. We estimate four different models:

1. The bias-corrected unconditional DEA model ($\tilde{\theta}$).
2. The bias-corrected conditional DEA model ($\tilde{\theta}^c$).
3. The unconditional DEA model with a second-stage regression to correct for external factors ($\tilde{\theta}^{Ray}$), like in Ray (1991).
4. As 3, with the second-stage regression that is used by the Norwegian regulator ($\tilde{\theta}^{NVE}$), i.e., as described in Amundsveen et al. (2014).

We run 2,000 replications to obtain the bias-corrected efficiency estimates by the m -bootstrap where drawing the reference sets depends on the respective kernel probabilities. We select $m = 60$ using the *leave-one-out-order-m* algorithm proposed by Daraio and Simar (2007a) and Simar (2003) based on Cazals et al. (2002).

Note that $\tilde{\theta}^{Ray}$ and $\tilde{\theta}^{NVE}$ are both based on $\tilde{\theta}$, and they differ only with respect to how the regression to correct for external factors is performed. For the estimator $\tilde{\theta}^{Ray}$, the unconditional efficiency scores are regressed on the level of the external factor values directly (Ray, 1991). For the estimator $\tilde{\theta}^{NVE}$, the independent variables in the regression consist of the differences between the DMUs' external factor values and the corresponding factor values of their peers (Amundsveen et al., 2014).

The first effect of conditioning, i.e., a different peer selection, is seen from the difference between $\tilde{\theta}$ and $\tilde{\theta}^c$. The second effect of conditioning, i.e., compensating differently for z -variables, is seen from the difference between, on the one hand, $\tilde{\theta}^c$ and, on the other hand, $\tilde{\theta}^{Ray}$ and $\tilde{\theta}^{NVE}$.

4.2. Data

Our dataset comprises 123 Norwegian electricity distribution firms regulated by NVE. Since the deregulation of the Norwegian power market in 1990, the regulatory regime for distribution and regional transmission firms¹⁵ has gone through several phases. In the first years after deregulation, firms were subject to a rate of return regulation with rather weak efficiency incentives. In 1997, revenue cap regulation with benchmarking was introduced with five-year regulation periods.

Firm-specific efficiency requirements in each of the first two regulation periods, were based on DEA analysis of historical cost and output data. After 2007, the efficiency incentives were further strengthened with the introduction of annual updates of the revenue caps based on DEA analyses.¹⁶ There are separate benchmarking models for distribution and regional transmis-

¹⁵The TSO is subject to a separate regulation, which we do not discuss in this paper.

¹⁶A more detailed discussion of the different regulatory phases is given by Bjørndal et al. (2010) and Amundsveen and Kvile (2015).

sion. In the following, we focus on the regulation model for the distribution networks.

Table 1 lists the input and output variables in our dataset corresponding to those used in the regulatory benchmarking model. We follow the practice of NVE and construct our benchmarks based on average data over a five-year period.¹⁷ The input variable measures total expenditures (TotEx) in Norwegian Kroner (NOK), which comprise five cost elements: the value of lost load (VOLL), thermal power losses, capital depreciation, operation and maintenance expenses, and return on capital. In order to take five-year averages of the data, we have adjusted the data to the price level of a base year.¹⁸ Output variables measure the number of customers, length of the high voltage network, and the number of network stations (transformers).¹⁹

We use three variables to measure heterogeneous operational conditions (see Table 1). The first, z_{Dis} , is the average distance between the road network and a firm's network, representing the increased difficulty of maintaining a network that is not easily accessible. The second, z_{Und} , is the share of underground cables which may imply cost disadvantages because of expensive installation as well as cost advantages in terms of fewer outages and lower VOLL. The third is the share of fast-growing forest, z_{For} , that may represent a cost disadvantage due to the added cost of forest clearing.²⁰

¹⁷NVEs justification for their practice is to remove noise from the data, as well as to allow firms to appear as super-efficient relative to their own historical performance (Amundsveen and Kvile, 2015). We have used the NVE model for 2014, with average cost and output data from 2008-2012. This corresponds to how NVE estimates efficiency scores based on lagged data. We have calculated revenues in 2012 prices.

¹⁸We use an industry-specific price index for adjusting operations and maintenance costs and the consumer price index for the VOLL costs. Thermal losses are valued at the average system price for the base year. Capital depreciation is based on reported (nominal) book values, and the return on capital is calculated using the nominal rate of return set by the regulator for the base year.

¹⁹Given that the rate of convergence for CRS-DEA models is $n^{-2/(p+q)}$, it is worth mentioning that a benchmarking model for 123 observations including one input and three output variables, converges at the same rate as a parametric model with convergence rate $n^{-1/2}$. Econometrically, the parsimonious model specification of the Norwegian regulator is beneficial.

²⁰The two composite geographical variables used by NVE for this period combine several sub-variables, by using principal component analysis (PCA). The data values for these

Table 1: Definitions of variables.

Variable	Name	Sub-variable	Unit
Input			
TotEx	x	O&M costs	1000 NOK
		Value of lost load (VOLL)	1000 NOK
		Thermal power losses	1000 NOK
		Capital depreciation	1000 NOK
		Return on capital	1000 NOK
Outputs			
Customers	y_{Cus}		No. of customers
High voltage lines	y_{HV}		Kilometers
Network stations	y_{Net}		No. of stations
Environment			
Average distance to road	z_{Dis}		Meters
HV lines underground	z_{Und}		Share of HV network (0-1)
Forest (coniferous)	z_{For}		Share of HV network affected (0-1)

Table 2 lists the summary statistics for the 123 distribution firms. We use average data for the period 2008-2012, and we set 2012 as the base year for adjustment of the cost data. Note that averaging the data does not affect the external factor variables, since their values are constant over time in our dataset.

Table 2: Summary statistics ($n = 123$).

	Input	Outputs			Environment		
	TotEx	y_{Cus}	y_{HV}	y_{Net}	z_{Dis}	z_{Und}	z_{For}
Mean	93974.5	22669.7	796.2	1006.6	227.8	0.3386	0.1191
Std. dev.	187511.3	59221.2	1328.0	1893.4	208.6	0.1755	0.0994
Minimum	8456.2	1014.2	54.0	59.0	70.4	0.0571	0.0000
Median	35728.8	6386.8	324.8	369.4	142.9	0.3060	0.1191
Maximum	1579179.2	547693.0	8494.6	13491.4	1056.4	0.8641	0.3916

Notes: Average data for the period 2008-2012; 2012 is base year for adjustment of the cost data.

variables are close together, with both positive and negative numbers. We could not obtain reasonable bandwidths that remain stable over multiple runs with these variables included. In order to obtain comparable results, we have excluded the composite variables from the analysis of all model variants.

5. Results

5.1. Selected bandwidths

Table 3 lists the computed bandwidths.²¹ Given that they are the smoothing parameters for estimating the conditional density of the multivariate kernel function, they inform us about the impact of the z -variables on the produced output y . Therefore, they indicate the z -variables' importance for selecting the reference set in our conditional efficiency estimation.

For the distance to road variable z_{Dis} , Table 3 shows a very large bandwidth compared to the median value of 142.9 shown in Table 2. Thus, this variable is smoothed out in the kernel density estimation, which implies that the overall impact of z_{Dis} on output y is limited, and therefore, the reference sets are not restricted by z_{Dis} . The magnitudes of bandwidths for the other two variables, 0.30 for z_{Und} and 0.13 for z_{For} , are considerably smaller. For interpretation, we again compare the bandwidths to the median values in Table 2, i.e., 0.31 for z_{Und} and 0.12 for z_{For} , respectively. We conclude that the externally induced cost disadvantages, which would be measured as inefficiency, if not controlled for, are most likely due to z_{Und} and z_{For} .

Table 3: Estimated bandwidths for z -variables

Variable		Bandwidth
Average distance to road	z_{Dis}	1,299,522,861
HV lines underground	z_{Und}	0.30
Forest (coniferous)	z_{For}	0.13

5.2. Reference sets and selected peers

To illustrate how estimated bandwidths affect the selection of the reference set when individual firms are benchmarked, we start by looking at a firm i_0 . Figure 1 shows the two-dimensional contour plots of the kernel density function for firm i_0 indicated by the yellow point. Its values for z_{Dis} , z_{Und} , and z_{For} are 162, 0.5236, and 0.1307, respectively. We only select

²¹The estimates have been scaled so that they are comparable to the corresponding z -values in Table 2. For the Epanechnikov kernel the scaling factor is $\sqrt{5}$.

firms with positive kernel density values as members of firm i_0 's reference set, i.e., any firm i in the sample for which the values of each z -variable $j \in \{Dis, Und, For\}$ satisfies $|z_{ij} - z_{i_0j}| \leq h_j$.

The shaded areas in both panels of Figure 1 represent the combinations of two z -variables for which firms obtain positive kernel density values. The left panel shows that the possible peers of firm i_0 will be firms with z_{Und} and z_{For} values within the range of $[0.2274, 0.8197]$ and $[0.0018, 0.2596]$, respectively, i.e., the 70 firms indicated by black dots. Gray dots indicate firms that are too different from firm i_0 to be selected in its reference set. The shaded area in the right panel is not limited in the dimension of distance, which is due to this factor's large bandwidth. Note that this shaded area also includes gray-colored firms because they are similar in terms of z_{Dis} but not in terms of z_{Und} , and therefore, do not belong to the reference set of firm i_0 . So a firm must be similar along all external factor dimensions to be considered a potential peer for i_0 .

Figure 1: Kernel densities and reference set for firm i_0 .

Given that the kernel centers around firm i_0 , the density estimation produces a reference set specific to each firm. Figure 2 shows that both the size of the reference set and the number of peers differ between the unconditional

and the conditional approaches. The blue crosses and the black circles indicate the number of selected peers for each firm, with the unconditional and the conditional models, respectively. The number of selected peers for firm i is the number of firms out of its reference set that serve as peers for this firm in at least one of the bootstrap samples.²² For almost all firms, note that unconditional DEA selects more peers for efficiency evaluation than the conditional DEA. The horizontal green bars show the number of overlapping peers, i.e., peers that are used in both models. In extreme cases the overlap of selected peers is very small or even zero, which is particularly true for firms with a small number of selected peers under conditional DEA. In fact, Figure 2 shows that when unconditional DEA is used, each firm in our sample is evaluated against a frontier spanned by peers with dissimilar operational environments (for a detailed list of the sizes of reference sets and identical and different peers in the unconditional and conditional models, see Appendix B).

5.3. Efficiency estimates

We focus on our four efficiency measures ($\tilde{\theta}$, $\tilde{\theta}^c$, $\tilde{\theta}^{Ray}$, $\tilde{\theta}^{NVE}$) to demonstrate the effect of differences with respect to reference sets and peer selection. Table 4 lists the descriptive statistics of the estimated efficiency scores. As expected, the level of the estimator $\tilde{\theta}^c$ exceeds its unconditional counterparts for $\tilde{\theta}$. For example, the median efficiency score is 0.6990 when external factors are accounted for, while it is 0.6696 in the unconditional case. The difference between the point estimates indicates that external factors indeed produce cost disadvantages for the median firm, shifting its feasible frontier such that its efficiency increases by roughly 3 percentage points. Hence, if $\tilde{\theta}$ scores are used for further analysis, the median firm has been evaluated based on a technically infeasible frontier.

Figure 3 shows the point estimates of $\tilde{\theta}$ and $\tilde{\theta}^c$ along with the confidence

²²If λ_{bij}^m is the weight of firm j in the reference set of firm i in bootstrap sample b , we let $\bar{\lambda}_{ij}^m = B^{-1} \sum_{b=1}^B \lambda_{bij}^m$ denote the average weight of firm j for efficiency measurement of firm i across all B bootstrap samples.. The number of selected peers for firm i counts the firms j for which $\bar{\lambda}_{ij}^m > 0$.

Figure 2: Number of peers.

intervals at the 1 percent level obtained according to equations 8 and 9. In both cases, the confidence intervals cover the respective point estimates, thus supporting their accuracy. The $\tilde{\theta}^{Ray}$ and $\tilde{\theta}^{NVE}$ scores are based on the $\tilde{\theta}$ scores, and are not reported separately here.

Considering again Table 4, comparing $\tilde{\theta}^c$ to $\tilde{\theta}^{Ray}$ and $\tilde{\theta}^{NVE}$ reveals a mixed picture of the second-stage corrections. The estimator $\tilde{\theta}^{Ray}$ gives lower efficiency scores than $\tilde{\theta}^c$, while the reverse is true for the relationship between $\tilde{\theta}^{NVE}$ and $\tilde{\theta}^c$. We interpret this finding as empirical evidence of biased performance measures and note that this bias results from both the underlying and economically meaningless $\tilde{\theta}$ estimates and the second-stage regression, which assumes that z -variables do not affect the frontier. Based on these results, it is likely that NVEs present model compensates both the frontier shifts due to external variables and the managerial inefficiency because the effects of both on the estimated technology are not appropriately separated.

The rank correlations shown in Table 5 further illustrate the relationship

Figure 3: Point estimates and confidence intervals for unconditional (left) and conditional (right) model.

Table 4: Descriptive statistics of efficiency estimates.

Estimator		Min	Median	Mean	Max*	Std. dev.
1. Unconditional	$\tilde{\theta}$	0.4072	0.6696	0.6879	0.9670	0.1210
2. Conditional	$\tilde{\theta}^c$	0.4338	0.6990	0.7206	1.0000	0.1372
3. Ray two-stage	$\tilde{\theta}^{Ray}$	0.4032	0.6758	0.6962	1.0000	0.1254
4. NVE two-stage	$\tilde{\theta}^{NVE}$	0.4387	0.7303	0.7470	1.0259	0.1233

* The $\tilde{\theta}$ and $\tilde{\theta}^c$ estimates are bias-corrected, and their maximum values can therefore differ from 1.

between the estimators. We see that the two-stage estimators $\tilde{\theta}^{Ray}$ and $\tilde{\theta}^{NVE}$ give similar rankings, and these rankings are similar to the ranking implied by the unconditional estimator $\tilde{\theta}$. On the other hand, the ranking under conditional DEA is less correlated with the rankings based on the other three estimators. This could imply that the *relative* effects of adjustment for external cost drivers are greater with the conditional DEA method than with the two-stage procedures.²³

²³Alternatively, the high correlations could be a result of noise, which is not directly accounted for by the models discussed.

Table 5: Rank correlations between efficiency estimates.

	$\tilde{\theta}$	$\tilde{\theta}^c$	$\tilde{\theta}^{Ray}$	$\tilde{\theta}^{NVE}$
$\tilde{\theta}$	1.00	0.92	0.96	0.95
$\tilde{\theta}^c$	0.92	1.00	0.91	0.90
$\tilde{\theta}^{Ray}$	0.96	0.91	1.00	0.97
$\tilde{\theta}^{NVE}$	0.95	0.90	0.97	1.00

5.4. Impact of the external factors on the frontier

Figure 4: Impact of external factors on the frontier.

In this section we analyze the impact of the external factors on the production process by using the ratio of conditional to unconditional efficiency scores. Bădin et al. (2012) emphasize that unconditional measures of efficiency are economically meaningless if z -variables impact the frontier since units are compared to infeasible production plans. Knowing that conditional measures control for this, the ratio, therefore, informs us about the local effect of z -variables on the attainable frontier. The ratio δ is defined as

$$\delta = \frac{\tilde{\theta}^c}{\tilde{\theta}} \quad (10)$$

where we refer to the bias-corrected measures. The ratio δ takes the value of 1 if both measures are equal, i.e. there is no frontier shift, whereas other

values imply that the feasible frontier shifts due to z -variables. A ratio larger than 1 implies that the conditional exceeds the unconditional efficiency score indicating cost disadvantages due to the external factors, whereas ratios with values smaller than 1 indicate cost advantages. To analyze the local effect of the external factors on the feasible frontier, we plot the ratio of $\tilde{\theta}^c$ to $\tilde{\theta}$. Both panels in Figure 4 sort firm-specific efficiency ratios in an increasing order of the values of z_{Und} and z_{For} .²⁴

Notably, the ratios of the two performance measures differ from 1 for almost all firms; the range is roughly 0.9 to 1.3. Hence, z -variables significantly affect production. Also, for 89 of the 123 firms, the ratio, based on bias-corrected efficiency estimates, is larger than 1. Therefore, the frontier-shifting factors mainly lead to cost disadvantages.

6. Effects for regulation

The four models also demonstrate that the choice of benchmarking model, including how to compensate for the effect of external factors, has important regulatory effects for firms' customers and owners. We start by looking at the impact on the aggregate efficient costs and then look in detail at firm-specific profitability (return on capital) and the average prices customers pay per kWh. We base our calculations on the current revenue cap model used by the Norwegian regulator (NVE). Similar to Equation 1, as well as the DEA-based mechanism proposed by Bogetoft (1997), the revenue cap for firm i is set by

$$R_i = \rho \cdot (C_i^* + \Delta_i) + (1 - \rho) \cdot C_i, \quad (11)$$

where $\rho = 0.6$ (Amundsveen and Kvile, 2015). The efficient costs C_i^* are calculated by $\tilde{\theta}_i^v \cdot C_i$, where $v \in \{c, Ray, NVE\}$. The yardstick formula 11 is applied every year to set the annual revenue caps, but in order to simplify the presentation we will drop the time subscripts in our notation²⁵.

²⁴From the bandwidth analysis we know that the frontier shift is mainly driven by these two factors.

²⁵We focus on the most important features of the Norwegian regulation. In practice, there is a two-year time lag in the reporting; the revenue caps in year t must be based on

NVE calibrates the revenue caps, by adding the amount Δ_i to the efficient cost of each firm²⁶, in order to ensure that revenue equals cost for the industry as a whole, i.e., $\sum R = \sum C$. The rationale for the calibration, as described in Bjørndal et al. (2010) and Amundsveen and Kvile (2015), is to allow the representative firm, with an efficiency equal to the industry (cost-weighted) average, to have a return on its capital equal to the regulated rate of return. Given the calibration scheme, firms that have above-average efficiency scores will earn more than the regulated rate of return, while firms with below-average efficiency scores will earn less.

6.1. Effect on aggregate efficient costs

Because of the revenue calibration performed by NVE, the compensation scheme for external factors will have no effect on aggregate revenue. In order to study aggregate effects, we will therefore focus on efficient costs. The first row in Table 7 shows the aggregate efficient cost from the unconditional DEA model, i.e., $\sum C^*(\tilde{\theta})$, and the second row shows what the compensation for the external factors will be when it is based on one of the estimators $\tilde{\theta}^c$, $\tilde{\theta}^{Ray}$, or $\tilde{\theta}^{NVE}$. The compensation of 771 million NOK (MNOK) under the current NVE regime is higher compared to the conditional DEA model, where the compensation equals 425 MNOK. This could indicate that the currently implemented regime overestimates efficient production costs. We cannot however, conclude that two-stage methods always will overcompensate for cost drivers given by external factors. The compensation with $\tilde{\theta}^{Ray}$ is only 268 MNOK, i.e., considerably lower than with $\tilde{\theta}^c$.

Under the current regulatory scheme, where the regulator calibrates the

the data available after year $t - 2$. We assume that the average of the data for 2008-2012 is representative of a typical year, and we do not consider the timing of the revenue stream. Many firms also own and operate part of the regional transmission network. However, we do not consider this part of their revenue caps.

²⁶In the present regulation model, the calibration takes the form

$$\Delta_i = \frac{\sum C - \sum C^*}{\sum BV} BV_i,$$

where BV_i is the total book value of capital for firm i . The use of book values in the calibration formula is done to correct for a suspected age bias in the capital costs.

Table 6: Aggregate effects (MNOK).

		Estimation methods (v)		
		c	Ray	NVE
Efficient cost w/o external factors	$\sum C^*(\tilde{\theta})$	8,517	8,517	8,517
Compensation for external factors	$\sum (C^*(\tilde{\theta}^v) - C^*(\tilde{\theta}))$	425	268	771
Efficient cost with external factors	$\sum C^*(\tilde{\theta}^v)$	8,943	8,786	9,288
Revenue calibration	$\sum \Delta$	2,616	2,773	2,271
Industry revenue	$\sum (C^*(\tilde{\theta}^v) + \Delta)$	11,559	11,559	11,559

final revenue caps by adding Δ_i to the efficient cost of each firm, the over-compensation for external factors will not affect the aggregate revenue. The revenue calibration, i.e., $\sum \Delta$, equals 2,271 MNOK under the current regulatory model, whereas it would be 2,616 MNOK with our alternative model. As shown on the last row of Table 7, the aggregate revenue will be unaffected by the choice of compensation scheme, and equal to total industry cost.

6.2. Effect on firm-specific profits and customer prices

Figure 5: Effect on firms' profits ($r_{NVE2012} = 4.2\%$). Dots = uncalibrated effects.

In this section we focus on the differences in firm-specific revenues from

using the conditional estimator $\tilde{\theta}^c$ instead of the efficiency estimates presently used by the regulator ($\tilde{\theta}^{NVE}$).²⁷ Figure 5 illustrates how the choice of compensation ($\tilde{\theta}^c$ versus $\tilde{\theta}^{NVE}$) affects the owners of the network firms. The columns represent the relative difference in firm-specific revenues between NVE's DEA model and the conditional DEA,

$$\frac{R_i(\tilde{\theta}^{NVE}) - R_i(\tilde{\theta}^c)}{BV_i},$$

where we have divided the firm-specific revenues by their total book values, BV_i , to see the effects on return on capital. In order to make the example interesting in a more general setting, we also show firm-specific effects without the calibration effect, i.e., with $\Delta_i = 0$, as dots in the figure. The effects on return on capital vary from -3.5 (-3.0) percent to +5.9 (+6.4) percent, where a positive value reflects a higher return on capital under the NVE model compared to conditional DEA, and where numbers in parentheses are effects without calibration. The main effect that we see from the figure is a reallocation of revenue due to the change in compensation method, and this is related to the difference in relative efficiency scores under the respective methods, cf., the difference in rank correlations shown in Section 5.3. The difference between the calibrated and the uncalibrated effects is related to the overcompensation of 346 MNOK (771-425) in the efficient cost level by the NVE two-stage method, as discussed in the previous section. According to formula 11, 60 percent of this overcompensation will be awarded to the firms, and this represents an increase of 0.6 percent in the return on capital for each firm.

Figure 6 shows the same revenue differences,

$$\frac{R_i(\tilde{\theta}^{NVE}) - R_i(\tilde{\theta}^c)}{Energy_i},$$

²⁷Similar analyses could be done for $\tilde{\theta}^{Ray}$, however, we choose to focus on the suggested alternative method for compensating for external factors as opposed to the present regulatory practice.

Figure 6: Effect on calibrated revenue caps $\left(\frac{\sum Revenue}{\sum Energy} = 0.155 \text{ NOK/KWh} \right)$. Dots = uncalibrated effects.

only now relative to the total quantity of energy delivered by each firm, in order to approximate the average effect on the prices charged by each firm. The differences range from -0.024 (-0.017) NOK/KWh to +0.045 (+0.050) NOK/KWh, where a positive value reflects higher prices for customers under the NVE model compared to conditional DEA, and where numbers in parentheses represent the differences without calibration, i.e., with $\Delta_i = 0$. The average revenue collected from the customers amount to 0.155 NOK/KWh, hence, the reallocation effect of the compensatory scheme is considerable. The overcompensation of 346 MNOK in efficient costs amount to an increase, on average, of only 0.0028 NOK/KWh in the customer prices, hence the overcompensation effect is small compared to the reallocation effect.

7. Conclusion

In its basic form, nonparametric benchmarking methods, such as data envelopment analysis, that is implemented by several European energy regulators, cannot disentangle the effects of managerial inefficiency and difficult

operational environments. Under revenue cap regulation, firms should be compensated for the latter, but not for the first. In particular, if external factors are suspected to cause non-controllable costs, a two-stage method is often used by energy regulators, including a second stage regression to correct for external factors that are not accounted for in the first stage DEA model. In this setting, the two-stage method is likely to capture multiple effects, leading to over- or under-compensation of individual firms. Conditional non-parametric benchmarking approaches have been designed to overcome this challenge, restricting the selection of firms used to compare against the firm of interest, to those with similar environments.

This paper proposes a conditional DEA benchmarking model for electricity distribution, and compares it to an unconditional model, as well as two variants of two-stage models; (i) with second stage regression based on Ray (1991), and (ii) with second stage regression based on the method presently used by the Norwegian regulator, NVE. A dataset of 123 Norwegian electricity distribution firms is used to illustrate how managerial inefficiency can be estimated in a meaningful way by comparing among firms operating in comparable environments, i.e., a homogeneous technology. The proposed model is used to compare the effect of conditioning on total efficient costs and revenues, firms' rate of return, and customers' prices. Based on the results, we observe that the use of conditional benchmarking methods may lead not only to a different peer selection and changes in aggregate efficient cost, but also to a reallocation effect that affects the relative profitability of firms and relative customer prices. Thus, conditional benchmarking methods may provide a fairer basis for setting revenue caps.

This insight is relevant when using benchmarks for revenue cap regulation, and, more generally, when nonparametric benchmarking is used to compare decision making units in different operational environments.

Acknowledgements

This paper is produced as part of the project Municipal Infrastructure Companies against the Background of Energy Policy and Demographic Change (KOMIED) funded by the Leibniz Commission (SAW-2013-DIW-5), as well

as the project Benchmarking for Regulation of Norwegian Electricity Networks (ElBench) at the Centre for Applied Research at NHH (SNF). The paper has been presented at the European Workshop on Productivity and Efficiency Analysis (WEPEA), 2015, and the INFORMS Annual Meeting, 2015. We gratefully acknowledge Pio Baake, Tomaso Duso and Christian von Hirschhausen for valuable comments and suggestions. We further thank Ann Steward for providing language help and proof reading the article. The usual disclaimer applies.

References

- Agrell, P. J. and Bogetoft, P. (2013). Benchmarking and regulation. *CORE Discussion Paper, Louvain-la-Neuve, Belgium*, 2013/8.
- Agrell, P. J., Bogetoft, P., Koller, M., and Trinkner, U. (2014). Effizienzvergleich für Verteilernetzbetreiber Strom 2013, Bericht im Auftrag der Bundesnetzagentur (BNetzA). Technical report, Swiss Economics and Sumic-sid.
- Amundsveen, R., Kordahl, O.-P., Kvile, H. M., and Langset, T. (2014). Second stage adjustment for firm heterogeneity in DEA: A novel approach used in regulation of Norwegian electricity DSOs. *Recent Developments in Data Envelopment Analysis and its Applications*, page 334.
- Amundsveen, R. and Kvile, H. M. (2015). The development and application of an incentive regulation model - a balancing act. *The ICER Chronicle*, Edition 3.
- Armstrong, M. and Sappington, D. E. M. (2007). Recent developments in the theory of regulation. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization*, volume 3, chapter 27, pages 1557–1700. North Holland/ Elsevier: Amsterdam, The Netherlands.
- Bădin, L., Daraio, C., and Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3):818 – 833.

- Banker, R. and Morey, R. (1986a). Efficiency analysis for exogeneously fixed inputs and outputs. *Operations Research*, 34(4):513–521.
- Banker, R. and Morey, R. (1986b). Use of categorical variables in data envelopment analysis. *Management Science*, 32(12):1613–1627.
- Bjørndal, E., Bjørndal, M., and Fange, K.-A. (2010). Benchmarking in regulation of electricity networks in Norway: An overview. In Bjørndal, E., Bjørndal, M., Pardalos, P. M., and Rönnqvist, M., editors, *Energy, Natural Resources and Environmental Economics*, Energy Systems, pages 317–342. Springer Berlin Heidelberg.
- Bogetoft, P. (1997). DEA-based yardstick competition: the optimality of best practice regulation. *Annals of Operations Research*, 73:277–298.
- Bogetoft, P. and Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Springer-Verlag New York.
- Bădin, L., Daraio, C., and Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2):633–640.
- Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1 – 25.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.
- Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24(1):93–121.
- Daraio, C. and Simar, L. (2007a). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Studies in productivity and efficiency. Springer, New York.

- Daraio, C. and Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, 28(1-2):13–32.
- Debreu, G. (1959). *Theory of Value*. Cowles Foundation Monograph. Yale University Press, New Haven, Connecticut, U. S. A.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.
- Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026.
- Jeong, S.-O., Park, B., and Simar, L. (2010). Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, 173(1):105–122.
- Johnson, A. L. and Kuosmanen, T. (2011). One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNEZD method. *Journal of Productivity Analysis*, 36(2):219–230.
- Johnson, A. L. and Kuosmanen, T. (2012). One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research*, 220(2):559 – 570.
- Joskow, P. (2007). Regulation of natural monopoly. In A., M. P. and Shavell, S., editors, *Handbook of Law and Economics*, volume 2, chapter 16, pages 1227–1348. Elsevier, 1 edition.
- Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24:1663–1697.
- Koopmans, T. (1957). *Three Essays on the State of Economic Science*, chapter 1. McGraw-Hill, New York, U. S. A.

- Kuosmanen, T. (2012). Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. *Energy Economics*, 34(6):2189 – 2199.
- Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1):11–28.
- Laffont, J.-J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation*. The MIT Press.
- Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press, Princeton.
- Li, Q. and Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics*, 26(4):423–434.
- Park, B. U., Jeong, S.-O., and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *The Annals of Statistics*, 38(3):1320–1340.
- Ray, S. C. (1991). Resource-use efficiency in public schools: A study of Connecticut data. *Management Science*, 37(12):1620–1628.
- Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, 90(3):553–565.
- Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, 111(3):461–469.
- Shleifer, A. (1985). A theory of yardstick competition. *The RAND Journal of Economics*, 16(3):319–327.
- Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

- Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61.
- Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.
- Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H. O., Lovell, C. K., and Schmidt, S. S., editors, *The measurement of productive efficiency and productivity growth*, pages 421–521. Oxford University Press, Oxford.
- Simar, L. and Wilson, P. W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36(1):33–53.

Appendix A. Unconditional DEA

The main idea is to estimate the unknown technology or production set Ψ from a given sample of observed production units $i = 1, \dots, n$. The set Ψ contains all input-output-combinations that are feasible to the production units and is defined as

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}$$

with $x \in \mathbb{R}_+^p$ and $y \in \mathbb{R}_+^q$ being the vectors of inputs and outputs. The boundary of Ψ is referred to as the frontier and serves as the benchmark against which each individual observation is compared in order to determine its efficiency. The production set assuming constant returns to scale (CRS), as proposed by Charnes et al. (1978), is defined as

$$\begin{aligned} \hat{\Psi}_{CRS} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \lambda_i y_i, \ x \geq \sum_{i=1}^n \lambda_i x_i, \\ \text{and } \lambda_i \geq 0 \text{ for } i = 1, \dots, n\}. \end{aligned} \quad (\text{A.1})$$

Using $\hat{\Psi}_{CRS}$, the level of inefficiency $\hat{\theta}_i$ for firm i can be estimated by solving the following linear program

$$\begin{aligned} \hat{\theta}_i = \min\{\theta \mid \theta x_i \geq \sum_{j=1}^n \lambda_j x_j, \ y_i \leq \sum_{j=1}^n \lambda_j y_j, \\ \text{and } \lambda_j \geq 0 \text{ for } j = 1, \dots, n\}. \end{aligned} \quad (\text{A.2})$$

The efficiency estimate $\hat{\theta}_i$ is the Debreu-Farrell input efficiency score and indicates how much the unit of interest can reduce its inputs for a given level of output. If $\hat{\theta}_i = 1$, observation i is considered fully efficient, while $\hat{\theta}_i < 1$ implies inefficiency.

Appendix B. Reference sets and peers in different models

Table B.1: Sizes of reference sets and identical and different peers in unconditional and conditional models.

Firm	Ref. set uncond.	Number of peers uncond.	Ref. set cond.	Number of peers cond.	Number of identical peers	Number of peers in uncond./ not in cond.	Number of peers in cond./ not in uncond.
1	123	36	58	19	14	22	5
2	123	30	61	18	13	17	5
3	123	25	63	12	11	14	1
4	123	38	60	16	12	26	4
5	123	31	46	11	10	21	1
6	123	24	59	8	8	16	0
7	123	25	28	4	4	21	0
8	123	34	84	28	21	13	7
9	123	34	58	8	7	27	1
10	123	34	58	18	14	20	4
11	123	38	60	15	8	30	7
12	123	28	63	20	9	19	11
13	123	41	59	18	16	25	2
14	123	34	81	31	18	16	13
15	123	40	15	5	5	35	0
16	123	36	82	27	19	17	8
17	123	35	73	21	14	21	7
18	123	36	51	14	11	25	3
19	123	35	67	17	13	22	4
20	123	27	53	12	10	17	2
21	123	33	61	16	14	19	2
22	123	38	59	18	16	22	2
23	123	40	73	20	15	25	5
24	123	40	60	19	16	24	3
25	123	30	19	5	5	25	0
26	123	38	53	16	12	26	4
27	123	35	71	23	16	19	7
28	123	37	8	1	1	36	0
29	123	37	7	2	1	36	1
30	123	37	71	20	15	22	5
31	123	35	8	1	1	34	0
32	123	39	67	24	19	20	5
33	123	35	84	26	20	15	6
34	123	27	46	8	6	21	2
35	123	21	50	8	8	13	0
36	123	36	84	24	19	17	5
37	123	38	68	22	16	22	6

Table B.1: Sizes of reference sets and identical and different peers in unconditional and conditional models.

Firm	Ref. set uncond.	Number of peers uncond.	Ref. set cond.	Number of peers cond.	Number of identical peers	Number of peers in uncond./ not in cond.	Number of peers in cond./ not in uncond.
38	123	36	58	18	13	23	5
39	123	36	78	25	18	18	7
40	123	34	73	15	13	21	2
41	123	18	51	4	4	14	0
42	123	31	49	12	11	20	1
43	123	26	60	10	9	17	1
44	123	16	47	4	4	12	0
45	123	30	36	12	4	26	8
46	123	35	59	19	14	21	5
47	123	36	63	17	14	22	3
48	123	37	98	36	24	13	12
49	123	33	92	23	17	16	6
50	123	27	61	12	11	16	1
51	123	38	76	27	21	17	6
52	123	30	65	16	12	18	4
53	123	24	60	14	11	13	3
54	123	33	74	17	14	19	3
55	123	37	15	3	1	36	2
56	123	28	47	8	8	20	0
57	123	35	47	11	9	26	2
58	123	37	65	25	17	20	8
59	123	40	81	31	24	16	7
60	123	33	98	27	21	12	6
61	123	25	64	12	11	14	1
62	123	35	106	39	27	8	12
63	123	39	104	29	26	13	3
64	123	40	105	37	31	9	6
65	123	37	63	20	16	21	4
66	123	31	65	9	7	24	2
67	123	37	68	24	13	24	11
68	123	23	57	13	7	16	6
69	123	27	70	22	10	17	12
70	123	32	56	13	12	20	1
71	123	33	39	11	5	28	6
72	123	36	90	34	19	17	15
73	123	38	61	22	17	21	5
74	123	17	53	8	7	10	1
75	123	36	65	20	14	22	6
76	123	35	4	1	1	34	0

Table B.1: Sizes of reference sets and identical and different peers in unconditional and conditional models.

Firm	Ref. set uncond.	Number of peers uncond.	Ref. set cond.	Number of peers cond.	Number of identical peers	Number of peers in uncond./ not in cond.	Number of peers in cond./ not in uncond.
77	123	33	79	23	17	16	6
78	123	37	91	28	22	15	6
79	123	26	65	13	13	13	0
80	123	32	54	13	12	20	1
81	123	35	50	5	5	30	0
82	123	36	67	28	14	22	14
83	123	32	46	13	11	21	2
84	123	32	28	10	5	27	5
85	123	30	56	14	6	24	8
86	123	40	69	20	17	23	3
87	123	33	47	9	8	25	1
88	123	35	81	23	14	21	9
89	123	38	93	29	25	13	4
90	123	36	73	20	14	22	6
91	123	38	62	18	15	23	3
92	123	32	56	10	7	25	3
93	123	39	41	12	6	33	6
94	123	35	60	15	12	23	3
95	123	33	88	24	13	20	11
96	123	25	56	14	7	18	7
97	123	36	62	18	15	21	3
98	123	39	88	29	23	16	6
99	123	37	60	20	17	20	3
100	123	40	65	20	16	24	4
101	123	34	76	23	16	18	7
102	123	31	24	8	5	26	3
103	123	36	44	10	7	29	3
104	123	37	66	23	17	20	6
105	123	30	70	16	9	21	7
106	123	37	64	18	14	23	4
107	123	36	44	13	10	26	3
108	123	37	61	20	16	21	4
109	123	36	50	11	11	25	0
110	123	37	81	28	20	17	8
111	123	35	25	9	4	31	5
112	123	38	97	28	23	15	5
113	123	32	43	12	9	23	3
114	123	25	13	5	5	20	0
115	123	37	62	17	14	23	3

Table B.1: Sizes of reference sets and identical and different peers in unconditional and conditional models.

Firm	Ref. set uncond.	Number of peers uncond.	Ref. set cond.	Number of peers cond.	Number of identical peers	Number of peers in uncond./ not in cond.	Number of peers in cond./ not in uncond.
116	123	36	61	19	14	22	5
117	123	35	62	21	14	21	7
118	123	39	58	20	15	24	5
119	123	32	59	16	12	20	4
120	123	14	16	2	2	12	0
121	123	34	28	10	7	27	3
122	123	35	91	23	19	16	4
123	123	28	76	22	14	14	8
Average		33	60	17	13	21	4
Min		14	4	1	1	8	0
Max		41	106	39	31	36	15