DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Controlling for environmental conditions in regulatory benchmarking

Emil Heesche
Mette Asmild

2020 / 03

IFRO Working Paper 2020 / 03

Controlling for environmental conditions in regulatory benchmarking

Authors: Emil Heesche, Mette Asmild

JEL-classification: C02; C14; C51; C52; C61; C67; L51

Published: March 2020

See the full series IFRO Working Paper here:
www.ifro.ku.dk/english/publications/ifro_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

Controlling for environmental conditions in regulatory benchmarking

Emil Heesche[1], **Mette Asmild**[2]

Abstract

Data Envelopment Analysis (DEA) is often used by regulators to create a pseudo-competitive environment for sectors with natural monopolies. In addition to develop a theoretically well-behaved model, regulators need to take into account several other factors, such as the political agenda and the historical context of the regulation. This sometimes results in some unconventional approaches, which furthermore are not easily changed. In this paper, we discuss the model used for DEA-based benchmark regulation of the Danish water sector. More specifically, we look at the characteristics of the method the regulator uses to take into account differences in the companies' environmental conditions.

We show how the approach currently used to control for differences in environmental conditions seemingly does not sufficiently control for the actual differences as intended since second stage analysis still reveals significant correlations between the efficiency scores and these external factors. To explain this, we reconsider the second stage analysis, using permutation-based approaches and also accounting for the fact that only those companies that in the DEA assign weights to those output measures adjusted for environmental conditions, will benefit from the adjustments.

Keywords Data envelopment analysis; Second Stage Analysis; Environmental Variables; Regulation; Permutation

[1] University of Copenhagen, Department of Food and Resource Economics, Rolighedsvej 25, 1958 Frederiksberg C Denmark. ehe@kfst.dk
[2] University of Copenhagen, Department of Food and Resource Economics, Rolighedsvej 25, 1958 Frederiksberg C Denmark. meas@ifro.ku.dk

1 Introduction

Regulators often use benchmarking models in sectors with natural monopolies such as water distribution, water treatment, electricity networks and gas networks. Without regulation, the monopolies will earn a high profit at the expense of the consumers, which according to standard economic theory is not socio-economically optimal. By using benchmarking, the regulator creates a pseudo-competitive environment in the sector, which forces the companies to become more efficient and reduces prices for the consumers.

The regulatory praxis differs among countries and sectors but the main idea is often the same: Several overviews of the different practices can be found in, for example, Agrell & Bogetoft (2017), Banker, Førsund and Zhang (2017), Benchmarkingekspertgruppen – Energistyrelsen (2017), Haney & Pollitt (2009) and Thanassoulis (2000). In this paper, we will mainly focus on the approach used by the Danish water regulator (KFST), but most of the principles are the same in the majority of the regulators' models across Europe in particular. This could be because it is the best setup for such regulation, but could also be because regulators copy each other and use the same consultants.

KFST use so-called revenue cap regulation, which does exactly what the name indicates: Sets an upper cap for the companies' allowed revenue. The idea is that KFST estimates a revenue cap equal to the expected revenue under perfect competition. It is hereafter up to the individual companies to make sure that their expenses do not exceed this revenue cap to avoid going bankrupt.

The revenue cap is set equal to the revenue cap from the last period but adjusted for inflation, expected industry-wide productivity improvements, new projects and finally the efficiency score. If the companies are not already fully efficient, KFST decreases the revenue cap with an amount dependent on the efficiency score.

The main challenge for KFST is to estimate the efficiency scores. Fortunately, KFST can apply benchmarking methods to do this. Unfortunately, KFST, like other regulators, work in a political world where the theoretically best model is not necessarily the most appropriate. Regulators need to take into account the political agendas, the lobbyists' agendas, the risk of bankruptcy for the regulated companies, which can result in consumers being without water for some time, and lastly, some kind of fairness, for example in the sense that no company can get a bad efficiency score just for being an outlier or because of statistical noise. Also, the starting point and previous development of the regulation is a factor, since changes in regulation, even to a theoretically superior model, can be difficult to implement if it results in just some companies being worse off.

In addition to the above, the use of benchmarking models means assuming that the companies are homogeneous which, especially in the water industry, is questionable. To adjust for differences in the operating environment the regulators tend to use some statistics, often estimated based on regression analysis of potential cost drivers. For these to be precisely estimated requires large datasets and data that do not vary a lot over time. In Denmark, there are only 74 drinking water companies included in the benchmarking model, which limits the amount of data the regulator can use in their analysis. It is therefore not possible for KFST to include all the environmental conditions when adjusting for the heterogeneity. The data furthermore varies over time because the companies' costs and production change a lot from year to year – for example due to heavy rain, an incident at a water plant or a pipeline breakage. In combination with the relatively small number of companies, this leads to results that may vary a lot over time for an individual company, which is not desirable for a regulator.

In terms of the Danish regulation of the water companies, KFST needs to control for especially two environmental conditions: The average age of the companies' assets (hereafter age) and the customer density (hereafter density) in the areas in which they operate. The reason for this is that companies with old assets tend to have higher costs of maintaining these. However, they cannot themselves decide the age because their assets have to follow the development of the areas in which they operate. The reasoning behind needing to control for the density is that it is expensive to operate in areas with high density. It will, for example, be more expensive to lay down pipes in the middle of a big city, where you have to stop and redirect traffic, dig up the road without damaging any other installations utilizing the same space and so on.

The purpose of this paper is to critically examine KFST's current benchmarking model in light of the specific context described above. We find that the regression models, currently used to adjust for environmental differences, have problems with misspecification and that the combined benchmarking model does not properly account for differences in environmental conditions. Furthermore, since the DEA models include heavily correlated outputs, they have some interesting characteristics which KFST needs to be aware of when using the results for further analyses. These include the relationship between the dimensionality and the number of peers and mean efficiency, the relative rankings of the companies and the interpretation of the results.

We propose several alternative methods to handle environmental conditions, which overcome some of the issues mentioned above. The new methods are, however, not without their problems. In addition, we develop a new method for using regressions as a second-stage analysis to assess if KFST's methods for handling environmental variables work. Using this new method, we realize that it is not necessarily a problem that the efficiency scores still depend on environmental conditions in a model where several precautionary measures influence the results.

The rest of this paper is structured as follows: Section 2 describes KFST's current benchmarking model. We explain the method where we focus on how they adjust for age and density. In Section 3 we discuss some of the characteristics of KFST's model due to their way of accounting for age and density. In addition, we question if their method does, in fact, sufficiently account for age and density. In Section 4 we propose an alternative model where the goal is to keep the same characteristics, but still improve the model. Section 5 concludes the paper. Note that all of the examples and discussions will focus on the drinking water sector, but KFST uses the same model setup for the waste water sector.

2 The current benchmarking model in the Danish water industry

KFST uses both a Data Envelopment analyses (DEA) model (Charnes, Cooper, & Rhodes, 1978) and a stochastic frontier analysis (SFA) model (Meeusen & Broeck, 1977, Aigner, Lovell, & Schmidt, 1977) to find the companies' efficiency scores. The higher of the two efficiency scores for each company is being used to set the revenue cap. This is a precaution against model uncertainty, which the regulator takes, to make sure that no companies can blame their efficiency scores on the choice of the modelling approach. In this paper, we will only focus on the DEA model and leave the discussion of the SFA model to future research.

Ideally, the benchmarking model should estimate the efficiency score by looking at the amount of resources used by the companies to deliver a given amount of water to the consumers. This would be a single input, single output model, where the input is the companies' total controllable cost (hereafter total costs) and the output is the total quantity of water delivered. This will, however, not result in fair benchmarks because of various external conditions, which the companies have to organize themselves according to. This can, for example, be how deep down or how clean the subsoil water is, how the soil conditions are, how much water is delivered to the industry, the length of pipes in different dimensions needed to deliver the water in the given area etc. The list of potentially relevant factors is simply too long to incorporate all these as separate variables in a DEA model with only 74[3] observations.

KFST has therefore chosen to aggregate all those production factors for which they have data. To aggregate the factors KFST uses a common unit of measurement, which is monetary. KFST have therefore defined a standardized unit price for each of these factors and combined them into a so-called net volume measure, used as an output in the DEA model instead of the delivered quantity of water. We can think of it as the capacity, which the companies make available for the consumers. If, for example, a company needs to get their water from deeper down than the other companies, they need bigger pumps, which will give them a higher net volume (capacity) measure and thereby count as a higher output quantity in the DEA model. The unit prices also account for different scale sizes, in the sense that having two identical components is, for example, not necessarily twice as expensive as having one component.

Furthermore, besides utilizing this, in DEA perhaps rather unconventional, output measure (a standardized cost of installed capacity), KFST has decided to split it in two: One net volume measure where the components are aggregated using operational costs as the unit prices (OPEX net volume) and one with the capital costs as the unit prices (CAPEX net volume). The reason for this split is that KFST does not have all the unit prices for the total costs needed to define a single net volume measure (TOTEX net volume). One could argue that the TOTEX unit prices are just the sum of the OPEX and CAPEX unit prices. The problem with this is that the unit prices for OPEX and CAPEX are calculated differently because they were originally defined for different purposes. If one method systematically gives higher unit prices than another, the TOTEX net volume will favour either OPEX or CAPEX over the other. To define new TOTEX unit prices is costly and is, therefore, not likely to been done if it is not absolutely necessary.

Second-stage analysis of the efficiency scores from a DEA model with total costs as input and OPEX net volume and CAPEX net volume as outputs have revealed that there are still two environmental conditions, which significantly influence the efficiency scores: The average age of the production assets and the density (cf. Table 11 in Appendix). KFST can, however, not control for them in the same way as some of the other differences in the operating environment since they do not result in differences in the used production factors that contribute to the net volume (capacity) measure. Age and density should, therefore, be included in the efficiency assessment in another way. To do this one could try to include them as separate variables. This, however, raises two problems: Firstly, age and density are intuitively not meaningful outputs as they are non-discretionary conditions (Ruggiero, 1998). They are, therefore, not something the companies produce but are rather characteristics of the environment in which the companies operate. Secondly, age and density do not scale in the same way as the other parameters (Dyson, et al., 2001). To illustrate this problem, consider the example shown in Table 1:

[3] KFST define four companies as outliers. The total number of companies in the analyses is therefore 70.

Table 1 Illustrative example of the problem with using both a volume measure and an index (age) as output variables

	Input	Net volume (output)	Age (output)	Efficiency scores (CRS)	Efficiency scores (VRS)
Company 1	100	100	20	1	1
Company 2	40	10	9	1	1
Company 3	50	20	10	0.91	0.93
Company 4	100	40	10	0.49	0.60

In the example in Table 1, company 1 and 2 are fully efficient in both CRS and VRS and are therefore peers for company 3 and 4. Company 3 and 4 should be equally efficient because company 4 is simply twice the size of company 3 (but with the same age). If, however, age is included as an output, then company 4 gets a lower efficiency score than company 3, regardless of whether a CRS or a VRS model is used.

KFST have attempted to solve this problem by adjusting the net volume measures based on age and density. Age and density will therefore not be separate outputs measures but will instead adjust the net volume measures, such that companies with high age and density will get higher adjusted net volume measures etc. Let $OPEX\ Costs$ be the companies' actual operational costs and $CAPEX\ Costs$ be the actual capital costs. KFST then makes the adjustments by using regression analyses as follows:

$$\frac{OPEX\ Costs}{OPEX\ net\ volume} = \hat{\alpha}_1 + \hat{\beta}_1 \cdot Age + \epsilon \Rightarrow$$

$$Age\text{-}adjusted\ OPEX\ net\ volume = \left(\hat{\alpha}_1 + \hat{\beta}_1 \cdot Age\right) \cdot OPEX\ net\ volume \qquad (1)$$

Similarly,

$$Age\text{-}adjusted\ CAPEX\ net\ volume\ = \left(\hat{\alpha}_2 + \hat{\beta}_2 \cdot Age\right) \cdot CAPEX\ net\ volume \qquad (2)$$

$$Density\text{-}adjusted\ OPEX\ net\ volume = \left(\hat{\alpha}_3 + \hat{\beta}_3 \cdot Density\right) \cdot OPEX\ net\ volume \qquad (3)$$

$$Density\text{-}adjusted\ CAPEX\ net\ volume = \left(\hat{\alpha}_4 + \hat{\beta}_4 \cdot Density\right) \cdot CAPEX\ net\ volume \qquad (4)$$

These four new net volume measures (given by equations (1) - (4)) are then included as separate outputs in the benchmarking model[4], together with the two non-adjusted net volume measures. The latter is kept as a precaution to make sure that no companies exclusively can blame bad efficiency scores on the choice of regression assumptions like, for example, the functional form. If the companies do not believe that the correlation between age (density) and costs match their specific characteristics, the model can instead assign a higher weight to the non-adjusted net volume measures. This way the companies can either benefit from the adjustments or ignore them. There are, therefore, six outputs in the model and one input.

Lastly, the benchmarking model for drinking water in Denmark is input-oriented and uses constant returns to scale. It is input-oriented because the companies cannot change their outputs: By law, they have to deliver water to all the consumers and should, of course, not deliver more water than the consumers demand. The constant returns to scale assumption is often up for debate, but the argument for this is that

[4] Note that KFST only includes those adjusted net volumes where β_1 is significant which is for three out of the four adjusted net volume measures. For simplicity, we will, however, ignore this point in the discussions.

the unit prices used in the net volumes already account for different scales as mentioned earlier. The models using net volume measures as output are thereby supposed to be neutral to different scale sizes. Another argument for using CRS models is that it is in the consumer's interest that the companies operate at the optimal scale size, and therefore the efficiency scores should reflect this.

Having now described the benchmarking model for the Danish water sector, we will discuss some of the characteristics of the model in the next section.

3 Discussion of the Danish water benchmarking model

The Danish water benchmarking model has several characteristics which are not typically seen in the DEA literature. In this section, we will describe and discuss those that are specifically associated with the environmental conditions, which are:

1. Too much benefit of the doubt
2. Model misspecifications in the regression analysis
3. Correlated outputs
4. Ranking
5. Second-stage analysis
6. Number of positive multiplier weights

In the following, we discuss each of these characteristics in turn.

3.1 Too much benefit of the doubt

Individually, each precautionary measure in the Danish water benchmarking model seems somehow valid, but when adding them all together, one might argue that the consumers pay too high a price compared to what is necessary. The main reason for the precautions is that KFST wants to reduce the risk of setting the revenue cap too low for one or more companies. A low revenue cap could be due to the companies not being as homogenous as expected or due to statistical noise. KFST, therefore, uses several precautions distributed over multiple steps in the benchmarking process. One of these precautions is, for example, the inclusion of the various adjusted net volume measures as separate outputs discussed in the previous section. The argument for all these precautions is that there is a probability that a specific step is unfair to at least one company. When this probability is deemed too high, KFST adds a precaution to the model.

The problem is that the extra precaution is applied to all companies and not only the ones where it is needed – KFST cannot distinguish between which companies should get the benefit of the doubt and which ones should not. When KFST use precautions in several steps of the model, chances are that the individual steps compensate a few companies appropriately but some companies more than needed. And perhaps even worse is that if it is not the same company that needs precautions in every step, all companies could end up being overcompensated in the final model.

KFST can, therefore, consider removing some individual precautions and instead look at the combined model when giving the companies some benefit of the doubt. This will, of course, make some steps seem unfair, but it will still reduce the risk that any company will get an unfair efficiency score in the final model even if it tends to be the same companies that are problematic in the different steps. If possible, KFST could consider removing all precautions in the model to get a more precise efficiency score, and hereafter add a combined precaution when setting the revenue cap. How such a precautionary measure can be created is a topic for further research.

3.2 Misspecification

KFST have used simple linear regressions to create the adjusted net volumes. However, they do not document that this is the correct specification or that they have even tested others. We have therefore tried several different functional forms and show the most interesting ones here. To compare the models we do, of course, want the standard MLR axioms to hold (linear in parameters, random sampling, no perfect collinearity, zero conditional mean, homoscedasticity and normality) [5]. In addition to these, we will appreciate if:

1. The functional form makes intuitive sense for the water companies
2. The results and statistical properties are consistent over time, so the regulator does not have to change the functional form every year.

We have summarized the results of KFST's regressions in Table 2.

***Table 2** Regression summary for KFST's adjusted net volumes*

	Models			
	OPEX		CAPEX	
	Age adj.	Density adj.	Age adj.	Density adj.
Constant	-0.424	-0.204	0.974	0.781
	$p = 0.046^{*}$	$p = 0.00000^{***}$	$p = 0.00000^{***}$	$p = 0.000^{***}$
Age	0.011		-0.003	
	p = 0.082		p = 0.586	
Density		2.116		1.407
		$p = 0.00001^{***}$		$p = 0.0003^{***}$
AIC	-11.09	-31.24	-33.08	-46.54
Shapiro-Wilk normality test	0.14275	0.32335	0.00624	0.01728
Observations	70	70	70	70
R^2	0.044	0.283	0.004	0.179
Adjusted R^2	0.030	0.272	-0.010	0.166
Residual Std. Error (df = 68)	0.217	0.188	0.186	0.169
F Statistic (df = 1; 68)	3.118	26.842^{***}	0.301	14.777^{***}
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

[5] See e.g. Wooldridge (2006), page 157-158

The first two columns show the regression results for the Age-adjusted OPEX and Density-adjusted OPEX respectively. The last columns show results for the Age-adjusted CAPEX and Density-adjusted CAPEX respectively. The results are all based on the simple linear regressions (1) - (4).

We observe several issues with these results. First, age is not significant at the 5 % level in either model. We will later show that age is important with another functional form even while it is not significant. Second, the adjusted R^2 in the age-adjusted models is very low. This indicates that the models do not fit the data very well. Third, the Shapiro-Wilk normality test shows that the residuals are not normally distributed in neither of the two CAPEX models.

To overcome some of these issues we have tried different functional forms in the regression analysis. We find that neither models with logarithms, non-linear functions of the independent variables, or allowing for different slopes and intercepts for companies with respectively low/high age/density did a much better job than the simple linear models. We find, however, that there is strong evidence of an interaction between age and density. When KFST is separating the age-adjustment and the density-adjustment in different regression models, it is likely to cause an omitted variable bias, especially because age and density are correlated (with a correlation coefficient of 0.41). We can, therefore, argue that the adjustments for both age and density should be done within one regression model, which gives a single adjusted net volume measure. If we do this, we can, in addition, add an interaction term into the regression model.

In Table 3 we show a summary for the age- and density-adjusted OPEX net volume (hereafter adjusted OPEX) and the age- and density-adjusted CAPEX net volume (hereafter adjusted CAPEX) with and without interaction between age and density. We observe that even while age is still not significant in the adjusted OPEX model, the interaction between age and density is. The interaction is, however, not significant in the adjusted CAPEX model and can, therefore, be removed. The argument for keeping it could be that it is preferable that the adjusted OPEX and adjusted CAPEX are similar. In addition, we observe that the adjusted R^2has substantially increased compared to the models in Table 2. However, the Shapiro-Wilk normality test still finds that neither the CAPEX models nor the OPEX models without interaction satisfy the assumption of normal distributed residuals.

Table 3 *Regression summary of alternative models to adjust the net volumes using simple linear regressions*

	Models			
	Adj. OPEX		Adj. CAPEX	
	Without inter.	With inter.	Without inter.	With inter.
Constant	0.880	0.339	1.170	1.458
	p = 0.00001***	p = 0.276	p = 0.000***	p = 0.00001***
Age	-0.002	0.014	-0.012	-0.021
	p = 0.731	p = 0.143	p = 0.017*	p = 0.015*
Density	2.221	9.232	1.792	-1.939
	p = 0.00001***	p = 0.007**	p = 0.00002***	p = 0.517
Age*Density		-0.192		0.102
		p = 0.036*		p = 0.211
AIC	-33.32	-36.05	-50.55	-50.23

Shapiro-Wilk normality test	0.03294	0.23142	0.01427	0.00497
Observations	70	70	70	70
R^2	0.305	0.350	0.246	0.264
Adjusted R^2	0.284	0.320	0.224	0.231
Residual Std. Error	0.184 (df = 67)	0.179 (df = 66)	0.163 (df = 67)	0.162 (df = 66)
F Statistic	14.667*** (df = 2; 67)	11.844*** (df = 3; 66)	10.939*** (df = 2; 67)	7.892*** (df = 3; 66)
Note:				*p<0.05; **p<0.01; ***p<0.001

If we cannot accept that the residuals are not normally distributed, we can change the functional form to a log-log model with interaction. This will, however, be at the cost of a much worse adjusted R^2 for especially the CAPEX models (cf. Table 4).

Table 4 *Regression summary of alternative models to adjust the net volumes using log-log linear regressions*

	log-log models			
	Adj. OPEX		Adj. CAPEX	
	Without inter.	With inter.	Without inter.	With inter.
Constant	0.681	2.952	1.639	-3.950
	p = 0.350	p = 0.274	p = 0.033*	p = 0.149
Age	-0.062	-0.709	-0.415	1.178
	p = 0.751	p = 0.354	p = 0.044*	p = 0.130
Density	0.178	1.044	0.115	-2.018
	p = 0.00001***	p = 0.292	p = 0.002**	p = 0.047*
Age*Density		-0.248		0.610
		p = 0.381		p = 0.036*
AIC	-33.63	-32.45	-28.32	-31.06
Shapiro-Wilk normality test	0.73476	0.77676	0.14935	0.06418
Observations	70	70	70	70
R^2	0.327	0.334	0.151	0.206
Adjusted R^2	0.307	0.304	0.125	0.170
Residual Std. Error	0.184 (df = 67)	0.184 (df = 66)	0.191 (df = 67)	0.186 (df = 66)
F Statistic	16.249*** (df = 2; 67)	11.056*** (df = 3; 66)	5.944** (df = 2; 67)	5.718** (df = 3; 66)
Note:				*p<0.05; **p<0.01; ***p<0.001

KFST should in the light of the above create a combined age- and density-adjusted net volume, for OPEX as well as for CAPEX, instead of two separate adjusted net volume measures for each. The functional form is, however, not easily determined, but we have discussed a few possibilities. By combining the age- and density-adjustment, the model will go from six to four outputs whilst still controlling for the same factors. This will probably reduce the number of fully efficient companies in the DEA analysis.

3.3 Correlated outputs

KFST's current benchmarking model contains both the non-adjusted- and adjusted net volumes. Equations (1)-(4) show that the adjusted net volumes are functions of the non-adjusted net volumes. The correlations between the non-adjusted- and adjusted net volumes are therefore extremely high; all over 0.987. However, even while the high correlations would indicate that there is almost no extra information available in the adjusted net volumes, the efficiency scores still depend on quite a lot of them (with the highest reduction being 3.3 percentage points if the adjusted net volumes is omitted). They are therefore not easily removed without opposition from the industry, especially not when they intuitively should be incorporated in the model (cf. equation (1) - (4)) and because we observe a significant correlation between the efficiency scores and age as well as density if they are omitted (cf. section 2).

A high correlation between outputs is problematic due to the assumption of free disposability. If two or more outputs are correlated, we cannot, holding everything else equal, reduce one output without reducing the other(s) (Mehdiloo & Podinovski, 2019). This means that the part of the technology which is defined based on the free disposability between the correlated outputs is questionable. Companies which are projected to a facet in this area of the technology are likely to get an unfairly low efficiency score. Mehdiloo & Podinovski, 2019 propose to use weak disposability instead of strong disposability to overcome this problem. We will, however, not go into details regarding this assumption because the problem only arises due to KFST's precaution of having both the non-adjusted net volumes and the adjusted net volumes as outputs in the model. It is only a company which benefits from the precaution that are projected to this questionable part of the facets, and the efficiency scores will, therefore, increase more due to the precaution than they will decrease due to the assumption not being satisfied. That the free disposability assumption is not satisfied therefore just means that the precaution does not have as strong an effect on the efficiency scores as one might expect.

It is also worth noticing that a model with correlated outputs tend to give fewer peers than one might expect given the dimensionality. A peer company that is doing well in one dimension is likely to also do well in all correlated dimensions.

In addition to this, the efficiency scores will be lower than expected, given the dimensionality. Here the argument is the opposite. The companies will not benefit as much (in terms of higher efficiency scores) if a correlated variable is added compared to an uncorrelated variable. Intuitively, it is the uncorrelated part of the extra variable that gives information to the model and thereby is likely to increase the number of peers and the efficiency scores.

Therefore, we will not expect the average efficiency scores to increase as much when also including the four heavily correlated outputs into the model with the existing two outputs (OPEX net volume and CAPEX net volume) compared to if the additional outputs were not as correlated. For some companies (the companies with high age and/or density), the efficiency scores still depend on the correlated outputs, cf. section 4.

3.3.1 Outputs are functions of other outputs

In the Danish water regulation models the outputs are not just extremely correlated; one is, in fact, a function of the other. This gives rise to some additional reflections. For simplicity, we illustrate this in an input-oriented DEA model with constant return to scale and 1 input (actual OPEX costs, x_i), 2 outputs (non-adjusted ($y_{i,1}$) and age adjusted ($y_{i,2}$) OPEX net volume), and 2 companies, of which one ($x_1, y_{1,1}, y_{2,1}$) is efficient. We estimate the efficiency for the other ($x_0, y_{0,1}, y_{0,2}$), which is inefficent and therefore the constraint for this observation will not be binding and is ignored in the following. The equation for the age-adjusted OPEX net volume (1) is inserted into the dual formulation of the model and the constraints rearranged as follows:

$$f(v_1, v_2) = \max y_0(v_1 + v_2(\alpha + \beta \cdot Age_0)) \qquad (5)$$

$$s.t.\ y_1\big(v_1 + v_2(\alpha + \beta \cdot Age_1)\big) \leq \frac{1}{x_0} x_1 = K \qquad (6)$$

Note that y_i now refers to the non-adjusted net volume $y_{1,i}$. We know that (6) is binding. We can, therefore, set the constraint equal to K, rewrite (6) for v_1 and v_2 respectively and insert the solutions in (5):

$$f(v_1) = \max y_0 \left(v_1 + \left(\frac{K}{y_1(\alpha + \beta \cdot Age_1)} - \frac{v_1}{(\alpha + \beta \cdot Age_1)} \right) (\alpha + \beta \cdot Age_0) \right)$$

$$f(v_2) = \max y_0 \left(\frac{K}{y_1} - v_2(\alpha + \beta \cdot Age_1) \right) + y_0 v_2 (\alpha + \beta \cdot Age_0)$$

By taking the first derivative of $f(v_1)$ and $f(v_2)$ we examine how much v_1 and v_2 contribute to the object function. The one that can contribute the most will be positive and the other equal zero. We find that

$$\frac{df(v_1)}{dv_1} \geq \frac{df(v_2)}{dv_2} \Leftrightarrow$$
$$Age_0 \begin{cases} \leq Age_1, & if \beta \geq 0 \\ \geq Age_1, & if \beta \leq 0 \end{cases} \qquad (7)$$

It is therefore the age of the assessed company relative to the age of the peer company that, together with the sign of β, determines whether a company has positive multiplier weight on the non-adjusted net volume or the age-adjusted net volume. More generally, we can say that the assessed company weighs the unadjusted net volume if it has a better environment than the peer company does and vice versa.

If there are several peer companies, the age of the assessed company should be compared to the ages of all the peers. By the same logic as above, it can be shown that If age_0 is greater than the age of all the peers or less than the age of all the peers the conclusion from (7)is still the same. If, however, age_0 is lower than than the age of some peers but higher than the age of others, both v_1 and v_2 will be positive. The intuition for this is that v_1 increases the objective function the most compared to the peer(s) with a higher age, but v_2 increases the object function the most compared to the peer(s) with a lower age. A combination of v_1 and v_2 will, therefore, maximize the program. This is illustrated in Figure 1. If an inefficient company is located above the striped line, like company C, it has a better environment than both company A and B. It will therefore only have a positive multiplier for the non-adjusted net volume. If a company is located between the striped and the dotted line, like D, it has a better environment than company A but worse environment than company B. It will therefore have positive multiplier weights on both the adjusted- and non-adjusted net volume. Lastly, if a company is below the dotted line, like company E, it has a worse environment than both company A and B and will therefore only have a positive multiplier weight on the adjusted net volume.

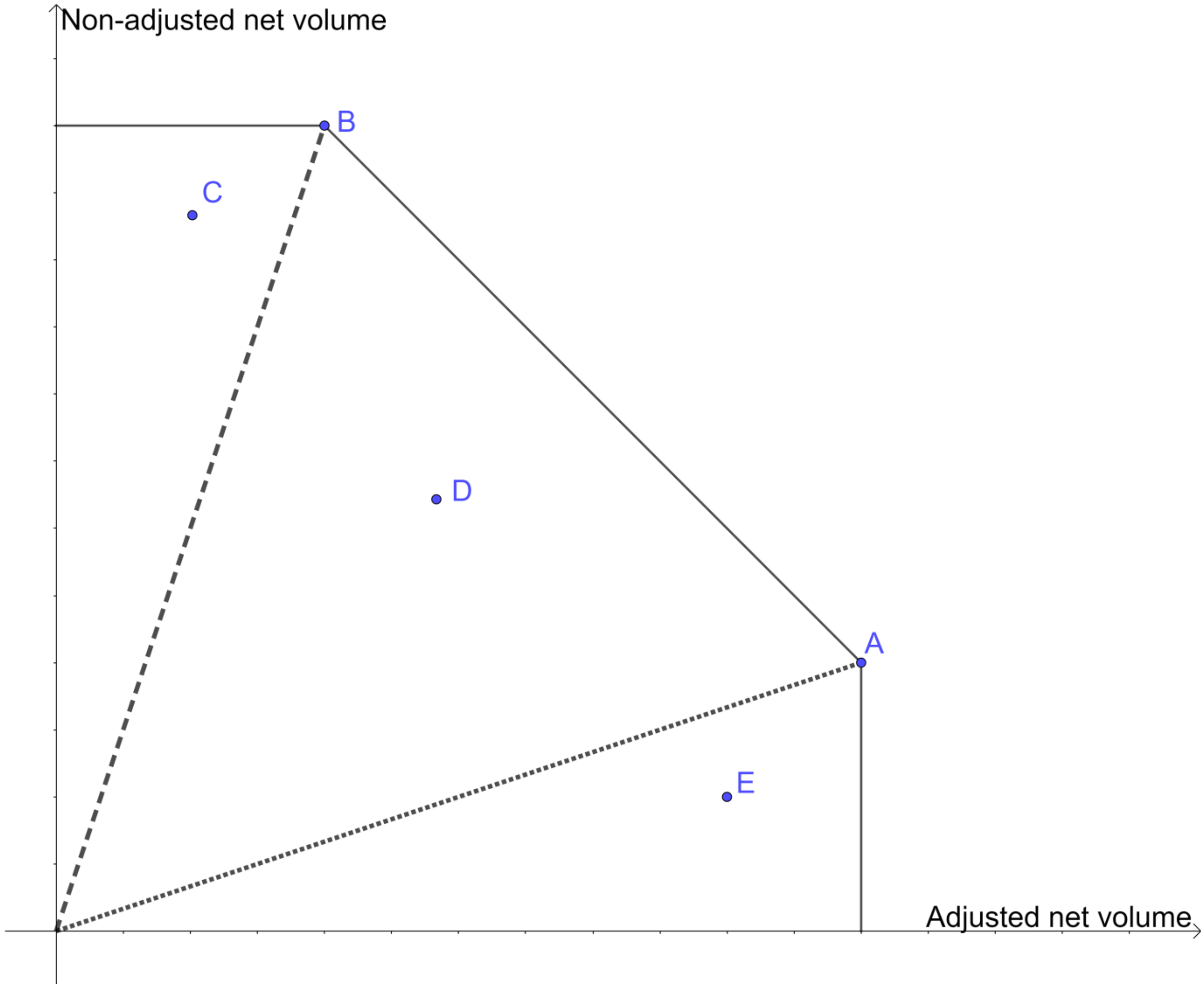

Figure 1 *Illustrative example of the relationship between a company's peers and age (density).*

Intuitively it make sense that the companies with the worst environment want to be compensated through the adjusted net volumes and that companies with good environments want the adjustments to be ignored. It is, however, interesting that the adjustments only have an effect for some companies and that whether they do so depends on the environment of the peers.

When we add more dimensions, the logic applies the peers spanning the facets on which the inefficient company is projected. If a peer, for example, is doing better than the evaluated company on the CAPEX net volume but not on the OPEX net volume, this peer's environment will not decide whether the evaluated company should have positive weight on the adjusted OPEX net volume or not. It will, however, decide if the evaluated company should have positive weight on the adjusted CAPEX net volume or not.

3.4 Ranking

Another problematic characteristic in KFST's current benchmarking model is that we cannot directly rank the companies, which have positive multiplier weight on the non-adjusted net volumes. This problem is not specific to KFST's benchmarking model (cf. Asmild, Hougaard, & Kronborg, 2013) but is enhanced due to the inclusion of the non-adjusted net volume. Even while company C in Figure 2, for example, has a higher efficiency score than company E, we cannot really say which one is the most efficient. Recall from section 3.3.1 that company C has a better environment than company E. The reason why company C seems more efficient than E could, therefore, be the better environment, which the non-adjusted net volume does not penalize. It is, therefore, only possible to rank companies in two scenarios: First, we can rank companies, which have zero multiplier weight on the non-adjusted net volumes. Second, we can rank two companies if the company with zero multiplier weight on the non-adjusted net volume have a higher efficiency score than the company with positive multiplier weight on the non-adjusted net volume. This scenario requires, however, that we do not doubt the relationship between the efficiency scores and age. If we do doubt this, the company with positive multiplier weight on the non-adjusted net volume can claim that the adjusted net volume compensate a high age too much.

3.5 Second-stage analysis

KFST added the adjusted net volumes because the efficiency scores otherwise were found to be significantly correlated with age and density (cf. section 2). The question is, therefore, whether KFST's inclusion of the adjusted net volumes in the DEA model has solved the problem.

To examine this, we use regressions in a second-stage analysis to check if the efficiency scores still depend on age and density after the inclusion of the adjusted net volumes in the DEA model. The use of regressions as second-stage analyses in the DEA framework is, however, problematic (cf. Simar & Wilson, 2007). However, they can still give an indication, if not an exact statistical conclusion. To overcome some of these problems, we use three different methods. Standard TOBIT regression (Tobin, 1958), permutation regression (Anderson & Robinson, 2001, Oja, 1987, Winkler et al. 2014) and permutation regression on efficiency scores calculated by an Order-M DEA model (Cazals, Florens, & Simar, 2002).

The standard TOBIT regression is often used in DEA models because it accounts for the efficiency scores being censored with a value of maximum one (Bogetoft & Otto, 2011). We use the permutation regression to account for the DEA scores not being normally distributed. Permutation regressions are non-parametric and are, therefore, not constrained to the F or t statistics (Winkler et al, 2014). This means that permutation regressions do not require the assumption of normality to be fulfilled. Lastly, we use the order-M efficiency scores to reduce the problem with the observations not being independent. Standard DEA scores depend on the peers. Order-M DEA does not completely solve this problem, but it will spread the dependency to several peers instead of the few peers from a standard DEA.

3.5.1 Second-stage analysis of KFST's current model

We use the efficiency scores from the current model (with 6 outputs) as the dependent variable and age and density as well as their interaction as explanatory variables in the three regression methods described above. The results are shown in Table 5. The results show that the current model does not account properly for age and density no matter if we use TOBIT-, permutation- or order-M regressions. All methods show that there is a significant correlation between the efficiency scores and density, as well as the interaction between age and density.

Table 5 Regression summary of a second stage analysis of KFST's current model

	Method		
	TOBIT	Perm	Order-M Perm
Constant	1.062	0.829	0.852
	$p = 0.000^{***}$	$p = 0.000^{***}$	$p = 0.000^{***}$
Age	-0.006	0.003	0.003
	p = 0.249	p = 0.283	p = 0.251
Density	-4.822	-0.598	-0.659
	$p = 0.008^{**}$	$p = 0.028^{*}$	$p = 0.016^{*}$
Age*Density	0.127	0.109	0.112
	$p = 0.011^{*}$	$p = 0.024^{*}$	$p = 0.020^{*}$
logSigma	-2.357		
	$p = 0.000^{***}$		
Observations	70	70	70
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

It is not surprising that the results above show significant correlations between the efficiency scores and age and density (through the interaction). Besides the possible misspecification described in section 3.2, the distribution of multipliers from the dual DEA program influences the results. The intuition for this is shown in Figure 2. Assume, for simplicity, that we have a model with only two outputs: A non-adjusted net volume and the density-adjusted net volume. Some companies will have a positive multiplier corresponding to the density-adjusted net volume, while other companies will ignore the density-adjusted net volume in the efficiency assessment and thereby give the corresponding multiplier zero weight (cf. section 3.3.1). In this scenario, we still expect the efficiency scores to be negatively correlated with density for the companies with zero weight on the multipliers for the density-adjusted net volumes (which we in the following will denote the non-density-neutral companies) and non-correlated for the companies with positive weight on the multipliers for the density-adjusted net volume (which in the following is denoted the density-neutral companies).

We showed in (7) that the companies are density-neutral if their density is higher than that of at least one of their peers. Assume there are two efficient companies in the benchmarking model. The companies with lower density than both these companies will have a negative correlation between the efficiency scores and density, because they are non-density-neutral. The companies with a density higher than both efficient companies will not have any correlation between the efficiency scores and the density, because they are density-neutral. The companies with a density between the two efficient companies will have positive weight on the multipliers for both the non-adjusted net volume and the adjusted net volume. They will thereby have a correlation between the efficiency scores and the density that is less than the non-density-neutral companies are but higher than the density-neutral companies are. In Figure 2, this group will have a correlation in the gray area.

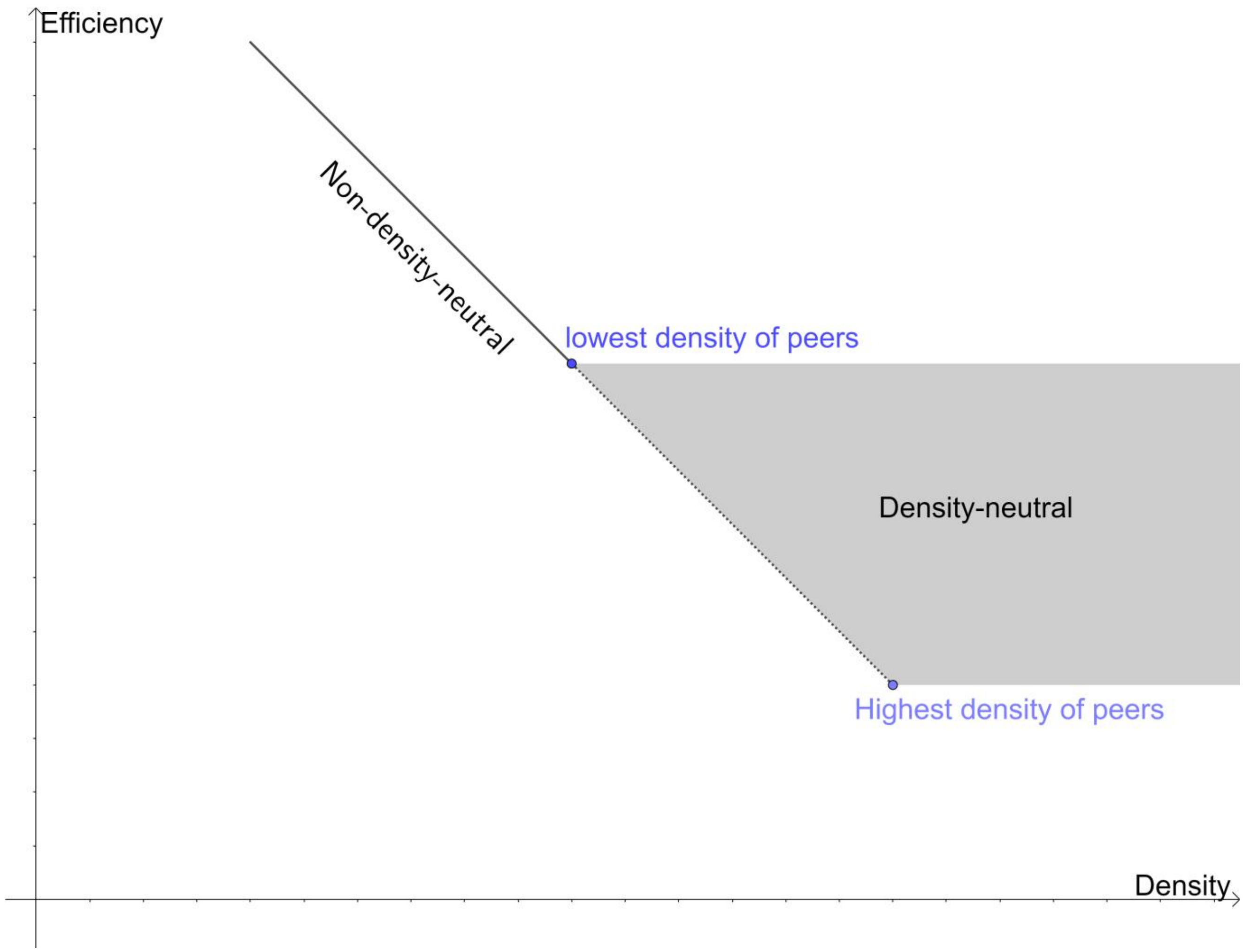

Figure 2 *Illustrative example showing when the companies are density-neutral or not*

By this logic, we will expect the results from Table 5 to be significant if we observe many companies that are non-age (density)-neutral. That is if the peers in the model have high ages and densities. To find out if the results are significant because, some companies ignore the adjusted net volumes, we need to explicitly account for this characteristic.

We, therefore, incorporate whether the companies are age (density)-neutral or not in the regressions[6]. Table 6 shows that 55 companies are not age-neutral (have zero weight on both the age adjusted measures) and 16 companies are not density-neutral (have zero weight on both the density adjusted

[6] Note that we do not distinguish between if the companies are fully age (density)-neutral or if they have positive weights on the multipliers for both the adjusted net volumes and the non-adjusted net volumes. This is a topic for further research.

measures), which might explain why there are still significant correlations between the efficiency scores from KFST's current model and age (density)-neutral in Table 5.

Table 6 *Number of companies, which are age (density)-neutral and non-age (density)-neutral*

	Non-neutral companies	**Neutral companies**
Age-adjusted net volumes	55	15
Density-adjusted net volumes	16	54

To formally control for this in the second-stage regression, we use the following model with a series of dummy variables:

$$Eff = \hat{\alpha} + \hat{\beta}_1 \cdot Density_{0,1} + \hat{\beta}_2 \cdot Density + \hat{\beta}_3 \cdot Density_{0,1} \cdot Density + \hat{\beta}_4 \cdot Age_{0,1} + \hat{\beta}_5 \cdot Age + \hat{\beta}_6 \cdot Age_{0,1} \cdot Age + \hat{\beta}_7 \cdot AgeDensity_{0,1} + \hat{\beta}_8 \cdot AgeDensity_{0,1} \cdot Age + \hat{\beta}_9 \cdot AgeDensity_{0,1} \cdot Density + \hat{\beta}_{10} \cdot Age \cdot Density + \hat{\beta}_{11} \cdot AgeDensity_{0,1} \cdot Age \cdot Density + \epsilon \quad (8)$$

where $Density_{0,1}$ is a binary variable taking the value of 0 if the company is density-neutral and 1 otherwise. The same principle goes for $Age_{0,1}$. $AgeDensity_{0,1}$ is a binary variable taking the value of 0 if the company is both age- and density-neutral and 1 otherwise. By doing this, we can observe if age and density have different influence on companies which weight the adjusted net volumes against companies that do not. Note that the dummy variables are defined opposite of what one may think intuitively. This is because we are mainly interested in the p-values for the group, which is age (density)-neutral. The p-values for the group where the dummies are equal to 1 (the non-age (density)-neutral companies) cannot be interpreted on their own. They should instead be interpreted as the probability that the coefficients are different from the coefficient for the group with the age (density)-neutral companies.

We present the regression summary from (8) in Table 7. For simplicity, we only show the coefficients that are significant or in other ways give us important information. In some models, the final set of variables depend on which variables are to be omitted first. In these situations, we present multiple scenarios.

Table 7 *Regression summary for a second stage analysis on KFST's current model when controlling for whether the companies are age (density)-neutral or not*

	Method					
	TOBIT	TOBIT	Perm	Perm	Order-M Perm	Order-M Perm
Constant	0.063	0.814	0.906	0.866	0.931	0.884
	p = 0.850	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
$Age_{0,1}$	0.561		0.080		0.083	
	p = 0.114		p = 0.008**		p = 0.006**	
Age	0.028		0.016		0.016	
	p = 0.016*		p = 0.009**		p = 0.008**	
$Age_{0,1}$* Age	-0.022		0.011		0.011	

	p = 0.070		p = 0.083		p = 0.080	
$Density_{0,1}$		0.047		-0.041		-0.035
		p = 0.508		p = 0.365		p = 0.443
Density		0.150		0.281		0.096
		p = 0.535		p = 0.793		p = 0.929
$Density_{0,1}$* Density		0.999		-0.207		-0.071
		p = 0.669		p = 0.846		p = 0.947
LogSigma	-2.368	-2.336				
	p = 0.000***	p = 0.000***				
Observations	70	70	70	70	70	70
Note:						*p<0.05; **p<0.01; ***p<0.001

The results in Table 8 shows that age is still significantly correlated with the efficiency scores even when we control for whether the companies are age (density)-neutral or not (as indicated by the Age variable). This means that the model does not fully account for age as expected by KFST. Furthermore, we observe that the companies, which are not age-neutral, have a significantly different slope and intercept at a 10 % significant level[7] in the permutation regressions than the age-neutral companies (as indicated by the $Age_{0,1}$ dummy). The efficiency scores, therefore, depend on whether a company is age-neutral or not. In addition, the results show that there is no longer any significant correlation between the efficiency scores and density. There is, however, not a significant difference between the density-neutral companies and the non-density-neutral companies either (as indicated by the $Density_{0,1}$ dummy). This means that we cannot explain the significant results from Table 5 by controlling for if the companies are density-neutral or not. The model does, therefore, not fully account for density as expected by KFST.

To better control for age and density we, therefore, propose to use the adjusted net volumes from section 3.2 in the benchmarking model instead of the current construction. We show the results from doing this in the following section.

3.5.2 Second-stage analysis on a model with proposed adjusted net volumes

In this section, we together with the non-adjusted net volumes include the adjusted net volumes from section 3.2 based on linear regressions but with interaction between age and density (c.f. Table 3) in the DEA model, instead of the adjusted net volumes from KFST's current benchmarking model. The results for this are shown in Table 8.

[7] Note that only 15 companies are age-neutral. We will, therefore, argue that using a 10 % significant level is acceptable.

Table 8 Regression summary for a second stage analysis on the model with the new proposed adjusted net volumes

	Method		
	TOBIT	Perm	Order-M Perm
Constant	0.985	0.841	0.863
	$p = 0.000^{***}$	$p = 0.000^{***}$	$p = 0.000^{***}$
Age	-0.004	-0.0001	0.001
	p = 0.458	p = 0.970	p = 0.802
Density	-2.012	-0.314	-0.372
	p = 0.254	p = 0.246	p = 0.170
Age*Density	0.051	0.052	0.065
	p = 0.287	p = 0.275	p = 0.178
logSigma	-2.347		
	$p = 0.000^{***}$		
Observations	70	70	70
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 8 shows that neither age nor density is significantly correlated with the efficiency scores[8]. There can be several reasons for this. We argue, for example, in section 3.2 that the previous net volume measures were misspecified. Furthermore, there are now only eight companies, which are not both age- and density-neutral. This is likely too few companies to make a significant difference in the results. In fact, we can examine this by estimating (8) as in the previous section, but this time for the new DEA model. If we do this, we find that there is still no significant correlation between the efficiency scores and age (density)[9]. In addition, we find that there is no significant difference between the age (density)-neutral companies and the non-age (density)-neutral companies. This is unexpected but is likely due to the fact that eight non-neutral companies are too few to make a significant difference in the results.

We can, therefore, conclude that the model with the new proposed adjusted net volumes are superior to KFST's adjusted net volumes, when the goal is to properly control for age and density. In addition, we find that there are no significant differences between the age (density)-neutral companies and the non-age (density)-neutral companies. This was unexpected but is probably because there are only eight companies that are non-age (density)-neutral. It is typically problematic to find any significant coefficients in a regression analysis when the corresponding variable only has few observations with non-zero values.

3.6 Number of positive multipliers

KFST's model identifies a large amount of slack. In fact, no companies have positive multiplier weights on more than two out of the six outputs. All companies have a positive weight on exactly one of the OPEX net volumes and one of the CAPEX net volumes. It follows almost directly from (7) that the companies will choose either the non-adjusted net volume or the age-adjusted net volume (and similarly for the density

[8] Even if we omit some of the covariates, the remaining coefficients are still insignificant

[9] The results are omitted in the paper due to lack of new information.

adjusted net volume versus the non-adjusted net volume). That there are fewer peers than might be expected, because of the high correlation (cf. section 3.3), also makes it easier for an inefficient company to have a higher age (density) than all their respective peers or a lower age (density) than all the peers. In this case, the company only have positive multiplier weight on either the non-adjusted OPEX (CAPEX) net volume or on one of the adjusted OPEX (CAPEX) net volumes cf. section 3.3.1. in addition, the few number of peers and high correlation between the outputs increases the likelihood that one peer have both the highest ratio of input to age-adjusted OPEX (CAPEX) net volume and input to density-adjusted OPEX (CAPEX) net volume. This is what we observe in KFST's current model. Here the companies are projected onto a non-fully-dimensional efficient facet, meaning that they will only have positive multiplier weights on either the age-adjusted OPEX (CAPEX) net volume or the density-adjusted OPEX (CAPEX) net volume. That observations are projected onto non-fully-dimensional efficient facets usually indicates some kind of misspecification (Olesen & Petersen, 1996). Normally the problem is that we want the companies to be measured in all available outputs because every output has different information. If the companies assign zero weights to some of these, we remove important information from the model. In KFST's benchmarking model, however, the outputs measure the same thing, the net volume, but do it with different assumptions. The presence of slack is therefore not necessarily problem for the regulator. In fact, when we have two or more outputs, which should have approximately the same information, it can be argued that there intuitively should be some slack in the model: Either we want the companies to be measured in the age (density)-neutral world or we want to compare them in the model, where we ignore age (density). A model in between will be hard to interpret.

4 Comparing the efficiency scores in different models

Having the non-adjusted net volumes included in the models (together with the adjusted ones) by design make some company non-age (density)-neutral (cf. section 3.5). In addition, the special structure creates several problematic characteristics (cf. section 3.3, 3.4 and 3.6). This is done as a precaution for the companies and can probably be removed (cf. section 3.1). In this section, we show how the efficiency scores depend on the non-adjusted net volumes in KFST's benchmarking model and in our model with the new proposed adjusted net volumes. We will later, in section 4.1.1 and 4.1.2 show that the new proposed adjusted net volumes are still superior to KFST's adjusted net volumes when the non-adjusted net volumes are omitted.

Figure 3 shows a Coxbox plot for the efficiency scores in the four models. The non-adjusted net volumes do almost not influence the results in KFST's model when we look at the average efficiency scores. In the new model, however, the efficiency scores are reduced substantially, when we remove the non-adjusted net volumes. It is not surprising that the non-adjusted net volumes influence the results more in the new model than KFST's model. That is because removing the non-adjusted net volume measures means going from 6 to 4 outputs in the original model, and from 4 to 2 outputs in the new model, and the impact of any additional variable is likely to decrease with the number of other variables considered.

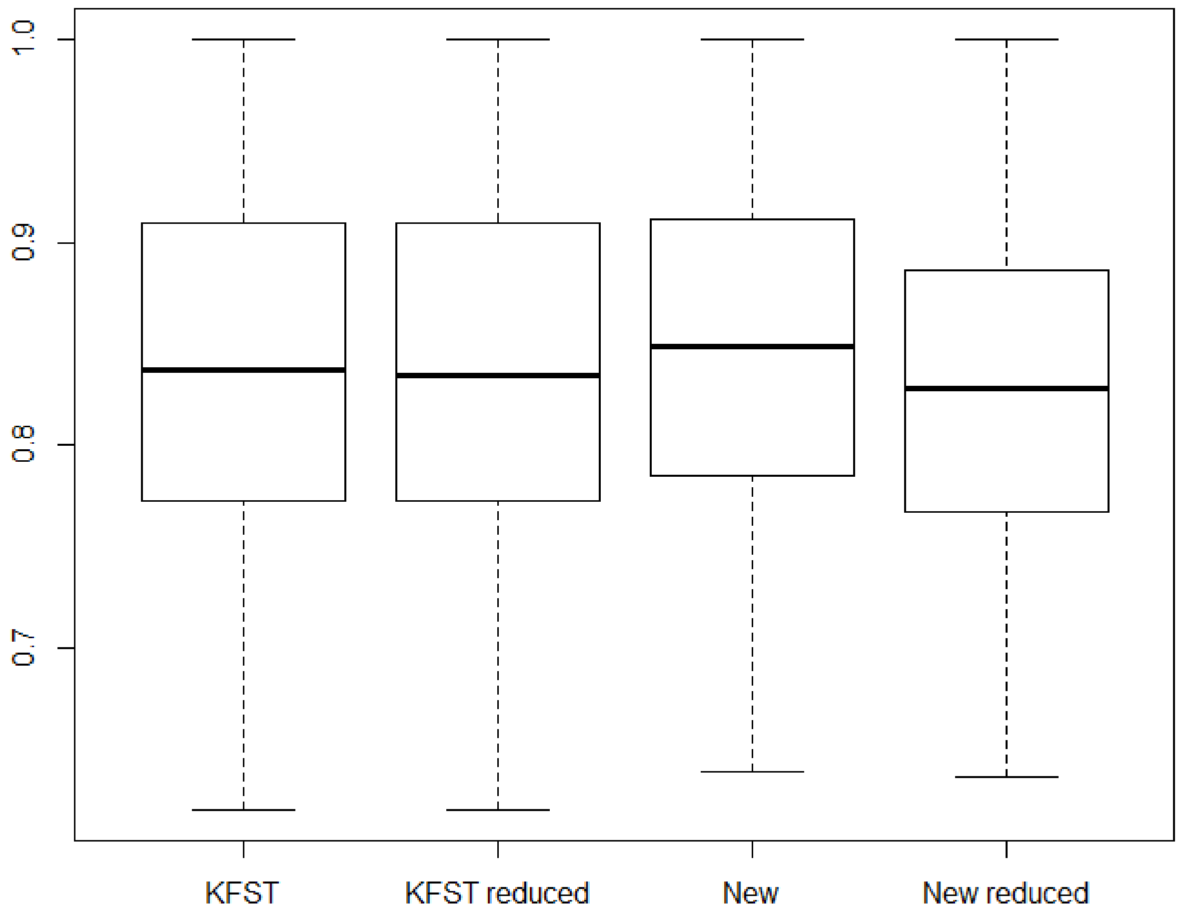

Figure 3 *Boxplot showing the efficiency scores for four models: KFST's current model, KFST's current model without non-adjusted net volumes, model with new proposed adjusted net volumes and model with new proposed adjusted net volumes without the non-adjusted net volumes*

KFST argue that inclusion of the non-adjusted net volumes is a precaution for a few companies rather than for the average company. It is therefore interesting to look at the changes for the individual companies. The left graph in Figure 4 shows the efficiency scores in the KFST model with and without the non-adjusted net volumes. Most companies have the same efficiency scores in both models but a few get a worse efficiency score in the model without the non-adjusted net volume. In the new model, however, we observe several companies which get a lower efficiency score when we remove the non-adjusted net volume, cf. the right graph in Figure 4.

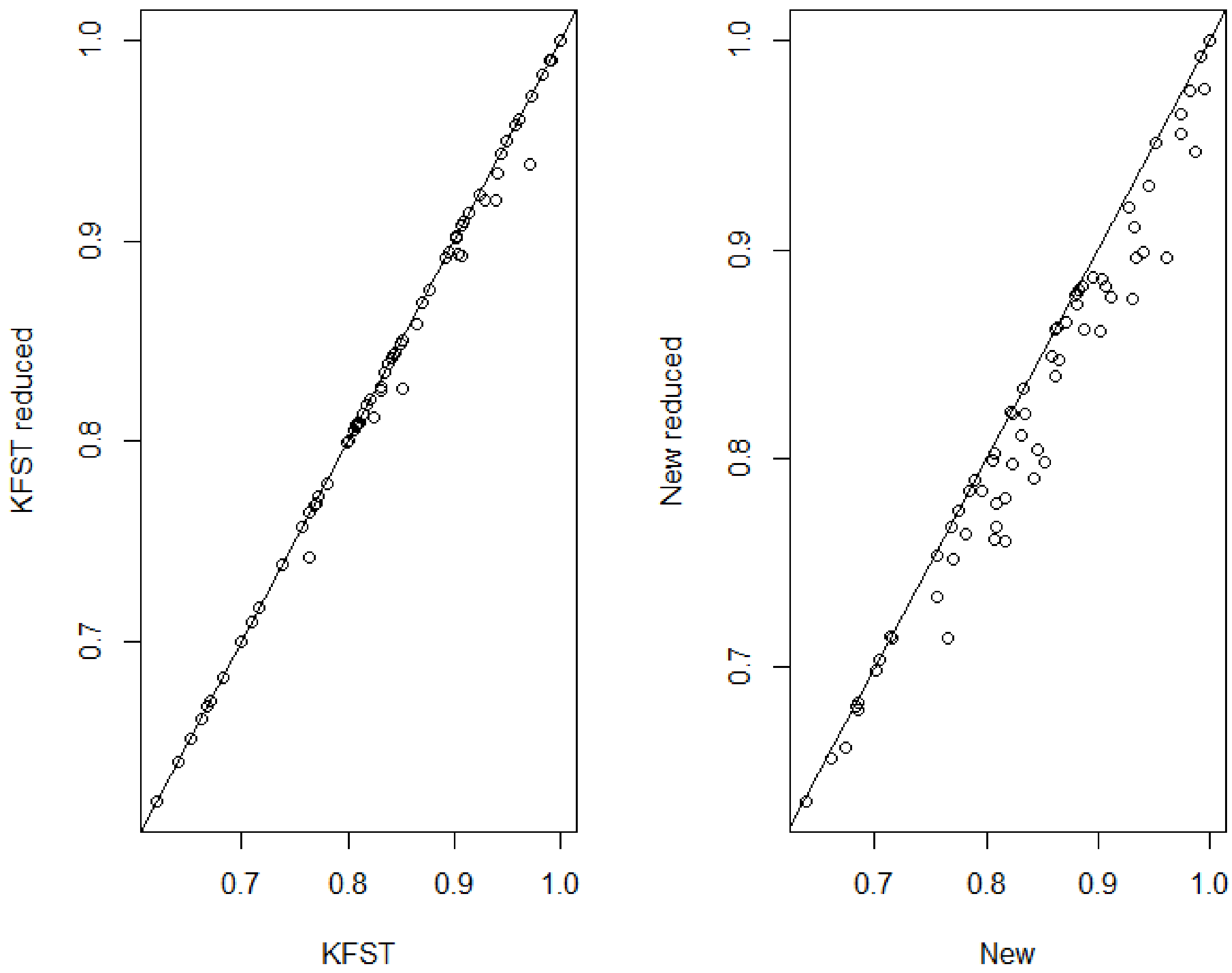

Figure 4 *Differences in the efficiency scores in the models with and without the non-adjusted net volumes, where "KFST" refers to and the current model, "New" to a model with the new proposed adjusted net volumes and "reduced" to a model without the non-adjusted net volumes*

The five largest changes in efficiency scores are shown in Table 9. Here we see that the changes for some companies are quite large. It is, therefore, reasonable to believe that the sector will use all their political power to keep the non-adjusted net volumes in the benchmarking model - especially if KFST chooses to change the adjusted net volumes to our newly proposed ones.

__Table 9__ Reduction in percentage points for the efficiency scores for the companies with the five highest reductions when omitting the non-adjusted net volumes

	Highest reduction	2. highest reduction	3. highest reduction	4. highest reduction	5. highest reduction	Mean of all companies
Model with KFST's adjusted net volumes	3.29	2.52	2.01	1.89	1.40	0.25
Model with new proposed adjusted net volumes	6.55	5.57	5.35	5.31	5.17	1.61

We have now shown that the non-adjusted net volumes are important outputs for some companies. It can, therefore, due to political influence, be difficult to remove them even while they theoretically perhaps should be removed (the non-adjusted net volumes are biased for age and density and are solely political precautions in the model, cf. section 3.1). In the next section, we show that a model with KFST's adjusted net volume but without the non-adjusted net volume still do not account properly for age and density. We show hereafter that the model with the new proposed adjusted net volumes but without the non-adjusted net volumes does account for age and density as intended.

4.1.1 Second-stage analysis of KFST's current model without non-adjusted net volumes

Table 10 shows the results from KFST's current model where we have omitted the non-adjusted net volumes. The table shows that there is a significant correlation between the efficiency scores and age as well as density even when the non-adjusted net volumes are omitted.

__Table 10__ Regression summary for a second stage analysis on KFST's current model without the non-adjusted net volumes

	Method		
	TOBIT	Perm	Order-M Perm
Constant	1.028	0.828	0.850
	$p = 0.000^{***}$	$p = 0.000^{***}$	$p = 0.000^{***}$
Age	-0.005	0.004	0.003
	$p = 0.325$	$p = 0.239$	$p = 0.250$
Density	-4.493	-0.543	-0.603
	$p = 0.013^{*}$	$p = 0.044^{*}$	$p = 0.026^{*}$
Age*Density	0.119	0.101	0.105
	$p = 0.016^{*}$	$p = 0.034^{*}$	$p = 0.028^{*}$
LogSigma	-2.361		
	$p = 0.000^{***}$		
Observations	70	70	70
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

We can, therefore, still conclude that KFST's adjusted net volumes do not take age and density sufficient into account. In the next section, we test if the new proposed adjusted net volumes account sufficiently for age and density, also when the unadjusted measures are excluded.

4.1.2 Second-stage analysis with new adjusted net volumes but without non-adjusted net volumes

Table 11 shows the results from a model with the new proposed adjusted net volumes but without the non-adjusted net volumes. The table shows that there is not any significant correlation between the efficiency scores and age nor density[10]. We have thus shown that our new proposed net volumes are superior to KFST's adjusted net volumes and that a model with these adjusted net volumes but without the non-adjusted net volumes also accounts properly for age and density.

Table 11 *Regression summary for a second stage analysis on the model with the new proposed adjusted net volumes without the non-adjusted net volumes*

	Method		
	TOBIT	Perm	Order-M Perm
Constant	0.977	0.826	0.842
	$p = 0.000^{***}$	$p = 0.000^{***}$	$p = 0.000^{***}$
Age	-0.004	-0.001	-0.001
	p = 0.370	p = 0.617	p = 0.758
Density	-1.439	-0.079	-0.064
	p = 0.409	p = 0.766	p = 0.811
Age*Density	0.041	0.042	0.049
	p = 0.387	p = 0.370	p = 0.298
LogSigma	-2.360		
	$p = 0.000^{***}$		
Observations	70	70	70
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

5 Conclusion

The Danish water regulator's benchmarking model use the so-called net volumes as output, in which they incorporate different environmental variables. This gives the benchmarking model some unusual characteristics. In this paper, we showed that the Danish water regulator's benchmarking model could be improved without changing the overall characteristics of the model. We did this by proposing new adjusted net volumes, which should replace the current adjusted net volumes.

[10] Even if we omit some of the covariates, the remaining coefficients are still insignificant

We developed a new second stage analysis approach to assess whether the benchmarking model, in fact, did control for environmental variables with the current adjusted net volumes and the new adjusted net volumes respectively. Using this new approach, we realized that it is important to incorporate which multipliers have positive weights, when using a second stage analysis of DEA efficiency scores. The second stage analysis showed that the new adjusted net volumes were superior to the current adjusted net volumes.

In addition, we discussed the necessity of giving benefit of the doubt to the companies and that the regulator should be careful when using both the non-adjusted net volumes and the adjusted net volumes as outputs as to a precaution. We saw that these outputs were highly correlated, which means that regulator should be careful with the interpretation of the mean of the efficiency scores and the number of peers, the ranking of companies and the interpretation of zero weight on the multipliers. In addition, the assumption of strong disposability can be problematic when having correlated outputs. This is, however, not necessarily the case for the regulator because the correlated outputs are incorporated in the model as a precaution.

Lastly, we showed that the new adjusted net volumes could be used without the non-adjusted net volumes in the benchmarking model. If the latter are removed, it will give less benefit of the doubt to the companies and remove the high correlation, but at the same time still control for the environmental variables. However, it will reduce the efficiency scores and for some companies it will reduce the efficiency scores with a substantial amount. Therefore, it might be hard for the regulator to remove the non-adjusted net volumes due to political influence, even while it theoretically will be a better-behaved model.

6 Appendix

Table 12 *– Regression summary of the second stage analysis for a benchmarking model with total costs as input and OPEX- and CAPEX net volumes as output. The regression methods are explained in section 3.5*

	Method		
	TOBIT	Perm	Order-M Perm
Constant	0.944	0.788	0.814
	$p = 0.000^{***}$	$p = 0.000^{***}$	$p = 0.000^{***}$
Age	-0.002	0.004	0.005
	$p = 0.727$	$p = 0.136$	$p = 0.104$
Density	-4.398	-1.460	-1.468
	$p = 0.009^{**}$	$p = 0.00000^{***}$	$p = 0.00000^{***}$
Age*Density	0.087	0.086	0.093
	$p = 0.054$	$p = 0.064$	$p = 0.047^{*}$
logSigma	-2.404		
	$p = 0.000^{***}$		
Observations	70	70	70
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

7 References

Agrell, P. J., & Bogetoft, P. (2017). Regulatory Benchmarking: Models, Analyses and Applications. *Data Envelopment Analysis Journal*(3).

Aigner, D. J., Lovell, C. A., & Schmidt, P. (1977). Formulation and estimation of stohastic frontier production functions. *Journal of Econometrics*, s. 21-37.

Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, s. 75-88.

Asmild, M., Hougaard, J. L., & Kronborg, D. (2013). Do efficiency scores depend on input mix? A statistical test and empirical illustration. *Annals of Operations Research*, s. 37-48.

Banker, R., Førsund, F. R., & Zhang, D. (2017). Use of Data Envelopment Analysis for Incentive Regulation of Electric Distribution Firms. *Data Envelopment Analysis Journal*(3).

Benchmarkingekspertgruppen - Energitilsynet. (2017). *Benchmarkingrapport.* Energitilsynet, https://forsyningstilsynet.dk/aktuelt/publikationer/elmarkedet/rapport-med-anbefalinger-til-en-ny-benchmarkingmodel.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R.* Springer.

Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, s. 1-25.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, s. 429-444.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, s. 245-259.

Haney, A. B., & Pollitt, M. G. (2009). Efficiency Analysis of Energy Networks: An International Survey of Regulators. *Energy Policy*, s. 5814-5830.

Meeusen, W., & Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, s. 435-444.

Mehdiloo, M., & Podinovski, V. V. (2019). Selective strong and weak disposability in efficiency analysis. *European Journal of operational Research*, s. 1154-1169.

Oja, H. (1987). On Permutation Tests in Multiple Regression and Analysis of Covariance Problems. *Australian & New Zealand Journal of Statistics*, s. 91-100.

Olesen, O. B., & Petersen, N. C. (1996). Indicators of Ill-Conditioned Data Sets and Model Misspecification in Data Envelopment Analysis: An Extended Facet Approach. *Management science*, s. 157-306.

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, s. 461-469.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*.

Thanassoulis, E. (2000). DEA and its use in the regulation of water companies. *European Journal of Operational Research*(127).

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, s. 24-36.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, s. 381-397.

Wooldridge, J. M. (2006). *Introductory Econometrics: A modern Approach* (4 udg.). Cengage Learning.