



Incorporating quality in economic regulatory benchmarking

Heesche, Emil; Asmild, Mette

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Heesche, E., & Asmild, M. (2020). *Incorporating quality in economic regulatory benchmarking*. Department of Food and Resource Economics, University of Copenhagen. IFRO Working Paper, No. 2020/13

IFRO Working Paper



Incorporating quality in
economic regulatory benchmarking

Emil Heesche
Mette Asmild

IFRO Working Paper 2020 / 13

Incorporating quality in economic regulatory benchmarking

Authors: Emil Heesche, Mette Asmild

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

Published November 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/

Incorporating quality in economic regulatory benchmarking

Emil Heesche¹, Mette Asmild²

Abstract

The Danish water regulator uses, among other things, Data Envelopment Analysis to create a pseudo-competitive environment for the water companies. The benchmarking results are used to set an individual revenue cap for each company. The benchmarking model is currently criticized for not including the companies' supply quality and thereby has an omitted variable bias problem. The regulator has, therefore, initiated an extensive effort to try to incorporate supply quality in the regulation. One problem the regulator has encountered is that incorporating supply quality in the benchmarking model tends to increase the revenue caps more than desired. The regulator does, however, not have any prior information about the quality variables and their trade-offs to the remaining variables which make it challenging to reduce the supply quality's impact on the revenue caps.

In this paper, we analyze the facet structure when incorporating three quality variables into the existing model. The facet structure gives important insights into the trade-offs between the companies costs and their level of quality. We argue that it is generally sensible to investigate the facet structure and ensure that it is trustworthy before calculating efficiency scores, in order to increase the credibility of the results.

By using an outlier detection model on the estimated trade-offs we use the insights for the facet structure to create weight restrictions between costs and quality, which gives the companies incentives to reveal private information about their true trade-offs. This can help the regulator incorporate quality in the model without allowing the efficiency scores to increase excessively due to the increase in dimensionality. In addition, we propose to set weight restrictions based on the consumer's willingness to pay for quality to avoid the companies choosing a level of quality that is higher than what the consumers are willing to pay.

Keywords Data Envelopment Analysis; Regulation; Facet structure; Weight restrictions; Trade-off

¹ University of Copenhagen, Department of Food and Resource Economics, Rolighedsvej 25, 1958 Frederiksberg C Denmark. ehe@kfst.dk (corresponding author)

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

1 Introduction

The Danish drinking- and waste-water companies are natural monopolies. The consumers cannot choose which provider to use and are therefore forced to use the local water company. Without regulation, the companies have incentive to set high monopoly prices and deliver poor quality. The Danish water regulator (KFST), therefore, regulates the sector by creating a pseudo-competitive environment with the use of benchmarking. Using benchmarking in regulation is not unique for the Danish drinking- and waste-water sector and is for example discussed in Agrell, Bogetoft, & Tind (2005), Dai & Kuosmanen (2014) and Ramanathan, Ramanathan, & Bentley (2018). The main idea in the Danish regulation is that a company is not allowed to set a higher price than other comparable companies do. KFST sets a revenue cap for each company to control the prices, by benchmarking the companies' costs against a measure for how much water they deliver, adjusted for several underlying conditions (Heesche & Asmild, 2020). The revenue cap is set equal to the most efficient companies' costs adjusted for the underlying conditions.

The benchmarking model used by KFST tries to control for as many differences in underlying conditions as possible. It does, however, not control for the quality of the water provision. For drinking water, quality is defined as safe and stable provision of good quality drinking water. For waste water, it is safe and stable management of wastewater, where the discharges of water have no negative impact on the environment. In this paper, we focus on the drinking water sector (hereafter water sector). Not considering quality in the existing model has led to two criticisms:

1) Companies with high quality, and thereby high costs to ensure this, can be compared to companies with low quality and thereby low costs. KFST only has information about the companies' total costs and can therefore not control for costs associated with ensuring high quality. The comparison, therefore, seems unfair. As a result, companies with low quality are allowed to have too high revenue and companies with high quality are forced to either lower their quality or be even more efficient than the low-quality companies. In other words, the model has an omitted variable bias.

2) Quality and economic efficiency are today regulated using two different regulations. This means, that neither regulation takes into account the correlation between the two.

The Danish politicians have, therefore, asked KFST whether it is appropriate to incorporate quality in their economic benchmarking model. However, this should only happen on the condition that KFST does not reduce the yearly cost-reduction requirement among the companies substantially.

KFST's uses both data envelopment analysis (DEA) and stochastic frontier analysis (SFA) in their benchmarking model. In this paper, we solely focus on DEA. It is well-known that DEA often estimates extreme multiplier weights to give the companies the best possible efficiency scores; DEA gives the benefit of the doubt to the companies. To make sure that quality does not influence the cost-reduction requirements too much, KFST could, therefore, try to reduce the number and size of extreme multipliers by only allowing realistic trade-offs between costs and quality. We define a trade-off as the ratio between two multipliers and it can be interpreted as how much you can improve one variable if you worsen the other. The problem for KFST is that they do not know which trade-offs are realistic.

In this paper, we aim to estimate a technology where the frontier has as few unrealistic trade-offs as possible without relying on prior information about which tradeoffs are inappropriate. Therefore, we first examine the

facet structure of the DEA-estimated frontier to study the trade-offs between quality and costs in the DEA framework. Hereafter, to improve the estimation of the trade-offs and to fulfil the political requirement that the inclusion of quality does not reduce the cost savings too much, we develop a set of weight restrictions. We base the weight restrictions on a facet outlier analysis, where we initially assume that all outlying facets have unrealistic trade-offs between costs and a quality variable. This will create a harsh model. Hereafter, for this method to work properly, we want the companies to reveal their private information about the true trade-offs to help us identify which outliers found in the outlier analysis do, in fact, have realistic trade-offs. If we allow the initial harsh model to be reestimated (and softened) by incorporating the additional information about true trade-offs provided by the companies, they will, according to agency theory, have incentive to disclose their private information without reducing the effectiveness of the regulation (too much).

It is important to note that we mainly focus on the estimation of the frontier. This means that the choice of the direction of the projection of the inefficient companies onto the frontier does not influence the results. Only after a technology with realistic trade-offs has been estimated does the projection become relevant.

The rest of this paper is structured as follows: Section 2 describes KFST's current benchmarking model, the new quality data and some changes made to the model. In section 3 we examine the facet structure and discuss whether trade-offs are realistic or not. In section 4 we incorporate weight restrictions based on the consumers' willingness to pay. In section 5 we discuss how to get the companies to reveal their true trade-offs by removing potentially unrealistic trade-offs. In section 6 we examine our method's impact on the efficiency scores and discuss the political agenda. Section 7 concludes the paper.

2 The current benchmarking model in the Danish water sector

The current benchmarking model in the Danish (drinking) water sector has been described in Heesche & Asmild (2020). In this paper, we only provide a short recap of the most important characteristics of the model incorporating the changes suggested by Heesche & Asmild (2020).

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) in a so-called "best-of-two" benchmarking model. This means that KFST uses the highest efficiency score from the two models for each company to set the revenue cap. However, in this paper, we only focus on the DEA model.

The DEA model consists of one input and two outputs. The input is the companies' total controllable costs (hereafter costs) and the outputs are measures for the capacity that the companies make available for their consumers. These are aggregations of the total length of water pipes, the quantity of water delivered and the number of consumers to be serviced among many other things. The first output is an aggregation of the operational costs associated with providing the capacity (OPEX) and the second is the capital costs of the capacity (CAPEX). Both outputs are adjusted for the companies' assets' average age and the population density³ in their supply area. We use the OPEX and CAPEX definitions from Heesche & Asmild (2020) where both measures are adjusted for both age and density as well as their interaction.

The model is input orientated because the regulatory goal is to reduce the companies' costs. It can also be argued that the companies have exactly the output they need, because they have to deliver the water

³ The population density is calculated as the number of consumers divided by the total length of water pipes.

requested by the consumers and cannot compete to increase or decrease market shares. KFST assumes constant return to scale (crs). There are two arguments for this. First, the aggregation of the outputs already take into account different scales through the prices used for the capacity components. Second, if the companies are not on the optimal scale size the regulation should provide incentives for them to become so by merging/splitting⁴.

The data consists of the 74 biggest water companies in Denmark. Following KFST four of these are characterized as outliers. For simplicity, we only use the remaining 70 companies in this paper. The outliers do, however, not influence the overall conclusions.

We consider three new variables as indicators of quality: Bacteriological Excesses (BE), Pipe Breaks (PB) and Water Wastage (WW). "BE" measures how often the companies exceeds the bacteriological limits set by the Danish Environmental Protection Agency. We assume that bigger companies have more BE than smaller companies as they have a larger supply area and therefore have more places where things can go wrong. "PB" measures how often the consumers are left without water because of some breaks in the system. The breaks are measured as downtime for the breaks multiplied with the number of affected consumers. "WW" measures how much water the companies waste before it reaches the consumers. "WW" does not directly affect the consumers in the short run, but can result in a local shortage of water resources in the long run. All three variables are volumes and therefore increase with scale, making it possible to include these variables in a model assuming constant returns to scale.

Intuitively quality should be incorporated as output. Due to the definition of these variables, they will, however, be undesirable outputs. Many attempts have been made to incorporate undesirable outputs in DEA models, but no one without any disadvantages. For an overview of different methods see, for example, Scheel (2001). In this paper, we incorporate the quality variables as inputs. The quality variables should, therefore, be thought of as costs necessary to deliver water. In other words, the companies may need to have some Bacteriological Excesses, Pipe Breaks and Water Wastage to deliver their water at a reasonable price. Note that a high value of a quality variable indicates bad performance, consistent with the inclusion as an input. By high quality, we mean good performance, i.e. low values on these input variables. The advantages of incorporating quality as inputs are, firstly, that we can easily examine the trade-offs between costs and quality, where we expect that high quality, i.e. low values on those inputs are associated with higher costs and vice versa. If one value is high we would expect the other to be low. Secondly, most methods to incorporate bad outputs require a transformation of the quality variables, for example multiplying them with -1 or subtracting the values from a large number. This affects the variance in the variables and therefore their influence on the final results. Thirdly, working with negative outputs is problematic when using crs and weight restrictions, which we will do later. Lastly, a common method for incorporating undesirable outputs is to assume weak disposability and use a directional distance approach. This will make the trade-offs between costs and quality hard to interpret and we have, in the current case, no obvious reason for challenging the assumption of free disposability for quality, since it is possible to worsen quality without necessarily impacting the other variables.

⁴ Due to high transporting costs it is not possible for all companies to merge (split) all the companies' activities. It is, however, possible to, for example, merge (split) the administration among other things. This is an ongoing debate which we will not discuss further in this paper.

2.1 Data description

The DEA model in this paper incorporates the six variables mentioned above. The data is from 2017 and is publicly available at KFST's website⁵. Note that the data from 2018 and 2019 have only recently become available. We will therefore use data from 2019 later as a robustness check. The variable definitions between the years are, however, not identical and the data from 2018 and 2019 have yet to be quality assured. The variables for 2017 together with a short description are listed in Table 2.1.

Table 2.1 – Description of input and output variables in the DEA model

Variable	Description	Type	Measurement unit ⁶
Costs	The companies' total controllable costs	Input	100.000 DKK
Bacteriological Excesses (BE)	Number of exceedances per m3 of water corrected for the number of samples	Bad output treated as an input	1.000.000 #
Pipes Breaks (PB)	The sum of the consumers time without available water	Bad output treated as an input	100.000 min.
Water Wastage (WW)	Amount of water the companies waste before it reach the consumers	Bad output treated as an input	100.000 m ³
OPEX net volume	An aggregation of the capacity that the companies need to supply the consumers based on operational standard prices	Output	100.000 units
CAPEX net volume	An aggregation of the capacity that the companies need to supply the consumers based on capital standard prices	Output	100.000 units

The summary statistics are shown in Table 2.2. We observe that the 25 % quantile for BE is 0 meaning that a lot of the companies do not have a problem with BE

Table 2.2 – Summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Costs	70	126.125	133.769	28.634	57.847	124.883	781.790
BE	70	1.141	1.440	0.000	0.000	1.669	5.625
PB	70	2.863	6.267	0.000	0.096	2.650	45.768
WW	70	1.987	1.685	0.068	0.845	2.586	9.481
OPEX	70	122.436	117.895	37.289	62.725	135.346	781.261
CAPEX	70	138.040	129.083	37.555	63.639	154.816	785.278

⁵ www.kfst.dk

⁶ We scale all the measurement units for two reasons: First, we want most solvers to be able to handle the program without changing their tolerance level for small numbers. Second, it is easier to interpret, analyse and report numbers close to one than numbers where the 10th decimals are important

3 The facet structure

The frontier in DEA consists of several piecewise linear relationships between the variables called facets. The efficiency score for a point inside the frontier measures its distance to one of these facets. If a company is located on a facet it may be fully efficient, but it may also exhibit non-radial slack on some variables. The relationships, or trade-offs, between variables, are often examined through the dual (multiplier) DEA formulation. Here the trade-offs can be seen from the multipliers (weights). However, multipliers are only identified for the facets that the inefficient companies are projected on to. There likely exists additional facets containing information about the possible trade-offs. Several studies have examined the DEA estimated trade-offs including Podinovski (2019) and Asmild, Paradi, & Reese (2006) To identify all the existing facets of the convex hull of the observations, we use the QHull algorithm (Barber, Dobkin, & Huhdanpaa, 1996), but from those only consider the subset of facets indicating the frontier in a CRS model and augmented with facets generated by the free disposability assumption in DEA. Petersen & Olesen (2015) offer an algorithm using QHull to find all possible facets in CRS and VRS.

With our 6 variables (1 cost input, 3 quality inputs, 2 outputs) we identify 115 facets generating the CRS frontier. Only 12 of these facets are fully dimensional facets meaning that they are defined from observations alone and not the free disposability assumption (Olesen & Petersen, 1996). These are also the only facets where the slope is different from zero in all dimensions. If the goal is to find some realistic trade-offs between the variables these will be the most interesting facets. However, it is not necessarily these facets that most companies are projected onto. The normals to the fully dimensional facets, indicating trade-offs between the variables, are shown in Table 3.1. For example, the trade-off between costs and BE on the first facet is

$$\frac{-0.0093}{-0.0418} = 0.22.$$

Table 3.1 - Normals to the fully dimensional facets. Note that the offsets to the normals are omitted because they are all zero in crs

	Costs	BE	PB	WW	OPEX	CAPEX
Facet 1	-0.0093	-0.0418	-0.0502	-0.9977	0.0132	0.0004
Facet 2	-0.0457	-0.0855	-0.1084	-0.9885	0.0423	0.0032
Facet 3	-0.0383	-0.196	-0.0579	-0.9773	0.0398	0.0002
Facet 4	-0.029	-0.0319	-0.0313	-0.9982	0.0291	0.0012
Facet 5	-0.0366	-0.0568	-0.0953	-0.9925	0.0367	0.0015
Facet 6	-0.0409	-0.0369	-0.1113	-0.9915	0.0388	0.0027
Facet 7	-0.0325	-0.0887	-0.1909	-0.9764	0.0347	0.0017
Facet 8	-0.0176	-0.1445	-0.3114	-0.9387	0.025	0.0008
Facet 9	-0.0346	-0.6832	-0.4432	-0.5788	0.0208	0.0125
Facet 10	-0.0167	-0.1666	-0.3291	-0.929	0.024	0.0011
Facet 11	-0.0205	-0.497	-0.4611	-0.7346	0.0182	0.0074
Facet 12	-0.0002	-0.0003	-0.9999	-0.0115	0.0003	0.0001

To ease the interpretation we in Table 3.2. show the trade-offs between costs and the quality variables. The results show that the trade-offs vary a lot and that the trade-offs on certain facets are quite extreme. The

trade-off between costs and PB is, for example, 4631 times higher on facet 4 than on facet 12. With such big differences, it seems inappropriate to use these results to estimate a single value for the trade-off, for example by taking the average. We can, however, use the results to say something about the appropriateness of each of the facets. If we have partial information about the true trade-offs we can use this to remove some of the facets. Even if we do not have any additional information we can argue that the most extreme facets should be excluded. We will use this kind of logic later on.

Table 3.2 - Trade-offs between costs and the quality variables

	Costs/BE	Costs/PB	Costs/WW
Facet 1	0.2225	0.1853	0.0093
Facet 2	0.5351	0.4221	0.0463
Facet 3	0.1954	0.6613	0.0392
Facet 4	0.9099	0.9261	0.029
Facet 5	0.6449	0.3842	0.0369
Facet 6	1.1078	0.3676	0.0413
Facet 7	0.3669	0.1705	0.0333
Facet 8	0.1219	0.0566	0.0188
Facet 9	0.0506	0.0781	0.0598
Facet 10	0.0999	0.0506	0.0179
Facet 11	0.0413	0.0445	0.0279
Facet 12	0.5545	0.0002	0.014

When we add the non-fully dimensional facets it is a well-known problem that the trade-offs can be extreme and often result in undefined ratios due to at least one dimension having a slope equal to zero. The distribution of the trade-offs between costs and the quality variables for all facets are shown in Figure 3.1⁷. The figure shows that most facets do not have a well-defined trade-off between costs and quality. The number of poorly defined trade-offs are shown in the upper left corner. Zero means that the slope on costs is zero, “Inf” means that the slope on quality is zero and “NA” means that both slopes are zero. These trade-offs come as a result of the free disposability assumption⁸. The remaining trade-offs are located in quite large intervals with a few extreme points. It is, for example, quite extreme that some trade-offs are more than 37,000 times larger than others, as is the case e.g. Costs/BE. In the absence of prior information, it becomes a subjective assessment whether such trade-offs are realistic or not.

⁷ Note that the y-axis is logarithmic because we want to compare the ratios between the trade-offs

⁸ A slope of zero can also occur if at least two of the efficient companies that span a certain facet have the same value on a quality variable. We observe, for example, several companies with a value of zero on a quality variable

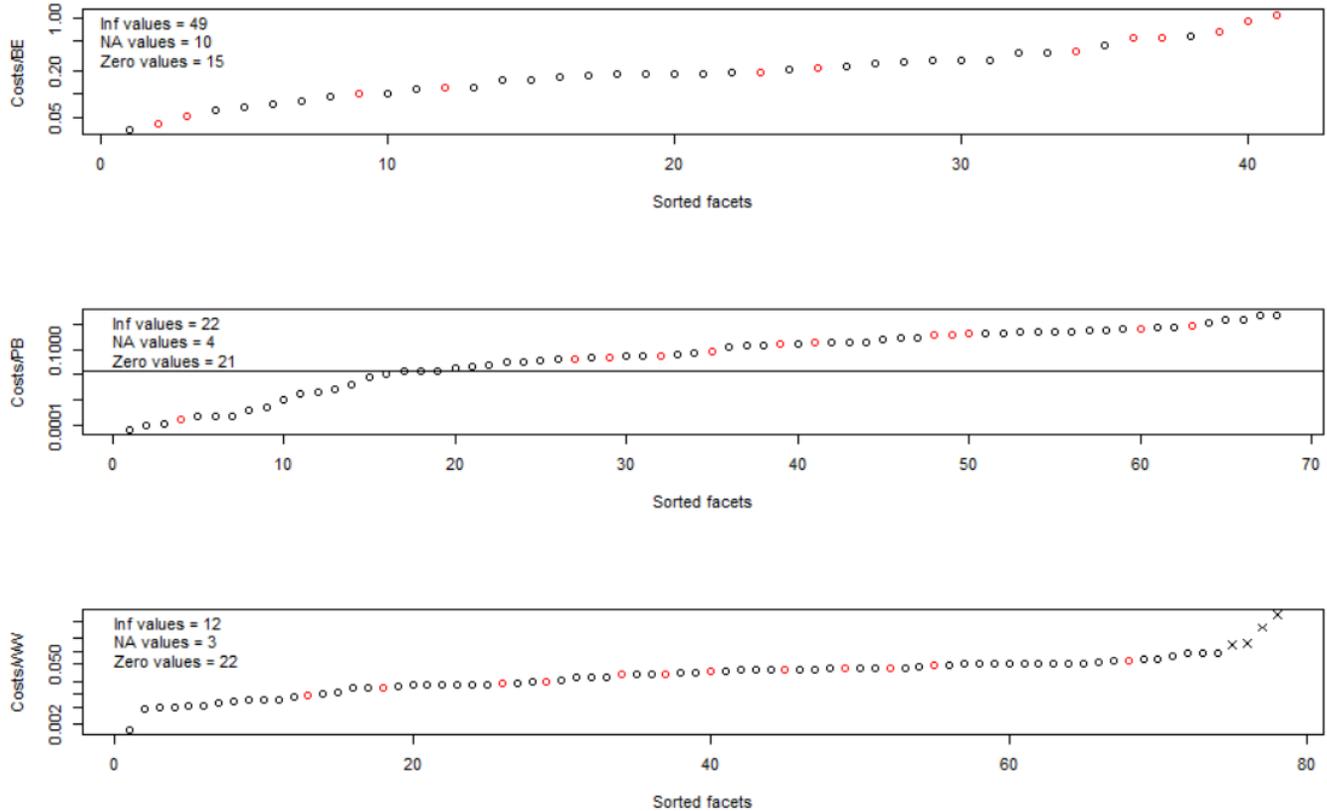


Figure 3.1 - Distributions of trade-offs between costs and quality. Note that the zero, Inf and NA values are omitted. The red dots indicate trade-offs from a fully dimensional facet. X indicates potential outliers and the y-axes are logarithmic. The horizontal line for $\frac{Costs}{PB}$ indicate the consumers willingness to pay cf. section 3.1.1.

3.1 Indications of unrealistic trade-offs

To remove unrealistic trade-offs we first need to identify them. If we do not have any prior information about the range of allowable trade-offs, we do not have any objective criteria for when a trade-off is unrealistic. We, therefore, define some subjective criteria that we want to be fulfilled, which practitioners (like a regulator) could use to evaluate if a trade-off is realistic or not when they lack prior information or expert opinions.

3.1.1 Willingness to pay

KFST do not have any prior information about the trade-offs between costs and quality. They know, however, the consumers' willingness to pay for (avoiding) PB (Konkurrence- og Forbrugerstyrelsen, 2020). This information cannot directly be used to identify unrealistic trade-offs but can, however, be used to identify trade-offs where the costs associated with improving quality means that these improvements are not in the consumers' interest. It can, therefore, be argued that the allowable trade-offs between costs and PB should not involve improvements in quality that are more expensive than what the consumers are willing to pay for them.

The willingness to pay for PB is 6.6 DKK. This means that the trade-off for cost/PB should be around 6.6 if we have the consumers' interest in mind. Figure 3.1 shows that most of the estimated tradeoffs are quite different

from this value. This is not surprising because there is not necessarily any relationship between willingness to pay and costs in a monopoly.

3.1.2 Expert opinions

KFST does not have any prior information about the true trade-offs between costs and quality. It is, however, likely that the companies themselves have at least some partial information about these trade-offs. This information is not something KFST can find in any existing sources and it is, therefore, necessary to elicit this directly from the companies. The companies do, however, not have any incentive to disclose their true trade-offs if KFST uses this information to remove unrealistic trade-offs which will lower the companies efficiency scores.

If KFST wants the companies to reveal their private information about true tradeoffs, they need a method that gives the companies an incentive to do so. This method should not only give appropriate incentives but should also be administratively easy to implement for both the companies and KFST. If it is not easy to implement, there is a risk that the administrative costs exceed the gain of getting the information. We propose such a method in section 5.

3.1.3 Regression/SFA

In lack of prior information about the true trade-offs, it would be obvious to estimate these with regression analysis or a stochastic frontier analysis (Meeusen & Broeck, 1977, Aigner, Lovell, & Schmidt, 1977). We could, for example, compare the DEA trade-offs with the maximum and minimum tradeoffs in these parametric models or look at the confidence intervals for the individual coefficients. We have, however, not succeeded in getting significant results in our parametric models and can, therefore, not use this method to identify when the trade-offs in DEA are unrealistic or not. There can be several reasons why we do not get significant results:

First, defining a parametric model can be a complicated task. In our case, we want to control for two environmental conditions, two net volumes and the three new quality variables (or at least one of the quality variables). With only 70 observations this relies heavily on appropriate choices of model specifications and parametric assumptions. We have, of course, experimented with several different models, but are for now leaving an extensive parametric analysis for further research⁹.

Second, intuitively we think that more quality requires higher costs. It might, however, be that the damage costs from repairing for example a pipe break are just as high as the costs of proper maintenance. This will result in trade-offs (coefficients in the parametric set-up) that are close to zero.

Third, some companies might be efficient due to superior management, which enable them to outperform others with respect to both economic efficiency and quality. Such companies will, therefore, have both lower costs and better quality than the inefficient companies, which complicates the estimation of trade-off between costs and quality in a parametric model.

Due to the above problems, we will not use this method to identify unrealistic trade-offs in the present paper.

⁹ KFST is currently working on exactly this challenge

3.1.4 Free disposability

The number of trade-offs that are zero, Inf or NA in Figure 3.1 is an indication of how well (poorly) the facet structure is estimated. If we observe a lot of these non-fully dimensional facets, many trade-offs¹⁰ are defined based on the assumption of free disposability¹¹ instead of from the observed data. These trade-offs are likely unrealistic. We will later show how the proposed method in this paper reduces the number of non-fully dimensional facets.

3.1.5 Outliers

An, admittedly somewhat subjective, criterion is that it is desirable to have a distribution of trade-offs without legible outliers. An outlier in this context is a trade-off that is very different from the rest. An outlier does, of course, not necessarily imply an unrealistic trade-off, but it might be worthwhile investigating such trade-offs further. We use the adjusted interquartile range method (Hubert & Vandervieren, 2008) to identify potential outliers amongst the logarithmic trade-offs. This method has the advantages that it can handle skewed distributions and is commonly used. For a review of different outlier-detection models see for example Hodge & Austin (2004). We use the logarithmic trade-offs because we define an outlier as a trade-off that is much bigger or smaller than the remaining tradeoffs as a ratio rather than in discrete numbers. We, therefore, want a trade-off that is twice as high as another but only half the value of a third to lie exactly between these two trade-offs. If we did not use the logarithmic trade-offs the middle trade-off would be closer to the lower trade-off than to the higher trade-off. The outliers identified this way are indicated by the x's in Figure 3.1, where we observe 4 potential outliers among the trade-offs, specifically for the tradeoff between costs and WW.

4 Removal of the unrealistic facets using the consumers' willingness to pay

In the previous section, we discussed how to identify unrealistic trade-offs. We will now show how one can remove such trade-offs by incorporating a weight restriction defined from the consumers' willingness to pay for quality, specifically for avoiding pipe breaks (PB). KFST's priority is to reduce the companies' costs. We will, therefore, only restrict the trade-offs in the direction where the quality variable is given to much weight. In other words, quality should never be given more weight than what the consumers are willing to pay for improving it. The multiplier DEA program is given in (1)-(5), where (4) is the weight restriction. $X_i = [Costs_i, BE_i, PB_i, WW_i]$ and $Y_i = [OPEX_i, CAPEX_i]$ are the input and output vectors for company i . $v = [v_{Costs}, v_{BE}, v_{PB}, v_{WW}]$ and $u = [u_{OPEX}, u_{CAPEX}]$ are the input and output weights, which are estimated when the problem is optimized. The program can easily be modified to assume any of the standard returns to scale assumptions.

¹⁰ Note, that we only analyze the facets and not the inefficient companies projection to the facets. By "many" trade-offs we, therefore, refer to the number of unique facets and not the number of times a specific trade-off is being used as benchmark for an inefficient company.

¹¹ Or in some cases, due to at least two efficient companies spanning a facet and having the same value on a (quality) variable.

$$\max \quad uY_0 \quad (1)$$

$$\text{s. t.} \quad vX_0 \leq 1 \quad (2)$$

$$-vX_i + uY_i \leq 0 \quad \forall i \in I \quad (3)$$

$$6.6v_{costs} - v_{pe} \geq 0 \quad (4)$$

$$v, u \geq 0 \quad (5)$$

To incorporate the weight restriction in the QHull algorithm, we add one artificial data point (X'_i, Y'_i) per DMU, by modifying the costs according to the willingness to pay for reducing PB to zero cf. (6).

$$X'_i = [Costs_i + 6.6PB_i, BE_i, 0, WW_i], Y'_i = Y_i \quad (6)$$

These new data points correspond to each company having traded all their PB to costs at a ratio of 1:6.6. If needed, the procedure can be made more time-efficient if we only include artificial data points corresponding to the efficient companies. For small data sets like in the present case this is, however, not necessary. The QHull algorithm described earlier can now run on the new extended data set. Note that it might be necessary to create more extreme artificial data points with values below zero if we have several weight restrictions as is the case in section 5. The new data points should have sufficiently low coordinates for every possible convex combination between the weight restrictions and the observed data that lies in \mathbb{R}^+ to be included in the technology set. In addition, all facets based on the weight restrictions and which exclusively is located in \mathbb{R}^- in the relevant dimensions are excluded from the analysis.

When we add the weight restriction (4), we reduce the number of facets to 87 and increase the number of fully dimensional facets to 13. The trade-offs on the fully dimensional facets are shown in Table 4.1. We observe that 6 of the fully dimensional facets are restricted by the weight restriction. This is evident from the trade-off 0.1515 for Cost/PB.

Table 4.1 - Trade-offs on the fully dimensional facets with weight restrictions based on the consumers' willingness to pay

	Cost/BE	Cost/PB	Cost/WW
Facet 1	1.1078	0.3676	0.0413
Facet 2	0.5351	0.4221	0.0463
Facet 3	0.3669	0.1705	0.0333
Facet 4	0.6449	0.3842	0.0369
Facet 5	0.9099	0.9261	0.029
Facet 6	0.2225	0.1853	0.0093
Facet 7	0.1954	0.6613	0.0392
Facet 8	0.0915	0.1515	0.0871
Facet 9	0.0772	0.1515	0.0825
Facet 10	0.2876	0.1515	0.0305
Facet 11	0.3261	0.1515	0.0318
Facet 12	0.2003	0.1515	0.0099
Facet 13	0.1079	0.1515	0.0076

Figure 4.1 shows the distributions of tradeoffs based on the weight restricted model (i.e. with (4) included). The figure illustrates how the weight restriction only restricts the lower values of $\frac{Cost}{PB}$. The number of NA and zero values have decreased a lot for all trade-offs because the multiplier for costs only can be equal to zero if the multiplier for PB is also zero. A single weight restriction between costs and a quality variable therefore highly influence the trade-offs between costs and the remaining quality variables as well.

By adding this weight restriction we have therefore reduced the number of unrealistic trade-offs discussed in section 3.1.4 and at the same time, we give the companies incentive to not spend more money on PB improvements than what the consumers are willing to pay. We observe, however, still several tradeoffs that according to the discussion in section 3.1.4 and 3.1.5 potentially are unrealistic. We will, therefore, in the following incorporate the outlier-based approach to removing potentially unrealistic trade-offs, with the underlying premise that the companies then have incentives to disclose private information about the true tradeoffs which might subsequently reintroduce some of these facets (tradeoffs).

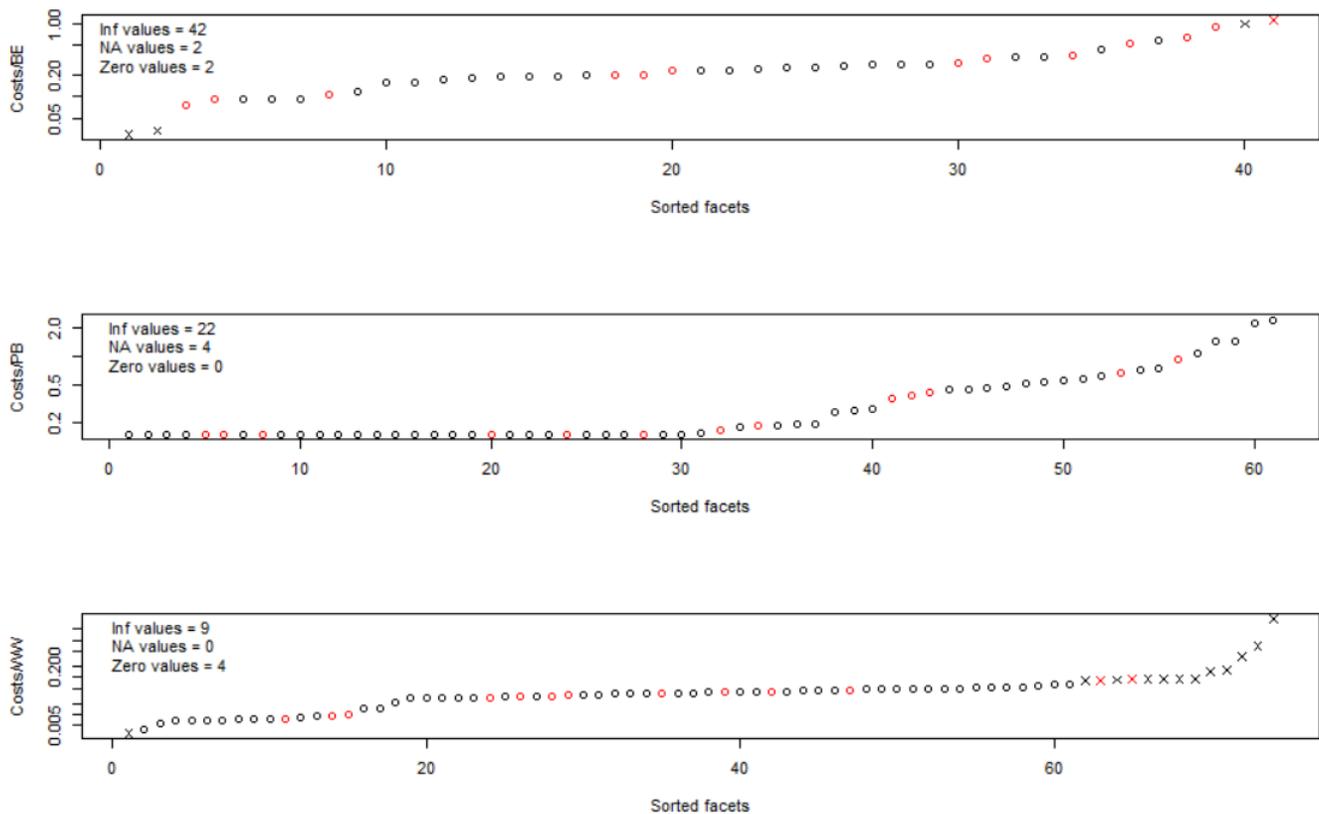


Figure 4.1 - Distributions of trade-offs between costs and quality. Note that the zero, Inf and NA values are omitted. The red dots indicate trade-offs from a fully dimensional facet. X indicates potential outliers and the y-axis is logarithmic

5 Removal of the unrealistic facets using expert opinions and outlier detection

We want to provide incentives for the companies to disclose their true trade-offs. Therefore, we propose to develop an initially harsh model but with a possibility for the companies to apply for the use of a more accommodating model, subject to providing additional information about their true tradeoff that can be included in the model. In this way, the companies have an incentive to reveal their private information about their true trade-offs, because it can potentially give them a higher efficiency score.

5.1 Simplified case

To illustrate the idea we first simplify the model such that WW is the only quality variable, and the OPEX net volume and CAPEX net volume are aggregated (added) into one single TOTEX net volume. By doing this we can graphically illustrate the differences in the facet structure before and after removing the unrealistic facets. Lastly, we for now assume variable return to scale to increase the number of facets in the 3D illustration to make it more interesting. Figure 5.1 shows the facet structure with all the facets. The red dots indicate the observations¹² and the blue dots show which of these are fully efficient. The green facets are the ones with the highest or undefined $\frac{Costs}{WW}$. According to the method explained in section 3.1.5 these are the ones that are outliers in the distribution. The orange facet has the fourth highest well-defined $\frac{Costs}{WW}$. There are 12 fully dimensional facets and 23 facets in total.

¹² In order to zoom in on the lower, interesting, part of the frontier, the figure is trimmed such that a few inefficient observations are not shown

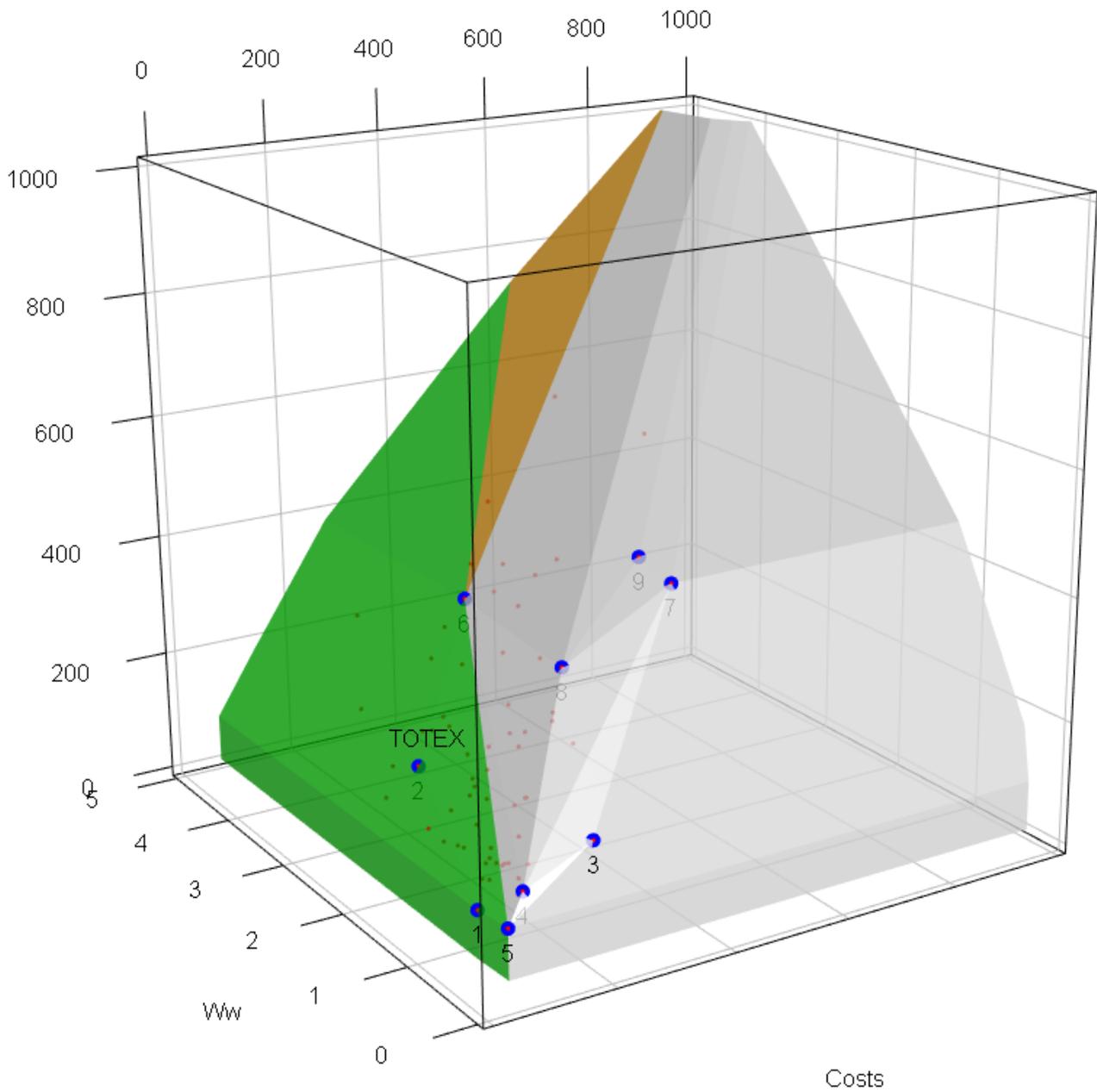


Figure 5.1 – Facet structure with two inputs (Costs and WW), one output (TOTEX) and variable return to scale

The green facets all have very high values for the tradeoff between Costs and WW, meaning that a company on these facets can increase WW without changing Costs, and still be close to the facet. And since these tradeoffs are classified as outliers according to section 3.1.5 we remove them by adding a weight restriction limiting the trade-offs between Costs and WW to be smaller than or equal to the trade-off given by the orange facet. This removes the green facets, fully-dimensional or not, through the introduction of 4 new facets generated by the weight restrictions as shown in Figure 5.2.

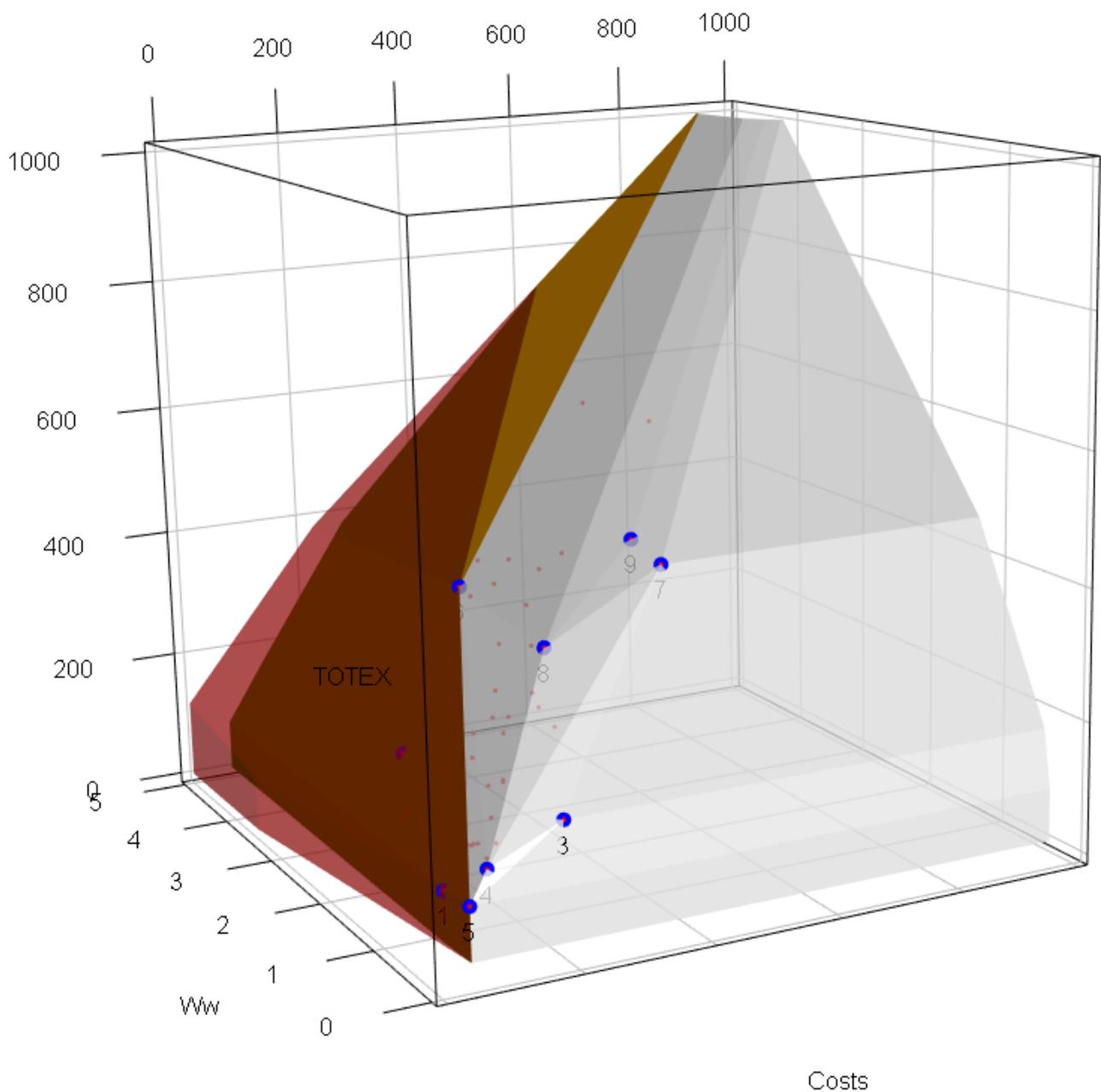


Figure 5.2 - Facet structure with two inputs, one output, variable return to scale and weight restrictions indicated by the red facets

Figure 5.2 shows how the weight restrictions extend the technology from the green area in Figure 5.1 to the red area. The number of facets based on the assumption of free disposability on WW (zero multiplier on WW) is now reduced to one and is pushed further away from origo (it starts at $WW = 2,5$). The part of the technology with zero multiplier on WW is therefore reduced and the part with a realistic trade-off between Costs and WW have increased substantially. The orange and grey parts of the frontier are not influenced by the weight restriction and therefore stay the same. Note that company 1 and 2 are no longer fully efficient even though they might still be visible behind the red facets.

The comparison between the two figures above illustrates how a weight restriction can reduce the zero-multiplier problem in a given dimension without compromising the well-behaved estimated grey part of the frontier.

From a regulator’s point of view, it is desirable to have a formalized mechanical approach to incorporating weight restrictions. We, therefore, propose to deem all tradeoffs classified as outliers as being unrealistic and remove them using the weight restrictions. This will still likely create a harsh model with low efficiency scores. The companies can hereafter apply for some of the removed trade-offs being allowed again. Their application will likely be administratively cumbersome, as they have to justify why the tradeoffs are appropriate, but ideally, it will only be relevant for a few companies and they can evaluate in advance if it will be worth it.

As discussed in section 3.1.5, there are many different methods for outlier detection. The methods that find most outliers will give the companies the strongest incentives to disclose private information but will also be the ones with the highest administrative costs. We here propose to use the adjusted interquartile range method as described in section 3.1.5. This method has the advantages that it can handle skewed distributions and is commonly used. We furthermore use an iterative process, where one outlier at a time is removed, and the model then reestimated. In each iteration, the outlier furthest from the whiskers in the adjusted Box plot is removed. Other approaches can be used as well for achieving the same overall purpose.

We remove the outliers by adding a weight restriction which constrains the tradeoffs to be no higher (lower) than the second highest (lowest) tradeoff in the relevant dimension. By doing this, the weight restrictions will expand the technology set more and more for each iteration until there are no longer any outliers left.

5.2 Full case

For a proper empirical illustration of the approach, consider again the full data set with all three quality variables and the two separate outputs, and assuming CRS. The iterations start with the model described in (1)-(5), with the corresponding tradeoffs illustrated in Figure 4.1. This model has, compared to the model in section 4, reduced the number of outliers for Costs/PB to zero but have, however, increased the number of outliers for some of the remaining trade-offs.

The iterative outlier deletion method uses 9 iterations before there are no outliers left. Note that 9 iterations does not mean that the method finds 9 outliers. We do not necessarily remove an outlier for every iteration because the weight restriction only restricts the trade-offs to the second-highest (lowest) trade-offs, which also may be identified as an outlier. At the same time, it is, however, possible for a single iteration to remove several outliers because it influences the entire facet which consists of several trade-offs that each can be an outlier. The weight restrictions developed from each iteration are given in Table 5.1. To implement the weight restrictions in a dual DEA program we add the constraints

$$P_j^T v \geq 0, \quad \forall j$$

to the program (1)-(5) where P_j is a vector corresponding to row j in Table 5.1. We observe, as expected, that the weight restrictions get more and more strict for every iteration. Note that the outliers can either be

extreme because they are too high or too low. If the weight restriction for a given iteration is -1 for Costs it means that the outlier is too high and the weight restriction, therefore, forces the trade-off $\frac{Costs}{Quality}$ to be lower and vice versa. Iteration 1 forces for example $\frac{Costs}{WW} \leq 0.7136$ and iteration 2 forces $\frac{Costs}{WW} \geq \frac{1}{247.9966}$.

Table 5.1 – Weight restrictions based on willingness to pay and hereafter an iterative outlier detection

	Costs	BE	PB	WW
Iteration 0	6.6	0	-1	0
Iteration 1	-1	0	0	0.7136
Iteration 2	247.9963	0	0	-1
Iteration 3	286.5374	-1	0	0
Iteration 4	153.6905	-1	0	0
Iteration 5	-1	0.9099	0	0
Iteration 6	-1	0	0	0.3973
Iteration 7	-1	0	0	0.2535
Iteration 8	-1	0	0	0.1523
Iteration 9	-1	0	0	0.1384

Adding these weight restrictions results in the trade-offs given in Figure 5.3. Due to weight restrictions most of the undefined trade-offs (non-fully dimensional facets) have been eliminated. The number of fully dimensional facets is now 29 and the total number of facets is 91. The non-fully dimensional facets are still present because we do not have any weight restrictions for the outputs, and we have facets along the axes which are based on free disposability rather than weight restrictions. We observe that approximately half of the trade-offs are restricted by weight restrictions while the remaining facets had well-behaved trade-offs to begin with. If the companies have information about true trade-offs and therefore can argue to remove some of the weight restrictions used to limit the allowable tradeoffs, the number of trade-offs eliminated through weight restrictions will, likely, be reduced. The number of trade-offs influenced by the weight restrictions here might seem a bit excessive. This is, however, intended as this method aims to provide incentives for the companies to reveal their true trade-offs. At the same time, however, we note that the ratio between the highest and lowest trade-off for $\frac{Costs}{BE}$, $\frac{Costs}{PB}$ and $\frac{Costs}{WW}$ are 140, 10 and 34 respectively. The method does, therefore, still allow a lot of flexibility in the DEA model when calculating trade-offs, which is arguably one of the strengths of DEA. The range of estimated tradeoffs has, however, decreased substantially for all three quality variables.

We note that the method described above results in approximately the same range and distribution of the trade-offs if we use data from 2019 instead of 2017, cf. Appendix A.

In section 6 we show that even though this final model might seem harsh, it can be argued that it fulfils the political desire to implement quality in the model without compromising the efficiency requirement for the companies' costs.

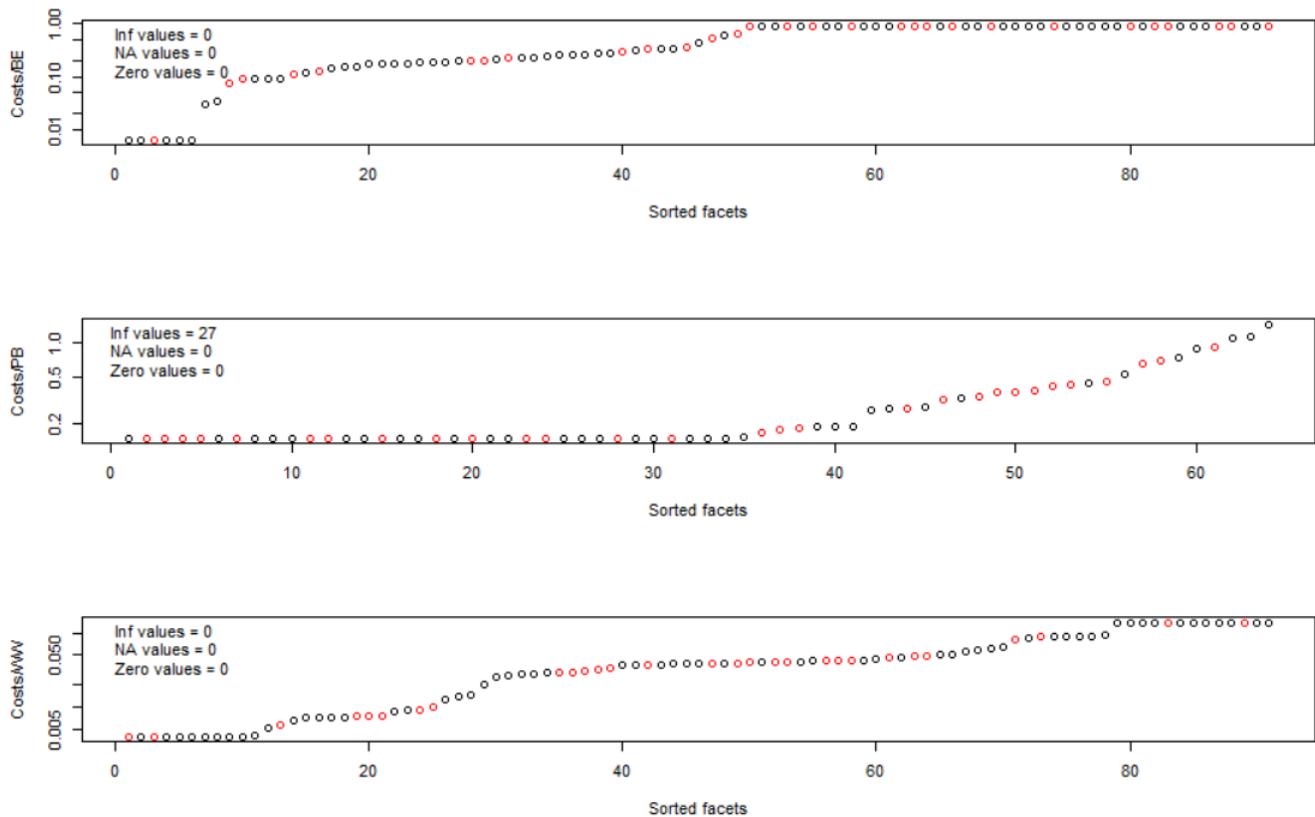


Figure 5.3 - Distributions of trade-offs between costs and quality. Note that the zero, Inf and NA values are omitted. The red dots indicate trade-offs from a fully dimensional facet and the y-axis is logarithmic

5.3 Example of argumentation for true trade-offs

Assume that a company wants to argue against an imposed weight restriction on WW. This scenario is fictional but the idea is based on interviews with a few companies. The company has recently maintained several km of pipes because their WW was too high in this area. The company can document the costs for this maintenance which, for simplicity, have not improved the pipes in any other way; this means that all the costs can be directly compared to the reduction in WW. The company has used 100,000 DKK on the maintenance and have subsequently reduced their WW with what corresponds to 20,000 m³ yearly right after this maintenance.

The company has, therefore, documented that if they did not use the 100,000 DKK their WW would have been 20,000 m³ higher. This corresponds to the company trading 1 DKK for 0.2 m³ WW, which means that the last two iterations in Table 5.1 probably are too harsh. In the company's documentation, they do of course not take into account their own inefficiency and that they probably extend the pipes life span among other things. It is, however, still a good indication that the trade-offs removed in the last iterations are realistic and, therefore, should be allowed back in the model

5.4 theoretical implications

While the reasoning for removing outlier trade-offs (facets) should be clear by now, the theoretical implications can be discussed. In the dual DEA formulation, it seems reasonable to add weight restrictions due to unrealistic trade-offs – we simply remove these trade-offs. In the primal space, however, the inclusion of weight restrictions leads to new artificial points in the technology. These artificial points occur due to the assumption that it is now possible to substitute one input in favour of another, exclusively based on the ratio given in the weight restriction and not based on the observed data set. This method, therefore, differs from other DEA outlier methods because points are, in effect, added in order to remove outlier trade-offs instead of simply removing outlying data.

The proposed outlier model is, however, enticing because it does not require any prior information about the trade-offs and thereby exclusively is based on the observed data. At the same time it has the disadvantage that DEA becomes less conservative by allowing non-observed data points in the technology.

6 Changes in efficiency scores

As mentioned earlier, there is a political desire to incorporate quality but without compromising the efficiency requirement for the companies costs. In this section, we use a standard radial DEA model based on the facets found in section 5 to assess, if this model could be suitable for KFST. Note that any other distance functions could be implemented as well, because we have ensured that every facet is realistic. The projection to these facets should, therefore, also be realistic no matter the choice of direction. For simplicity, we here only use the radial DEA model. It will obviously be relevant for KFST to explore additional directions and especially the one where only costs are discretionary. Figure 6.1 shows the companies efficiency scores in three different models. The dots show the efficiency scores in a model without quality. The upper bar shows the efficiency scores in a model with quality but without any weight restrictions and the lower bar incorporate all the weight restrictions discussed in this paper.

The figure shows that most companies' efficiency scores increase a lot when we go from a model without quality to the model with quality but without any weight restrictions. This is expected, given the inclusion of three additional variables, yet not consistent with the political agenda. When we add the weight restrictions, first for the willingness to pay and hereafter for the outliers, most efficiency scores get closer to the initial model without quality. The inclusion of the three quality variable does, however, still increase the efficiency scores a lot for several companies, even with the weight restrictions, which means that the model may not be harsh enough from a political point of view, as the intention was not to reward companies for quality (in terms of lower reduction requirements on costs) but mainly to punish, or at least not provide incentives for moving towards, poor quality.

However, for some companies, the efficiency scores are lower in the model with quality and weight restrictions than in the model without quality. This seems to go against the standard DEA ideology giving companies the benefit of the doubt such that the inclusion of more variables will never decrease the efficiency scores. From this point of view, the model may seem too harsh. If KFST chooses to only use the weight restriction from the willingness to pay, this problem does not occur for any of the companies. Another solution for this potential problem could, for example, be to use a so-called "best-of" model, where we use the highest individual efficiency score for the model without quality and the model with quality and all the weight restrictions. The

use of “best-of” models is often used in regulations and it is therefore likely, that such a model can be accepted for both KFST and the sector.

The mean efficiency scores in the model without quality is 0.06 lower than the mean in the model with all weight restrictions. If we compare the companies’ revenue caps in the two models, the proposed model with weight restrictions will increase the average revenue cap with 9.19 %. This means, that the consumers’ water expenses will increase by the same 9.19 % assuming that the companies charge the maximum of what they are allowed and do not change behaviour in other ways. However, the companies will now get an incentive to preserve high quality and are, therefore, expected to change behaviour. This will likely improve the quality in the sector and it is beyond the scope of this paper to assess if the price of achieving this is too high.

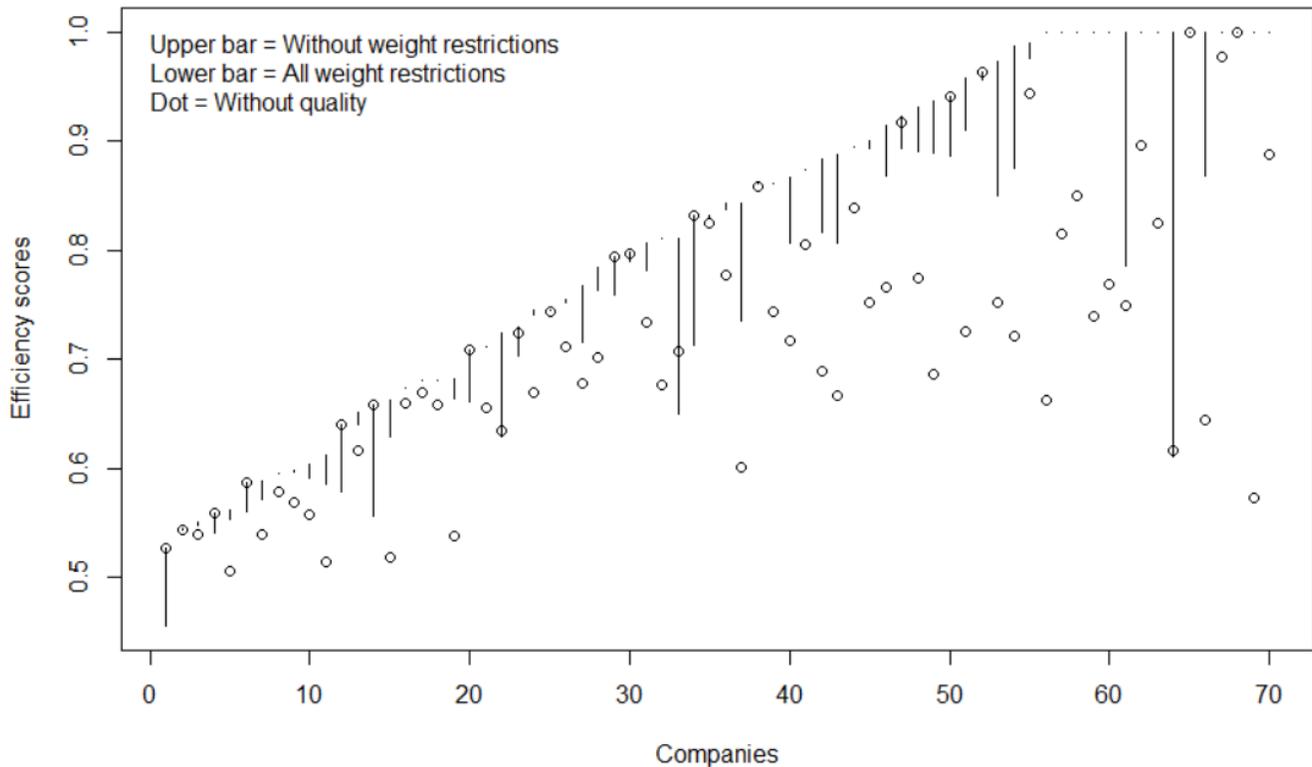


Figure 6.1 – Changes in the efficiency scores

7 Conclusion

The benchmarking model currently used to regulate the Danish water companies is likely to suffer from an omitted variable bias due to not considering the quality of the provision. Including the proposed quality variables is, however, not a trivial undertaking mainly due to two issues. First, the political agenda is that the quality variables should not reduce the cost-efficiency requirements too much. Second, the regulator lacks prior information about the relationship between the quality variables and the remaining input and outputs. This reduces the regulator's options to include the quality without compromising the cost-efficiency requirements to much.

We proposed to analyze the facet structure using Qhull to gain valuable information about the estimated trade-offs between quality and costs. We used this information in an outlier analysis. The outliers were later used to expand the technology using weight restrictions and thereby reduce the influence that quality had on the efficiency scores.

In addition, we showed how the incorporation of the customers' willingness to pay in the form of a single weight restriction improved the facet structure in several dimensions and forced the model to not assign more weight to quality than the customers are willing to pay.

We proposed to allow the companies to apply to reintroduce some of the tradeoffs, i.e. remove some of the weight restrictions eliminating outliers, if they can argue and documents that the trade-offs are in fact realistic. This gives the companies an incentive to disclose their private information about their true trade-offs without having an administratively cumbersome regulation.

Lastly, we argued that one of the advantages of analyzing and specifying the facet structure independently from the efficiency measurement is that any direction of projection onto the frontier can be used on this technology without risking getting unrealistic trade-offs. Multiple efficiency measures can therefore be compared in the same technology.

For further research we suggest to use a multi-dimensional outlier detection method. For the iterative trade-off outlier detection approach proposed in this paper, we have used a single-dimensional outlier method - the adjusted interquartile range. This method is simple and commonly used, making it an obvious candidate for our proposed idea, where outlier trade-offs are removed in order to subsequently get the companies to reveal private information about their true trade-offs. It is, of course, trivial to instead use another single-dimensional outlier method to identify the outliers and still use weight restriction to remove them (iteratively or not, and in some pre-specified order).

It is, however, more complicated to use multi-dimensional outlier detection methods, but we might need to then look for outlier facets rather than outlier trade-offs. To remove an outlier facet we could extend another similar facet using, for example, the k-means algorithm to identify similar facets (Hartigan & Wong, 1979).

8 Bibliography

- Agrell, P. J., Bogetoft, P., & Tind, J. (2005). DEA and Dynamic Yardstick Competition in Scandinavian Electricity Distribution. *Journal of Productivity Analysis*, 23, pp. 173-201.
- Aigner, D. J., Lovell, C. A., & Schmidt, O. (1977). Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics*, 6, pp. 21-37.
- Asmild, M., Paradi, J. C., & Reese, D. N. (2006). Theoretical perspectives of trade-off analysis using DEA. *Omega*, 34(4), pp. 337-343.
- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996, December). The quick hull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4), 469-483.
- Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp. 429-444.
- Dai, X., & Kuosmanen, T. (2014). Best-practice benchmarking using clustering methods: Application to energy regulation. *Omega*, 42(1), pp. 179-188.
- Hartigan, J. A., & Wong, M. A. (1979). K-means clustering algorithm. *Applied Statistics*, 28(1), pp. 100-108.
- Heesche, E., & Asmild, M. (2020, 3). Controlling for environmental conditions in regulatory benchmarking. *IFRO Working Paper*.
- Hodge, V., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22, pp. 85-126.
- Hubert, M., & Vandervieren, E. (2008). An Adjusted Boxplot for Skewed Distributions. *Computational Statistics and Data Analysis*, 52(12), pp. 5186-5201.
- Konkurrence- og Forbrugerstyrelsen. (2020). *Forbrugernes betalingsvilje for forbedringer i vandsektoren (In Danish)*. Konkurrence- og Forbrugerstyrelsen. Retrieved from <https://www.kfst.dk/media/igmpq3o4/20201009-forbrugernes-betalingsvilje-for-forbedringer-i-vandforsektoren.pdf>
- Meeusen, W., & Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, 18(2), pp. 435-444.
- 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*, pp. 157-306.
- Petersen, N. C., & Olesen, O. B. (2015). Facet Analysis in Data Envelopment Analysis. In Z. J., & Springer (Ed.), *Data Envelopment Analysis: A Handbook of Models and Methods* (Vol. 221, pp. 145-190). International Series in Operations Research & Management Science.
- Podinovski, V. V. (2019). Direct estimation of marginal characteristics of nonparametric production frontiers in the presence of undesirable outputs. *European Journal of Operational Research*, 279, pp. 258-276.

Ramanathan, R., Ramanathan, U., & Bentley, Y. (2018). The debate on flexibility of environmental regulations, innovation capabilities and financial performance – A novel use of DEA. *Omega*, 75, pp. 131-138.

Scheel, H. (2001). Undesirable outputs in efficiency valuations. *European Journal of Operational Research*, 132(2), pp. 400-410.

A. Appendix - Robustness

We have shown that our proposed method reduces the trade-off intervals and reduce the large increase in the efficiency scores from the addition of the quality variables. As a robustness check, the method has been applied to data from 2019. Note, however, that a direct comparison between the years is problematic due to differences in the definition of the variables between the years and the lack of quality check of the data from 2019. Especially the results for BE should not be compared directly due to a radical new definition of Bacteriological Excesses..

If we compare Figure 5.3 and Figure 8.1, we first observe that the number of trade-offs for Costs/BE has been reduced a lot. This is because most companies have zero BE with the new definition and its influence on the final results have, therefore, decreased.

Second, we observe that the lower bound of the weight restriction for $\frac{Costs}{PB}$ is the same because we have used the same willingness to pay in the two years. The upper bound is almost the same, but have occurred naturally in the first year and with a weight restriction in the second. This indicates that the high trade-offs in the second year were in fact outliers.

Lastly, we observe that the interval for the trade-offs between costs and WW are also almost the same for the two years. The interval is, however, a bit larger in the second year where there is no weight restriction on the lower bound. That the lower trade-offs yet are similar could again indicate that the outliers found in the first year are in fact outliers.

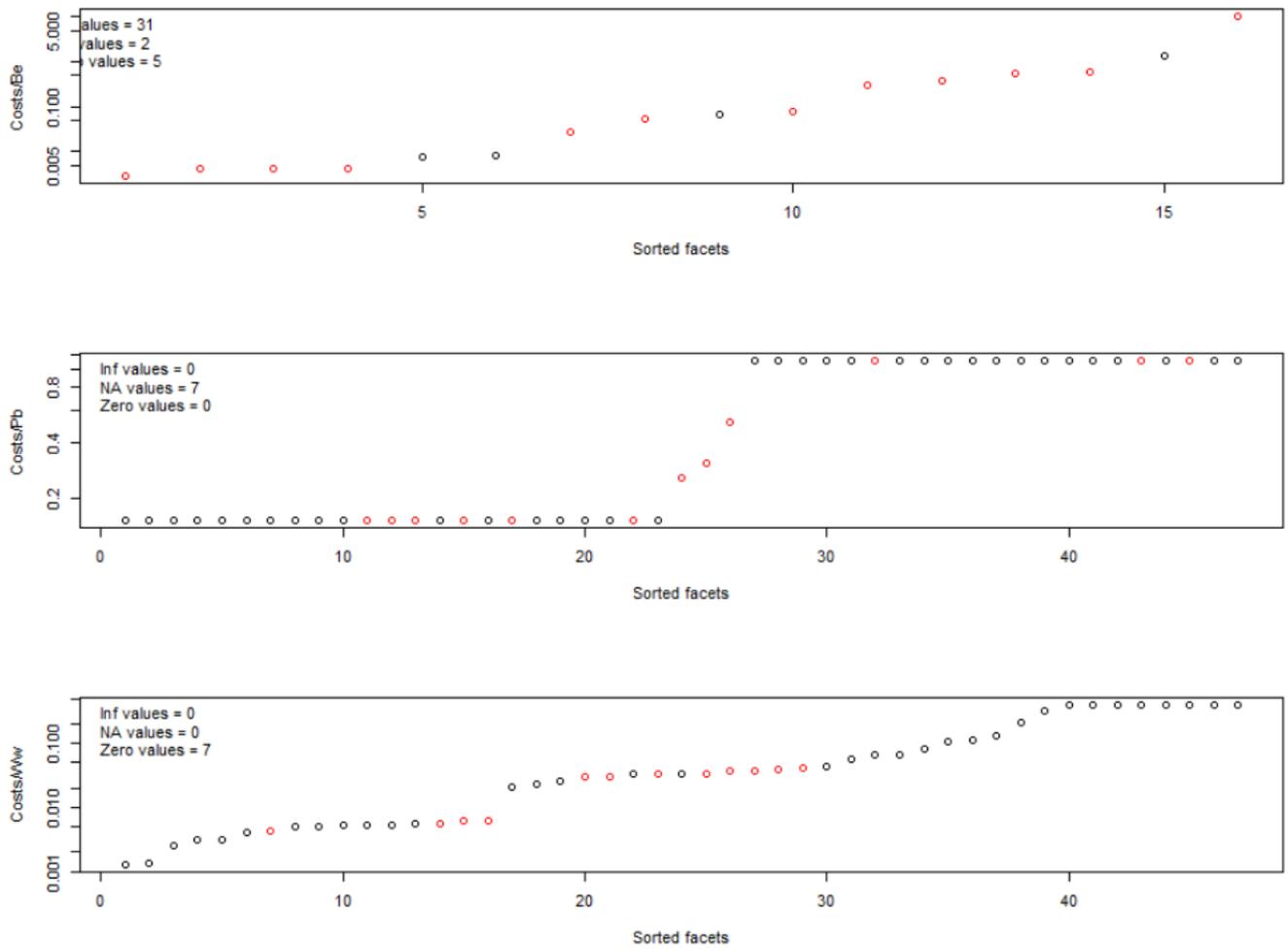


Figure 8.1 - Distributions of trade-offs between costs and quality in a new data set. Note that the zero, Inf and NA values are omitted. The red dots indicate trade-offs from a fully dimensional facet and the y-axis is logarithmic