Efficiency and Quality in the English and Welsh Water and Wastewater Industry

John E. Peirce, Cardiff University

A Thesis submitted for the fulfilment of the requirements for the Degree of Doctor of Philosophy at Cardiff University

The Economics Section of Cardiff Business School, Cardiff University

January 31, 2024

Abstract

In 1989, the English and Welsh Water and Sewerage Industry privatised, on account of Thatcherite liberalisation efforts of public utilities, motivated in part by the nationwide failure to meet multiple European Economic Community directives for Bathing and River Water Quality. Since then, the industry has been regulated, with periodic updates to the regulations set by Ofwat, the regulating body.

The research literature has concerned itself with the incorporation of quality in the industry since Saal & Parker (2000)'s use of quality indices to adjust the final water and sewerage service outputs of companies. However, as Saal himself reports in Saal et. al. (2017), these indices have 'stagnated' – they are not longer useful measures of quality, on account of the whole industry reaching near full compliance in these quality standards. These measures also face issues around the selection of and assumptions on their data, such as assumed fixed quality before the indices' reference year, or little variation in the measures selected for their indices, as well an assumed exogeneity of quality in the industry given the measures' applications as scaling factors to production outputs.

This thesis aims to develop a new, Composite Indicator of overall industry quality, utilising some of the newer regulatory targets, the Common Performance Commitments, introduced in 2014. This use of newer regulatory targets allows for the measurement of industry quality over a long time period, using targets common to all companies in the industry, with consistent data under current regulatory scrutiny, rather than traversing the difficulties of, say, the individualised K-factors used for price cap regulation. Using DEA modelling, the thesis first intends to see if the addition of the new indicator as an additional production output significantly changes the Technical Efficiency scores of companies, compared to older DEA models using the allegedly superfluous measures. The thesis then aims to perform the same exercise in DEA models with Quasi-Fixed

Capital, as to determine if the addition of a quality investment output addresses Capex Bias in the industry, as reflected through Allocative Efficiency. Finally, the thesis looks at and discusses various extensions and future directions for the composite indicator, focusing on its dynamic properties, and interactions with measures of extreme weather. Focus on how Welsh Water, the only non-profit company, compares to its for-profit counterparts is also given.

The thesis finds that the new composite indicator of quality is significantly more volatile and less complied with on average, finding a 42.5% difference in quality compliance on average between the old and new measures. When used in DEA models, the thesis finds a significant change in companies' technical efficiency scores; this is not true, however for allocative efficiency in the dynamic DEA models. Limited evidence about the composite indicator's correlations with extreme weather is found, and the dynamic properties of the indicator suggest that the overall quality improvement over time is limited, if not negligible. Finally, though non-profit behaviour seems to yield lower allocative efficiency in dynamic models, there is evidence of greater technical efficiency over time with the quality indicator as an output, compared to for-profit companies on average.

Acknowledgments

I am first and foremost grateful to the ESRC Wales DTP programme, whose scholarship enabled this thesis in the first place – it is thanks to their opportunity that I get to achieve my dream. I am next grateful to my supervisors, Prof. Kent Matthews and Dr. Charlotte Beale, whose guidance and patience with me throughout my thesis has been invaluable. I am grateful to my viva examiners, with whom I enjoyed a deep and thorough discussion of my work, and who have also helped develop this thesis into its final state. I am also greatly appreciative of the many academic colleagues who have attended my presentations of my research in the Welsh Postgraduate Research Conference 2023, the Gregynog Colloquium 2023, and the various Economics PhD Workshops in Cardiff Business School, whose feedback and discussion of my work has helped frame what I have done better than I could have done alone. A special thanks is also owed to people of Cardiff Business School themselves, specifically Denise Brereton, who have helped provide me a place that I can call home, and who have been a great help through the more turbulent parts of the thesis, as well as Jon Osborne, who introduced me to the subject ten years ago, and was the support I needed at the start to end up here.

I am of course grateful to my Mum, Dad, Sister, Grandma and Step-Parents, whose support has surely helped keep me afloat in those times where it has been the hardest. Similarly, my best friends Kieran Pinks and Matthew Miles, as well as my colleagues and close friends Suzanna Nesom, John Poole, Yuanli Guo and Julia Diniz, have all been invaluable in keeping me just well enough to complete this work through their friendship, support, and kindness. I offer this same thanks to the many members of D46 in the business school, who have made this experience one with a fantastic community, despite the difficulties faced throughout the project.

A particularly special thanks goes to Nicola Thompson, my partner, who

has in the past year provided a blinding, brilliant light in the dark that I have followed to a better place – her support, perhaps more than any other, has become the motivator to finish this thesis, and to carry on working towards my goals thereafter. *Dw i'n dy garu di i'r sêr ac yn ôl, fy nghariad.*

Contents

Key Acronyms

1SDEA: One-Stage Data Envelopment Analysis

3SDEA: Three-Stage Data Envelopment Analysis

AMP: Asset Management Plan

APR: Annual Performance Review

BoD: Benefit of the Doubt

CCR-DEA: Charnes-Cooper-Rhodes Data Envelopment Analysis

CI: Composite Indicator

CMP: Cost Minimisation Problem

CNLS: Convex Non-Parametric Least Squares

C^{2}NLS Corrected Concave Non-Parametric Least Squares

COLS: Corrected Ordinary Least Squares

CPC: Common Performance Commitment

CPI-H: Household Consumer Price Index

CRS: Constant Returns-to-Scale

DEA: Data Envelopment Analysis

DMU: Decision-Making Unit

DWI: Drinking Water Inspectorate

EEC: European Economic Community

FDH: Free Disposal Hull

FOC: First Order Condition

FOCS: Fixed Opex-Capex Share

FP: For-Profit

H4CSF: Homoskedastic Four-Component Stochastic Frontier

IQR: Inter-Quartile Range

(I)RDM : (Inverse) Range-Directional Model

JAR: June Annual Return

KDE: Kernal Density Estimation

MEA: Modern Equivalent Asset (Ch. 4)

MEA: Multidimensional Efficiency Analysis (Ch. 3)

NDRS: Non-Decreasing Returns-to-Scale

NFP: Not-For-Profit

NIRS: Non-Increasing Returns-to-Scale

NRA: National Rivers Authority

NRW: National Resources Wales

ODI: Outcome Delivery Incentive

OLS: Ordinary Least Squares

PC: Performance Change

PCA: Principal Component Analysis

PDF: Probability Distribution Function

PMP: Profit Maximising Problem

PR: Price Review

QP: Quadratic Program

RCV: Regulatory Capital Value

RoR: Rate of Return

RPI: Retail Price Index

RWA: Regional Water Authority

SFA: Stochastic Frontier Analysis

StoNED: Stochastic Non-Smooth Envelopment of Data

SWC: Statutory Water Company

VAR: Vector Auto-Regression

VRS: Variable Returns-to-Scale

WACC: Weighted Average Cost of Capital

WaSC: Water and Sewerage Company

WOC: Water Only Company

List of Water Company Acronyms

Company:	Acronym:	Company:	Acronym:
WaSCs:		**WOCs:**	
Anglian Water	ANG	Affinity Water	AFW
Hafren Dyfrdwy	HDF	Bristol Water	BRL
Northumbrian Water	NWL	Portsmouth Water	PRT
Southern Water	SRN	Sutton & East Surrey Water	SES
Severn Trent Water	SVT	South East Water	SEW
South West Water	SWT	South Staffordshire	SSC
Thames Water	TMS	& Cambridge Water	
United Utilities Water	UUW		
Welsh Water/Dŵr Cymru	WSH		
Wessex Water	WSX		
Yorkshire Water	YKY		

Chapter 1

Introduction

The water industry is widely considered to be the first public utility and, in its provision of what is, according to Lampard (1973), "the *sine qua non* of the city," there is a need to best provide water and wastewater services to the industry's customers in the best manner possible: at the cheapest price, causing the least external damage in its production and provision, and of the best quality producible. In the efforts to meet all of these standards, the industry faces consistent trade-offs between its customers, the environment, and company financing, which have pervaded water and wastewater management both historically and in the current day.

Such requirements of the industry are under constant scrutiny in England and Wales where, after the privatisation of the industry in 1989 in part because of wholly inadequate water environmental quality, both industry policy and performance measurement have been evaluated in both a regulatory and academic environment, with the then-founded regulator Ofwat requiring periodic Price Reviews throughout the industry as to constantly review and compare company achievements in five-year periods, and various academia studying the impacts privatisation has had on costs, efficiency and productivity, amongst other economic factors.

This thesis wishes in spirit to return to the issue of Quality in the industry. Since privatisation, there have certainly been vast improvements in the water quality issues of the time. Yet, as a report by Saal et. al. (2017) finds, such measurements are 'stagnant', in that their inclusion in industry decisions are no longer important, as the industry as a whole has almost uniformly fully complied with those environmental quality standards. As the report suggests, newer measures of quality should now be used instead, and though Ofwat has introduced such measures in the Common Performance Commitments (CPCs) for each company, this thesis seeks to further investigate how quality is incorporated in the evaluation of the industry, and by extension how quality affects industry performance.

To introduce the place of the thesis in the industry literature, the necessity of quality in regulatory decisions will be used to motivate a place for the forthcoming research questions, which will then introduce what this thesis seeks to explore, and how this research is to be carried out.

1.1 What is Quality?

The notion of 'Quality' is terribly broad, even after reducing it to matters in the water and sewerage industry. This section will briefly describe what is meant by quality in this chapter, and the thesis hereafter.

Prior to privatisation in England and Wales, the historical notions of quality often discussed pertained to drinking water and whether any effluent was treated at all (Hassan (1988)). As water infrastructure was first developed in the 19th century, the key issues of quality were the assurance that people in the municipalities at the time could receive water and, eventually, that the waste produced by the people could be taken away and treated in some fashion.

As municipal control of water and sewerage failed, primarily on the basis of insufficient water supply and sanitation in rural areas (Ofwat (2008)), poor industry management, and a lack of accountability from firms to polluters (Hassan (1988)), the industry became controlled by more and more centralised powers, eventually becoming a nationalised industry. By this point, it was not just the provision of water and the notion of treating sewerage that defined quality, and instead the definition shifted to clean, unpolluted waters and sewerage treated many times, so that its return to the water cycle did not also pollute.

Once again, this failed to be upheld, so much so that the industry had to privatise in 1989, in part as a commitment to recover English and Welsh waters from the polluted state they were in. Quality at this stage was defined by the regulations that the industry had summarily failed: the European Economic Community (EEC)'s standards for River and Bathing Water quality. This fed into what the research literature would utilise as the primary adjustment for water and wastewater quality - the quality indices defined by Saal & Parker (2000), whose ad hoc measures of quality followed limited earlier attempts to incorporate some account of quality (Lynk (1993), Hunt & Lynk (1995), Cubbin & Tzanidakis (1998)).

Quality then became a more prolific point of interest, at least from the academic perspective. Research added the convention of accounting for quality via the exogenous output adjustment indices of Saal & Parker (2000), and often made further examination of the effects of quality adjustments in various areas of literature, such as total factor productivity (Maziotis et. al. (2016)), profit decomposition (Maziotis et. al. (2014)), and allocative efficiency (Pointon & Matthews (2016)), amongst others. On the regulatory front, the standards once failed by the industry at large were met to near-uniform complete compliance, and in 2014 further significant quality-related regulatory targets were introduced, known as Common Performance Commitments (CPCs), which were targets that all water companies had to meet, related to issues such as flood resilience, pollution incidents, water leakage and water use per capita. Furthermore, in recent literature, some focus on service quality variables has been had (Molinos-Senante et. al. (2015)), furthering the exploration of facets of quality and how they impact the water and wastewater industry.

1.2 Motivations for the Thesis

As has already been mentioned, one of the reasons for privatisation of the water industry in 1989 was the failure to meet the required standards of quality, as provided by the government and the European Economic Community. One reason for this inadequacy can be explained by the Natural Monopoly structure of the industry: for each company, there exists a region of England and Wales in which they are the sole or majority provider of water services, not out of monopolistic behaviours forcing competitors out, but out the prohibitively expensive infrastructure required to enter the regional markets. So it was, and still is to a large extent, that each Regional Water Authority (RWA) could afford to behave as monopolists, due to their 'natural' right as the only authority for water and sewerage services in that region. Despite the nationalised state of the industry at the time, no government pressure succeeded in meeting the environmental standards of the time, and to address this problem, the industry was privatised.

So, in 1989, these RWAs were privatised as part of Margaret Thatcher's efforts to privatise public utilities, and were placed under Ofwat, the regulatory body for the water industry. Price Caps were introduced, as to curb any excessively high prices for the companies' services, and such caps were defined by the $RPI + K$ system, the K factor of which being used as a measure of a company's improvements in quality and in relative performance. This measure, as Littlechild (1988) highlights, is successful only if the factor for improvement K can be correctly identified for each company, and so it is one objective of the price reviews to update and review company price caps by evaluating this allowance for improvement, K.

This factor for improvement can be further decomposed, giving the following price cap:

$$\bar{P}_i = RPI + Q_i - X_i \tag{1.1}$$

Where, for each company i, Q_i is the allowance for the price cap specific to Quality Improvements, and X_i is a quantity that accounts for Productivity - to best improve their price cap, then, a company should improve quality, and improve productivity such that the determination for the X factor is minimised. Both factors then adjust a price index, which is historically RPI[1]. What this shows, particular to the motivation of this thesis, is that quality is in one sense important for regulation, in that it is used to determine a company's price cap allowance over a five-year period covered by the price review in which it is set.

That isn't to say that this is the only appearance of quality as a point of concern for companies, however. In PR14 - 2014's price review - Ofwat introduced the CPCs as a sub-class of Outcome Delivery Incentives (ODIs), all of which are targets set for companies to achieve with financial benefit upon achievement, or more commonly financial penalty upon their failure. CPCs, as their name suggests, are common to all companies to whom they are valid: those CPCs that refer to sewerage services are not relevant to Water Only Companies (WOCs). These targets aim to incentivise improvements in a variety of quality factors, be it infrastructural, environmental, customer service, or otherwise and, in a similar sense to the Q factor in the price cap, can change the allowances given to each company, based on the fulfilment of these common commitments to quality.

The CPCs are certainly a good point of focus for modern day quality concerns, but other measures precede them, such as Saal & Parker (2000)'s water and sewerage indices. One of the significant motivations for this project is based on both the fact that the CPCs, except for their introduction into the regulatory process, have been relatively unexamined, but another point of motivation lies in the fact that these older indexed quality measures have, as Saal et. al. (2017) points out, all but become completely uniformly met by the entire industry. To best promote quality improvements, then, it would be useful to find new measures that are far more volatile and therefore require far more improvement

[1]As of PR14, however, CPI-H has been used for price indexing instead.

than these older quality measurements.

One trade-off of this inclusion of quality in (1.1)'s price cap definition is the potential for capex bias. Averch & Johnson (1962) define such a bias as an over-investment of companies into short-term capital projects, and find that one solution to the issue is via rate-of-return regulation, such as price caps. However, the paper only considers an adjusment accounting for productivity - the X-factor in (1.1). The addition of a further quality adjustment factor, therefore, may have unintentional biasing effects that offset the proposed solutions to capex bias. This is further compounded by the shift in focus to ODIs from 2014: this significant regulatory change may imply a greater importance placed on Q, rather than X, amplifying the potential quality-related capex biases.

A final point of motivation comes from the aforementioned inclusions of quality, and is best expressed by a question: How should quality be included?

From an academic point-of-view, the older quality indices were used to adjust water and wastewater outputs, respectively. This supposes that quality is externally measured, and then is accounted for in the outputs through scaling down the produced services. However, with the CPCs acting as an even larger point of focus for companies than quality might have been before, is it right to suggest that quality is only treated as an exogenous adjustment to outputs?

This thesis intends to test the idea of quality as a more significant part of the production process. One point-of-view is that, alongside outputs, quality improvements could also be treated as an output, in that companies can invest in water service production, wastewater service production, and quality improvements, rather than only the former two as has been the convention in previous research. Another perspective, when looking at a dynamic model, is to assess how quality changes over time, as to see if any dynamic patterns emerge, and what might be some influencing factors on its behaviour year-on-year, or in each price review period. This increase in the importance of quality in the production process can be justified quite readily by the collection of contemporary issues in the water of England and Wales, such as the decline of native fish populations and increase in chemical pollution in the waters, or the increase of sewage overflows that further damage the industry's water sources and create risk for its customer base (Wye Salmon Association (2019), House of Commons (2022)). Atkins & Peirce (2024) provide a case study on the River Wye, highlighting these exact problems, which are all concurrently harming the river in a manner that is both urgent and potentially irreversible.

With this all in mind, this thesis proposes research questions with the aim of evaluating how new approaches to quality and quality measurement could be used in the modelling process, as to see if quality has significance elsewhere in industry regulation.

1.3 Research Questions and Contributions

This thesis proposes five research questions:

1. To what extent can a new measurement of Quality be derived, which accounts for the newer, broader definitions of factors of quality, as illustrated by the Common Performance Commitments?

2. Does this new measure of quality, when included as an Output in production, yield significantly different Technical Efficiency Scores for companies, compared to older models?

3. Using Dynamic Models, to what extent have the recent regulatory changes affected measures of Efficiency and Capex Bias over time?

4. To what extent are there Dynamics of quality, and, in dynamic models, to what extent does the novel inclusion of quality affect Capex Bias, by way of affecting Allocative Efficiency?

5. Throughout the previous research questions, how does Welsh Water, the only Non-Profit company in the industry, differ in terms of results from its other industry counterparts?

Research Question 1

This first question broadly covers the motivations above that are concerned with how quality is to be measured in the future, as to provide far greater scope for improvements, now that the older measures do not significantly provide such targets for companies. Accounting for the quality of water and wastewater services in efficiency measures allows for more accurate determinations of firm performance, and so it is imperative to have measures of quality that are useful both in empirical exercises and more widely in the industry.

One of the difficulties in addressing this question relates to how best to incorporate the multitude of different facets of quality that are now observed through the CPCs and other measures. The objective of this question, then, is to contribute to the industry and its literature by proposing a new method to measure quality which has the capacity to account for various factors of quality, and need not be necessarily limited to one facet of the industry, as the older indices were by design.

The thesis develops a Composite Indicator of quality, and finds a significant 42.5% difference between its measure of overall industry quality, and the older Saal & Parker (2000) measures of water and wastewater quality. The new measure is more volatile than its predecessors, and shows a slight downward trend over time, suggesting an urgent need for improvement in quality across the industry, and a large capacity for improvement of the quality measure.

Research Question 2

This question, based off of the results of the prior question, seeks to assess how a novel measure of quality affects models of production via its inclusion not as an adjustment to outputs, but as its own unique, additional production output. Using the non-parametric Data Envelopment Analysis (DEA) models, which are common to the literature, the aim is to test older specifications of quality against a new inclusion of quality.

This question aims to contribute to the empirical literature by evaluating how research considers quality in its models. Some recent papers, such as Molinos-Senante et. al. (2015), do include customer service quality measures as outputs, and so this question continues along that line of thinking, while also testing the novel measure of quality that covers multiple facets of quality in the industry.

The thesis finds that the addition of the new composite quality measure developed for the first research question, when added to DEA models as a production output, yields significant differences in technical efficiency when compared to previous DEA models which use the older, exogenous output-adjusting quality indices. It is also found that those older measures do not have a significant difference in efficiency scores from models that have no measure of quality, affirming the notion that these older measures are stagnant.

Research Question 3

An important facet of the research output of this topic is to assess how effective regulatory changes through the price reviews were in hindsight. This kind of evaluation is often a consequence of different research using yet-more recent spans of time in their research, and is no different in this thesis.

The aim of this question is to evaluate the models of Pointon & Matthews (2016) with a more recent window of time, covering the more recent price reviews. In doing so, particular attention is paid to the years following PR14, where the CPCs were introduced, but also to the period where Ofwat ordained a focus on Totex in their modelling, rather than Capex and Opex separately, so as to try and lower Capex Bias over time.

The thesis finds that, with the caveat of having to employ less reliable proxies for prices and capital than previous research, there is no significant differences in allocative efficiency between dynamic and static DEA models, suggesting that quasi-fixed capital has been identified properly. However, technical and overall efficiencies are significantly different, implying that, though the capital may be correctly allocated, it is not best employed throughout the industry.

Research Question 4

Following on somewhat from the previous question, this question seeks to extend the motivations for testing the inclusion of quality in the production process by employing Dynamic DEA models, as in Nemoto & Goto (2003) or Pointon & Matthews (2016).

The aim of this question, and to some extent the previous question also, is to provide some new potential empirical models by which the industry's comparative efficiency scores can be measured for use in the regulatory process. More specific to this question, the aim of these models is also to see how, via changes in Allocative Efficiency, the novel inclusions of quality might change Capex Bias in the industry, on the idea that, by incorporating quality explicitly into the model, some of the over-investment into capital could be accounted for because it is used for quality improvements.

The thesis finds that, much like the models of the previous research question, there is no significant difference in allocative efficiency between static and dynamic DEA models that incorporate a composite measure of quality, and that technical and overall efficiencies do demonstrate significant changes. Further, by comparing these models to the DEA models without quality measurements, it is found that there are no significant differences between efficiencies.

Research Question 5

This final question is both over-arching and secondary to the other questions, but nonetheless could provide some interesting insight into how, if at all, non-profit behaviour yield different results in the models of the thesis with respect to quality improvement and inclusion. Were non-profit behaviours found to be strictly better from a quality-improving standpoint, for example, then there is scope to incentivise the re-production of those behaviours in the other for-profit members of the industry, if possible.

The aim of this last question is to observe such behaviours if they exist, and however they exist, to see whether there are meritorious behaviours in quality improvement and production that not-for-profit companies might promote through changes in efficiency, capex bias, and the evaluations of the new quality measurement.

Broadly, the thesis finds that there are some differences between non-profit and for-profit efficiency on average, namely that non-profit efficiency tends to be lower on average than for-profit average efficiency, but has less variance. In some models, such as those that account for differences in operating environments in the industry, non-profit technical efficiency can be found to be higher than the average for-profit score.

The thesis is split into two broad parts, which are in turn defined by three chapters each. The first part goes through the prerequisite knowledge for the second part: Chapter 2 covers the literature of the industry from a historical point-of-view, outlines how the industry functions in its current state, and seeks to evaluate the industry when compared to other water industries, and other public utilities; Chapter 3 is another literature review of sorts, but covers the economic theory and empirical literature that is required to explain why the models of the research questions are used, and how they represent the production process; Chapter 4 goes over the data used throughout the literature, as to inform the choices of data chosen in the contributing research of the thesis.

The second part, then, contains three contributing chapters: Chapter 5 introduces a novel measurement of quality, and first tests this measure by comparing its inclusion in a DEA model as an output to older typical DEA specifications, and so answers the first two research questions; Chapter 6 answers the third and fourth questions, by developing dynamic DEA models to evaluate the impact of the new quality measure on capex bias and its behaviour in a dynamic setting, as well as the regulatory impact, if any, on the older models. The second part also addresses the last research question in all chapters, distinguishing non-profit results from the rest of the industry, as to see if the new models demonstrate any significant changes because of non-profit behaviour. Chapter 7, in a departure from the most of the research questions, capstones the thesis by discussing future research directions, assessing the plausibility of future industry behaviour based on the conclusions drawn from the previous contributing chapters.

Part I

Prerequisite Chapters

This part of the thesis seeks to introduce the necessary contexts required for the latter research chapters.

To do so, Chapter 2 first sets up the context of the industry overall, looking at its history and regulatory development post-privatisation from a Quality-focused point-of-view. Chapter 3 then covers the requisite economic theory needed for the eventual models of the research chapters, while also covering some of the other methodologies that have produced results elsewhere in the literature. Chapter 4 concludes this part by outlining what data has been used in the literature, and what data will be chosen to use in the chapters thereafter.

Chapter 2

Literature Review

To understand the context in which Quality has played a role in defining a lot of the English and Welsh water and sewerage industry's developments, it is useful to first have a chapter covering the literature and history surrounding the industry. This chapter, then, seeks to first run through how the industry has historically concerned itself with quality, and then through how the industry has developed into its contemporary design.

Since one of the industry's most significant changes was its privatisation in 1989, and since much of the research in the field has sought to evaluate this decision after the fact, many papers of the literature will be discussed, as it pertains to the choice of privatisation, but also to how quality was considered and evaluated as well. Much of the empirical modelling used in these papers will be looked over in Chapter 3, which will help to determine how this project will address its research questions.

2.1 History of Quality in the English and Welsh Water and Sewerage Industry

The history of the water and sewerage industry in England and Wales can be broadly defined by three parts: the pre-privatisation and post-privatisation periods, within which the pre-privatisation period can be further separated into a pre-nationalisation and post-nationalisation period. As it pertains to the investigation into quality, much of the former time period will address the inadequacies that led to the privatisation of the industry in 1989, which itself has a section devoted to its incursion, and how it related to quality standards at the time. The latter section will utilise the bulk of the literature in the field, relating to findings about the privatisation, and results that pertain to quality in the industry.

2.1.1 Pre-Privatisation

As far as the concern about quality goes in the industry, such a matter has been pertinent since at least the 1800s, with Hassan (1998, pg.18) highlighting that the initial development of municipal water suppliers mostly in the 1861-81 time period came about due to a want for cleaner water to prevent disease, among other things. Despite environmental costs, these new suppliers succeeded in developing increased potable water of acceptable quality, without failing to properly supply both industries and trade (ibid., pg.24-25).

However, persistent issues with environmental quality began soon after, with the deterioration of this era's river quality attributed to an over-prioritisation of water services relative to care for wastewater services, with as many as 132 of the 178 providers with contemporary sewage systems failing to treat sewage before its disposal - water, it seemed, far outweighed wastewater in the minds of the regional authorities (ibid., pg.26, 28-29).

These problems were compounded by further issues related to regional coordination and ownership, with quality being of less importance than land value, land ownership, company shares, and so on (ibid., pg.48). By 1945, there were more than 1000 water suppliers in the industry, and approximately 1400 sewerage and sewage disposal bodies. At this point in the industry, there became a sentiment post-war to consolidate these local bodies, for the purposes of better planning, control over pollution, and to better meet the increasing demands for water.

Post-war legislation, in addition to the aforementioned advantages, also sought extend water and sanitation services to the country's more rural areas via public investment. Following droughts in 1959 and floods in 1960, the 1963 Water Act introduced an administrative system for abstraction permits - permissions to take water from its sources for various uses - so that there was more security with respect to the conservation of current and future water resources.

The issues around pollution in the industry were not solved by this consolidation, however. Following these failures and further increases in forecasted water demand, the 1973 Water Act prompted the industry's nationalisation, establishing 10 Regional Water Authorities (RWAs), responsible for both water and wastewater activities, and 29 privately operated Water Only Companies (WOCs). Per Saal & Parker (2001), the RWAs covered about 75% of the water operations in the country, with the remaining water services achieved by the WOCs. Control over the investment into these authorities was given to the government, instead of the local authorities, with the hope that government planning can facilitate the forecasted water demand, and improve upon the industry's water and wastewater quality.

RWAs were also set to operate under a Cost Recovery Base, wherein the companies sought to recover the costs of maintenance and development, rather than basing their activities on gaining profit. To allow this base to succeed, the government set financial constraints on the companies, as well as performance objectives for the companies to achieve, and allowed for the RWAs to borrow from the government.

These changes in the industry's composition still failed to meet the required expenditures of the RWAs, due to outstanding debts and tight fiscal controls brought about by economic instability in the 1960s and 1970s. Continued discontent in the quality of the services provided in part motivated the 1983 Water

Act, which reduced local government decision-making power, and allowed the RWAs to access private capital markets on top of government borrowing, with the intention of meeting the required investments for adequate water and sewerage services. However, this legislation also ultimately failed to lower the amount of pollution incidents occurring in the industry, prompting a need for further change.

2.1.2 Privatisation

The period within which privatisation was organised and then carried out was from 1986 to 1989, with the principal evaluation of the industry coming from Littlechild (1988).

In the 1980s, the Thatcherite government sought to induce a wave of privatisation across public utilities industries, with some examples prior to the water industry's privatisation being the railway industry, the gas and electric industry, and the telecommunications industry. The principles behind the want for mass privatisation were centred around the notion that private ownership would allow for more efficient production by companies, whose newly-appointed freedom to invest capital would lead to a resolution of the industry's issues which, in context, would mean an improvement in water service quality. These justifications for privatisation of the water industry were exacerbated by the industry's failure to meet the EEC's directives of bathing and river water quality, with the proposed privatisation solution being a method by way the previously under-invested environmental improvements would become targeted investment goals, incentivising the private companies to invest more capital quality-related projects.

Despite the eventual choice to privatise, the government had a selection of choices from which changes to the industry could be performed. Were they to remain nationalised, the government could increase their use of public finance, via increases to taxation and borrowing which would then fund the RWAs. This would have kept in line with many European countries whose water industries persist via nationalisation or via municipalities, as the industry was originally. Other alternatives include financing via water rates, as in Scotland, or Australia's quasi-independent public companies, among others. Ultimately, however, privatisation was chosen given the proposed improvements in efficiency, regulation and quality it would provide, leading to an international industry outlier as the only fully private water industry in the world.

The 1986 Proposals

The initial proposal to privatise the industry was published in 1986, with some inspiration drawn from the recent privatisation of British Telecom and British Gas in 1984 and 1986, respectively, and from the idea that these changes would be a better method with respect to securing efficiency improvements in the industry. Though most reasons to privatise are similar throughout all of industries mentioned, the water and sewerage industry had distinct differences which made the choice yet more appealing:

- The privatisation of the water and sewerage industry would involve 10 RWAs, instead of a single company, as in the cases of British Telecom and British Gas.

- The water and sewerage industry had a particular distinction, in that they also had duties to protect the environment.

- As there were only local and regional monopolies at the time, with no national distribution network, the conditions for Natural Monopolies were more prevalent.

The concerns surrounding Natural Monopoly, in this case, are due to the idea that a natural monopoly is, by design, better if there are less companies in the industry, and therefore it is better to have less competition.

The 1986 proposals sought to give the RWAs private ownership, without a change in their responsibilities - RWAs would still have to provide the necessary water and sewerage services, and be responsible for Flood Control, River Water

Quality, and Control of the Abstraction of Water. The government also believed that the appropriate incentive for the companies would be Profit, instead of government controls, with respect to improving management performance. To protect the interest of the customers, however, the Director General of Water Services was given the responsibility of setting Price Limits and Performance Standards for each company, to prevent overcharging customers for their services or providing poor quality services. The main proposals for economic regulation were the following:

- $RPI-X$ Price Cap Regulation, similar to British Telecom, wherein the growth in prices is limited by the Retail Price Index (RPI), and a compensatory term that accounts for improvements in a company's efficiency. This was chosen due to its relative simplicity, cost effectiveness and ability to preserve efficiency incentives.

- Controls on Quality, as well as prices, since price controls can be undermined by a decrease in the quality of services provided.

- Comparative Competition, wherein the regulator would compare the Costs and Performance Quality of the companies. A system would be developed to measure and assess the performance of the companies with respect to setting their prices.

- Competition in the Capital Market, such that management was sufficiently disciplined and innovation was encouraged. Competition in the Product Market for the industry remained limited, however.

- Furthermore, since some companies were relatively small and vulnerable to takeovers, it was recommended that the government retained a 'Golden Share' in each company, to prevent unwanted takeovers.

With respect to the method behind the price cap regulations, Rate of Return (RoR) Regulation, as an alternative to the $RPI - X$ method, was also considered. RoR regulation defines the price cap by the prices required to cover the costs of production through the company's Revenue, under Perfect Competition, and is generally defined as:

$$RequiredRevenue = (BaseRate \times RoR) \; + \; OperatingExpenses \\ + \; Depreciation \; + \; Taxes \quad (2.1)$$

Though simple in method, one major disadvantage of this method arises as the Averch-Johnson Effect, where Averch & Johnson (1962) found that companies under RoR regulation chose to inflate their use of capital, so that their rate of return would be unnecessarily high, compared to their actual required rate. Such an overuse of capital is called a Capex Bias, and $RPI - X$ regulation was chosen over RoR regulation as a way of avoiding the Averch-Johnson effect via individualised company targets which can facilitate the removal of price cap distortions caused by capital over-investment.

Ultimately, the 1986 proposals were criticised at a fundamental level, where it was widely considered that private monopolies should not have responsibility over both making profits from essential services and a duty to perform an environmental regulatory function. So, six months after publication, the proposals were withheld.

The 1989 Water Act

Subsequently, in 1989, the industry went under the process of privatisation, following the 1989 Water Act. This legislation sought to privatise the water and sewerage industry in much the same fashion as the 1986 proposals, but with the separation of environmental responsibilities from the companies to other, regulatory bodies.

With this Act, the 10 RWAs were publicly quoted as Water and Sewerage Companies (WaSCs), whilst the 29 WOCs were re-established as public limited liability companies. Furthermore, the government chose to cancel all long-term debts owed by the WaSCs, at a cost of £4.9bn in 1989 prices. The WaSCs also received, overall, a cash injection of £1.5bn in 1989 prices known as the 'Green Dowry', and was provided a total further Capital Tax Allowance of £7.7bn, so that companies who were relatively behind in the industry could build up their capital allowances.

The issue of natural monopoly still remained, however, due to the regional ownership and control of the services' networks and a lack of substitutability in the services offered. Furthermore, ownership of the networks causes high barriers to entry within the industry, due to the large costs of distributing water, high capital intensity for the distribution networks, a lack of feasibility of multiple networks, and economic infeasibility due to the low value of the services relative to the costs of infrastructure and distribution.

To mitigate these issues, the 1989 Water Act established the Drinking Water Inspectorate (DWI), National Rivers Authority (NRA) and National Resources Wales (NRW), to manage drinking water quality, pollution and the environment. The Director General of Water Services was also given the responsibility of regulating the industry, and established the Office of Water Services, Ofwat.

As in the 1986 proposals, and with support via a report on regulation in the water and sewerage industry by Littlechild (1988), $RPI-X$ price cap regulation was employed. However, since the services provided by the industry were still of an inadequate quality, water prices had to increase to allow for the companies to appropriately invest. So, $RPI+K$ regulation was instead used, where the K-Factor, $K = Q - X$, is composed of the same efficiency improvement factor, X, and a positive factor, Q, that accounts for any allowed in the price cap for a company to due to the required increases in the quality of their services. This additional price cap factor, though useful for the specific incentivisation of quality improvement, may undo the Averch-Johnson effect abatement caused by the $RPI-X$ measure, as an additional, positive factor within which capital can be invested may lead to an ultimately detrimental over-investment into, say, pollution reduction projects of water improvement projects.

The new regulations also required companies, in an effort to monitor their continued improvements in efficiency, to produce Asset Management Plans (AMPs), which assess the conditions of a company's fixed assets and their expected expenditure levels over a 20-year period. Each WaSCs had to distinguish between operating expenditure and expenditures for maintaining and enhancing infrastructure, whilst meeting the demands of new customers and growth in demand from existing customers. In addition, each company had to produce detailed 10-year financial projections of their revenue, costs and capital expenditure.

Further on the topic of efficiency, a comparative review of the companies' performances was undertaken, which involved an examination of each company's operating costs, with weights assigned to factors regarded to have significant effects on costs, such as Raw Water Quality, Regional Wage Rates, and Population. Each company was then placed into an efficiency band[1], which dictated the scope for improvements in the company's efficiency. This band-designated improvement was added to a further, industry-wide requirement of a 1% efficiency improvement per year, to account for efficiencies associated with the privatisation and improvements in technology (Ofwat (2008)).

2.1.3 Post-Privatisation

Following privatisation, much of the research literature about the water industry began to bloom, with the initial primary question concerning the effectiveness, and any consequences, of privatisation.

[1]Four efficiency bands were proposed by advisors to the Department of Environment in 1989. Consultant teams visited companies to examine operating costs, weighted by significant cost factors as mentioned. The companies were then placed in one of these bands, to indicate the scope for improvements in its efficiency.

The 1994 and 1999 Price Reviews, PR94 and PR99

As aforementioned, following the introduction of the $RPI + K$ price caps, it was found that companies saw excessive profits, to the point of public outcry, suggesting that the complication of the $RPI - X$ solution to over-investment in capital was used, as supposed, to channel funds into short-term capital projects by way of the quality improvement targets. As Saal & Parker (2001) find, there were in this period increases in outputs prices that far outweighed increases in input costs, which led to supernormal profits post-privatisation.

Citing Ofwat (1994), the PR94 assessment proposed that the K-Factor should, overall, remain reasonably high for the next five years, until the previously neglected environmental obligations and excessive profiteering were accounted for. Ofwat supposed, that the weighted average K-Factor in the forthcoming years would first change by 1.4%, until 2001 where it drops to 0.4%, yielding a ten-year industry weighted average K-Factor of 0.9%.

Table 2.1: Average Annual Price Cap[1] Comparisons: 1989, 1994, 1999

Company	1989 Average Annual Limit (%) (min, max)	1994 Average Annual Limit (%) (min, max)	1999 Average Annual Limit (%) (min, max)
WaSCs	3.9	1.5	-2.0
	$(0.0, 5.5)$	$(0.5, 4.0)$	$(-4.6, -0.5)$
WOCs	1.9	0.6	-2.8
	$(-1.6, 6.0)$	$(-2.0, 2.5)$	$(-5.1, 2.0)$
Industry Average	3.7	1.4	-2.1

1: Price Cap limits are averaged over the five-year review period following the year of Price Review publication.

Looking at Table 2.1, which compares the initial price-cap determinations following privatisation with that of PR94, it can be seen that there was significant belief in lowering the maximal annual limits, under the justifiable assumption that the industry would, on average, commit to its environmental obligations and better utilise its expenditures for service improvements. Notably, where the WaSCs all saw a non-negative determined limit in both 1989 and 1994, multiple WOCs were given negative limits, therefore requiring price decreases over the review period, owing to the need for large reductions in Base Operating Costs and Returns on Existing Assets, as accounted for in the X component of the K used in the determination.

Extremities aside, the price limits were said to be in deceleration, owing to the progress in fulfilling statutory environmental requirements, and a regulatory tightening on efficiency improvement for each company. According to Table 3 of Ofwat (1994), most of the quality improvements were determined to be within the Sewerage services, leading to large overall Q-components in the industry's K Factors. The average industry X-components were relatively similar across both sides of the industry, though slight relief was given to reductions in base costs to the sewerage services, alongside slightly more compensation in accounting for Growth, Levels of Service and Capital Maintenance.

The following price review, PR99, saw for the first time an overall decrease in the prices of services provided to customers. Now known to be a more severe determination than its predecessors and following reviews, Ofwat (1999) sought to introduce a drastic immediate decrease in the price limits for all companies, with a determined industry average of -12.3% for the 2000-01 period, denoted by P_0 in the document, in order to strongly promote a reduction in the operating costs of each companies and cause a significant improvement in customer bill sizes and cost efficiency.

Shown in Table 2.1, the immediate large decrease in price limits led to negative price limits over the PR99 review period, with few exceptions for WOCs arising due to lower relative P_0 decreases, and higher sustained annual increases thereafter, captured by Table 1 of Ofwat (1999). In terms of qualifying components for the determined K factors, little was changed from PR94. Rather, it was the severity of the efforts required by the companies, as determined by Ofwat, that led to the negative average K factors in this period.

Amongst the first of the post-privatisation papers are Lynk (1993) and Hunt & Lynk (1995), who look at the effects of privatisation on industry efficiency. Lynk (1993) uses an empirical model of costs to find the impacts of joint production on costs, for pre- and post-privatisation private companies. It is found that, on average, those companies that were already private prior to industry-wide privatisation were more inefficient, and that there is strong evidence of improvements in productivity in the RWAs over time, which may be lost post-privatisation due to the separation of joint production, such that the environmental duties were taken by the NRA.

Hunt & Lynk (1995) use similar empirical models, and look at long-run elasticities of joint production upon costs for the 1980's prior to privatisation, and find a loss in efficiency-enhancing effects through economies of scope, reasoned to have been lost due to the failure to self-regulate under public ownership.

Ashton (2000, 2003) estimate cost models for primarily the post-privatisation period, with Ashton (2000) including some years pre-privatisation, and Ashton (2003) instead choosing years around PR94. Ashton (2000) also assumes a composite error with the usual white noise term and an individual inefficiency term, and find that there is an 84% overall average cost efficiency, with a range of 23% and standard deviation of 8%, which all suggest a both a moderate variation in efficiency across companies, and a moderate scope to reduce operating costs.

Ashton (2003) instead focuses on a Variable Cost function, and finding the economies of scale, economies of capital utilisation, and the level of capital utilisation over the time period, finding significant and consistent positive diseconomies of both scale and capital utilisation, and that, though increasing over time, the levels of capital utilisation in the time period were 30% on average, which is remarked as relatively low for utilities at the time.

Bottasso & Conti (2003) use a operational expenditure model with composite errors as in Ashton (2000), looking specifically at the post-PR94 period, where yardstick competition was introduced into the industry. They found a steady decrease in average cost inefficiency over the 1994 - 2001 period, to the point that they recommend technical improvements, rather than cost reduction, as the main way to improve efficiency thereafter.

Finally, and most relevant to the more recent regulations and the interest of quality in this thesis, Saal & Parker (2000) introduce in their cost model the notion of quality adjustment indices to their water and sewerage outputs, defined as, respectively:

$$
\begin{aligned}
Q_W &= \overline{\%CompliantZones} : \overline{\%CompliantZones}_{1990} \\
Q_S &= \left(\frac{S_R}{S_R + S_B}\right) R_Q + \left(\frac{S_B}{S_R + S_B}\right) B_Q
\end{aligned}
\tag{2.2}
$$

Where Water Quality, Q_W, is the ratio of the average percentage of complaint zones of each DMU to water regulations, to the average percent compliance in a base year of 1990. Wastewater Quality, Q_S, is defined by shares of company water defined as river water or bathing water, S_R and S_B respectively, which weight indices of river and bathing water quality, R_Q and B_Q respectively. Importantly, as well as confirming the earlier findings of Hunt & Lynk (1995), they find that the estimated joint production parameters change sign between a quality-unadjusted and quality-adjusted cost model, implying the existence of 'quality-driven scope economies', where quality improvement of one output might decrease the cost of another.

The 2004 and 2009 Price Reviews, PR04 and PR09

The 2004 Price Review saw the incidence of customer research, jointly commissioned by Ofwat and stakeholders in the industry, so that the interests of customers and the wider environment can be best accounted for. As shown in Table 2.2, this review also saw an average industry price limit over the review period that was positive. The change in regulatory behaviour, compared to the previous review which saw drastic negative annual price caps, was cited from Ofwat (2004) as a substantial financial requirement for a capital programme designed to both counter high company costs and further environmental and service improvements.

Table 2.2: Average Annual Price Cap[1] Comparisons: 2004, 2009

Company	2004 Average Annual Limit (%) (min, max)	2009 Average Annual Limit (%) (min, max)
WaSCs	4.3	0.5
	$(2.4, 6.9)$	$(-0.8, 1.9)$
WOCs	3.1	0.3
	$(-0.5, 4.8)$	$(-2.1, 1.8)$
Industry Average	4.2	0.5

1: Price Cap limits are averaged over the five-year review period following the year of Price Review publication.

PR09, the 2009 Review, also saw increasing price limits on average, with some companies facing a 5-year average decrease in their annual limit. The reasoning behind this review's final determinations, according to Ofwat (2009), include re-assessments of the relative efficiency of operating expenditures between companies, and the need to ensure strong and persistent incentives to improve efficiency.

Fundamentally, the calculation of the K factors for the price limits have remained similar for the pre-2000 and post-2000 reviews. Such limits are calculated, in these more recent reviews, by considering the Output Requirements of a company, which includes Operating Expenditure, Capital Expenditure, Capital Returns and Taxes, and adjusts the Required Revenue for each company based on the achievement of expected performance targets and efficiency improvements, giving yielding the annual price limits as a result.

At this time, research was expanding into retroactive models of efficiency, factor growth, and the effects of the regulations on the industry in hindsight. Saal & Reid (2005) estimate Opex productivity growth from 1993 to 2003, and by using a variable cost model with quasi-fixed capital, find that opex productivity growth declines over the sample time period, from 2.15% on average to 1.44%. After correcting for bias in the measurements of an average industry DMU, the rates of decline alter to 2.02% and 1.76%, respectively. Additionally, they find that the net impact on productivity growth from 1995 to 2003 was 0.577%, reflecting a net improvement over time, which offsets the declining growth rates of the period.

Saal & Parker (2005) explore the use of input distance functions to measure overall performance in the industry, finding that productivity growth can be decomposed into technical change, efficiency change, and scale change. They also find that, owing to their earlier findings in Saal & Parker (2000), WOCs and WaSCs have different production frontiers, likely because of the non-separability of WaSC water and sewerage operations. Saal et. al. (2007) continue this area of work by determining the contributions of each of the three productivity components to productivity growth in the 1985 - 2000 time period, looking at the effects of privatisation on these new facets of productivity growth. They find that privatisation has mixed results: WaSCs are defined by decreasing returns-to-scale, as Saal & Parker (2000) found, and additionally find that this behaviour has consistently negatively impacted industry productivity growth before and after privatisation; on the other hand, technical change, which measures the change in productivity strictly due to changes in industry technology, has increased post-privatisation, suggesting that the change in industry structure, given its premise to resolve environmental damages, may have led to beneficial environmental regulation.

Ultimately, however, they further find limited evidence of efficiency improvement in their model. Bottasso & Conti (2009a), looking at measures of scale economies, technology and technical change in only the water sector of the industry, find small but significant improvements in companies' economies of scale at all company sizes, and that there exist improvements measured by technical change in the 1995 - 2005 period.

Looking further at the effects of regulation on efficiency, Erbetta & Cave (2007) address the impact of the tightening price caps of efficiency. Concluding with remarks on policy effectiveness and changes in efficiency over the sample period of 1993 - 2005, they find that, though PR94 had a continued limited productivity improvement, PR99 showed significant technical efficiency improvements, corroborating Bottasso & Conti (2009a)'s findings; looking at allocative efficiency, and the specific distortions of inputs, they also found that allocative efficiency improved over time, with a reduction of over-utilised labour over time. Finally, a trend of continuous managerial efficiency improvements over time is found, for both technical and allocative efficiency.

Further on the effects of regulation at the time, Bottasso & Conti (2009b) look at a 'ratchet effect' in the industry, where it is found that companies tend to increase cost-cutting activities early in the regulatory cycle, and weakens as the next price review approaches. Using a model of variable cost for the 1995 - 2004 period, they find evidence of this cyclical pattern, which suggests that incentives relating to efforts in cost-cutting are biased, towards the start of the period. Therein, they also find that, compared to the PR94 period, the PR99 period has seen significantly lower technical change according to their model, implying that the strict price cap regulations given to companies did not lead to the desired performance improvements, although this notion could partially arise due to most of the scope of performance improvements from this kind of regulation having already been achieved - a more radical change to industry structure might, according to these conclusions, led to more of the desired improvements over time.

Research in the PR09 period continued the fashions of the period prior, with ramping interest in indices that both measure industry performance, and can be decomposed into detailed, explanatory factors. Portela et. al. (2011) find, using a meta-index model that allows for the incorporation of all industry companies into the analysis for the 1993 - to 2007 period, that there are increasing cost efficiencies over the time period, where the causes of which can be partly explained by the different regulatory periods: tight price cap regulations in PR94 and PR99 appeared to spur on greater cost efficiency improvements at an industry level, and increases technology gaps, which suggest that the frontier companies for each group of companies in the meta-frontier analysis are also improving in reaction to these regulations.

More pertinent to the discussion of quality, which has been used, incidentally, in much of the literature after the introduction of output-adjusting quality measurements by Saal & Parker (2000), is the measurement of profit change and its components by Maziotis et. al. (2014), for models with and without output quality. They find a general negative profit change over the 1991 - 2008 period, with some periods such as the PR99 providing major downward trends, and other regulatory periods such as PR94 providing some positive change. As it relates to output quality, they also find that, compared to quality-unadjusted results, greater emphasis is placed on effects due to reorganisation of activities, rather than effects due to productivity, in the quality-adjusted components of the profit change.

One point of note throughout the regulatory periods post-privatisation is the collection of other factors in regulation that were affected, and how these other targets were influenced by the price caps detailed above. One crucial factor for almost all of the companies operating in the industry is profitability, a factor that was believed to be the cause of some of the negligence towards quality. Taking price caps into account, firms had to contend with regulatory pressures on their profits, requiring efficiency and quality improvements that had to come from cost reduction. This too affects another important regulatory facet, that of capital investment. To account for stricter price caps, for example, firms reduce their total costs, which may include lowering capital investments which could reduce their efforts and investments into improving quality. All of these issues put together imply a degree of tension between companies and their price caps, and therefore Ofwat, which may have caused insufficient quality improvements consequently.

2.2 Modern Regulations in the Industry

As of writing, the industry is currently in the PR19 regulatory period, but in this section on the most recent regulations as it pertains to quality, the previous price review - PR14 - is of interest as well.

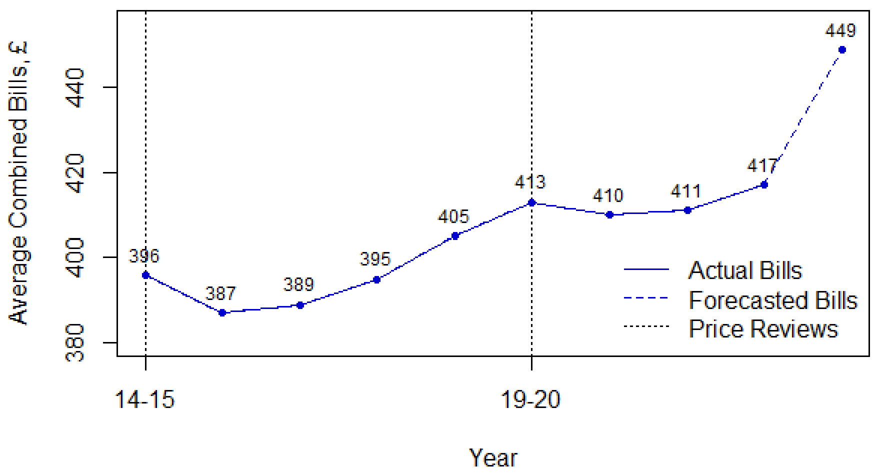
The 2014 and 2019 Price Reviews, PR14 and PR19

The most recent reviews have seen significant expansions in the determination of the price controls for each company, with a primary focus, as was alluded to in PR04, to better consider the customers of the companies in setting both price limits and efficiency incentives. To that end, PR14 introduced Outcome Delivery Incentives (ODIs), whish were designed to create improvement targets for each company that provided a financial reward for sufficient success in the target, and a penalty otherwise. Not all ODIs had rewards, and were instead to be reached so as to avoid financial penalty. Other substantial changes in regulation also occurred in PR14, such as Ofwat's change in definition on costs from Totex, total expenditures, to Botex, base total expenditures, which are less any enhancement costs. This change is an interesting empirical distinction, but has the drawback of making any previous cost assessments, using totex, difficult to compare to newer estimates. Other changes at the time included the change in regulatory focus, from one where regulations were to best regulate the industry's natural monopolies, to a paradigm where regulation is instead designed to promote competition.

The introduction of quality measures was not new to PR14, given for example the quality-improvement factor in the price caps for companies, but provided clearer incentivised targets that could be overseen and planned around by Ofwat. Previous targets were set by the director general, and included, amongst others, the targets of river and bathing water quality that were failed pre-privatisation. An interesting facet of this shift to ODIs is the change in incentives for companies. ODIs offer financial rewards, or more typically financial penalties upon failure, which directly impact company revenue. This could alter company behaviour to one which focuses on targets for profit's sake, rather than for the sake of quality improvement; on the other hand, there could be difficulties for the companies if the regulator cannot accurately determine the capacity of companies to meet these ODIs, or if they do not incentivise the targets properly - for instance, if a quality target costs more to meet than the rewarded it yields, will companies be likely to commit to that facet of quality?

A subset of these ODIs were Common Performance Commitments (CPCs), which were to be achieved by all WaSCs or all companies in the industry, depending on the ability to achieve sewerage targets. Such targets include, for example, reducing the number of Mains Bursts in a company's jurisdiction or reducing the Average Consumption per capita of water in a household, but also encompassed more customer-oriented goals, such as increasing customer satisfaction by reducing waiting times for provided support services.

Figure 2.1: Real Average Combined Bills, 2014/15 to 2023/24



As for the Price Caps in these periods, PR14 marked the time where the caps previously displayed in Tables 2.1 and 2.2 were no longer monitored - instead, price caps were split into various components over different facets of the industry. However, an analogous way to observe at least the industry average change in prices over the regulatory periods can be seen in Figure 2.1, which plots the Average Combined Bills of customers in both PR14 and PR19, with the last year of PR19 being a forecast at the time of writing. PR14 sees an overall bill increase of 2.27%, which is a moderate average out of all price review periods thus far. PR19, on the other hand, is forecast an increase overall of 8.72% - more than twice the increase of the next highest price cap of 4.2% on average over PR04.

The sudden increase in total bills are, according to a recent BBC article (Espiner (2024)), due to a supposed commitment by the industry to improve quality throughout the industry, such as through reductions in pollution incidents and the reduction of leakages via pipe repairs. However, as the article notes, such price increases are not only considered insufficient by water companies for these quality improvements, but are also in the wake of bonuses given to leading figures in the companies. The supposed increases in bills are restricted by Ofwat to a national rise of 21% by 2030, and is two thirds the increase demanded by water companies. Though this price increase could reflect an eventual interest in remedying the current issues in quality in the industry, it more cynically appears to arise from increases in bonuses, given the current quality issues. Atkins & Peirce (2024)'s case study on the river Wye highlights the contemporary issues of the water industry.

At this time in the research literature, the attitudes and methods surrounding quality measurement began to change as well. The most notable paper at the time, in my opinion, is the report by Saal et. al. (2017), who make the strong conclusion that the current measures of water and wastewater quality - referring to the Saal & Parker (2000) indices - are now 'stagnant', in that they do not allow for companies to improve their quality because all companies have achieved almost full compliance in these measures. To that end, the report calls for, and explores somewhat, the scope for newer measures using new or alternative measurements taken from industry data.

Research at the time began to reflect the changing opinions on quality as well. Applications of various factors that would becomes CPCs became more present, such as the use of Leakage (Mocholi-Arce et. al. (2020, 2021), Molinos-Senante & Maziotis (2019, 2020)), Mains Bursts (Mocholi-Arce (2019), Molinos-Senante & Maziotis (2018, 2019)), and Sewer Collapses (Molinos-Senante & Maziotis (2018)). Service quality aspects, such as in Maziotis et. al. (2014) also became more utilised as economic Bads: facets of production that firms seek to minimise in their production process - Molinos-Senante et. al. (2015b, 2016, 2017b) are all examples here, as well as Brea-Solis et. al. (2017) which covers water losses.

2.3 Conclusion

This chapter hopes to demonstrate how the English and Welsh Water and Sewerage industry has developed over its history, from local, municipal management to the current era of regulatory cycles for the few natural monopolies that remain. The chapter has also made the point of focusing on the perception of Quality in the industry, highlighting not only the historical shortfalls in quality that led to great industry change, but also the recent attempts to demonstrate the efficacy of regulatory outcomes on quality as of late.

Quality has always, it seems, been a principal cause of industry change, from municipal ownership to nationalisation, and from nationalisation to privatisation in 1989. Saal et. al. (2017)'s recent report on 'stagnation' in quality measurement should then highlight that the industry is need of another change, by way of re-designing how quality should be measured across all aspects of the industry, lest the industry face another axiomatic change in its structure from once-again ignored issues pertaining to quality.

In identifying this trend, research contemporary to each price review also hoped to add more depth to the performance evaluation of the price reviews and privatisation more generally, again focusing on quality issues when relevant. The literature supports the same argument that the thesis intends to uphold: that quality, in some fashion, does cause changes in the empirical results of research when accounted for, and therefore should be a concern in any industry model. By extension, the need for updated and more useful quality measures should also be a primary concern for the field, as stagnant quality measures will betray issues not otherwise reflected by the empirical outcomes.

This theme of exploring how measures of quality matter to empirical models will be a focal point of the contributing chapters later in the thesis - Chapters 5 and 6 in particular. To complement this, the following chapter, Chapter 3, will look at the research literature from a modeling perspective, covering the theory behind the models, and what methods are popular in the research, to best provide an insight into what model might be best appropriate for assessing industry quality.

Chapter 3

Prerequisite Theory

The previous chapter covered much of the literature that this thesis is built upon, and in doing so also outlined some of the methods used by researchers to analyse the water and sewerage industry. In building up the required prior knowledge for the contributing chapters, it is reasonable to also delve further into why such models were used in the preceding research, and in covering these topics also delving into other methodologies that could be employed in projects similar to this thesis and those previously published papers.

So, this chapter will cover the necessary theory that underpins the methods used in the literature. In particular, this chapter will first cover Producer Theory, which is the economics of firms producing their products subject to various constraints and assumptions. In addressing this area of economics, the idea of Cost Functions and their estimation can be covered, which relates to some of the aforementioned models in the literature.

Acknowledgement of this theory then allows for the coverage of what this thesis will primarily use as its models - Non-Parametric Models, and in particular Data Envelopment Analysis, which is also used in previous literature as well as in here. In looking at this type of method, other ways of analysing efficiency will also be looked at as both a comparatison of methodologies, and for the sake of completeness in the range of plausible methods available to this area of research. Such methods include not only other non-parametric models, such as Multidimensional Efficiency Analysis (MEA) and Range-Directional Measures (RDM), but also Parametric Methods such as Stochastic Frontier Analysis (SFA) and other areas such as Semi-Parametric models that lie somewhere in between.

The aim of this chapter is to prepare an understanding of why certain models are used, with the theory behind them as evidence, while also evaluating the suite of available models to this thesis, and the justifications behind why DEA, in particular, is used as the basis for the models in the following contributing chapters.

3.1 Producer Theory

If, in microeconomics, Consumer Theory is the study of how individuals choose what and in what quantities to consume goods and services subject to preferences and various constraints, then Producer Theory is analogous for firms who seek to produce those goods and services. Producer Theory is therefore the study of how firms, treated as individual entities, choose what, and in what quantities, to produce goods and services that they are able to feasibly produce, subject to constraints such as Cost Budgets and the like.

Broadly speaking, the problems in Producer Theory are looked at from two perspectives: Profit Maximising Problems (PMP) and Cost Minimising Problems (CMP), which happen to produce the same results in theory but through different information. To explain how these problems are addressed, much of Chapter 5 of Mas-Colell et. al. (1995) and the first six chapters of Varian (2009) are used here, as they serve as good fundamental textbooks for post-graduate microeconomics.

3.1.1 Production Sets

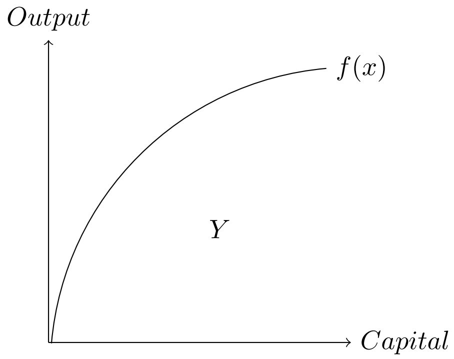
In general, the primary aspects of producer theory can be represented by mathematical sets, the first of which is the Production Possibilities Set, denoted by

Y. This set contains the vectors $(\mathbf{y}, -\mathbf{x})$, which consists of s Outputs $\mathbf{y} \in \mathbb{R}^s$ and m Inputs $\mathbf{x} \in \mathbb{R}^m$, which are considered to be negative quantities. In this representation, the set Y contains Netputs, which are based on the outputs produced less the inputs required for production:

$$Y = \{(\mathbf{y}, -\mathbf{x}) \in \mathbb{R}^{s+m} \mid y \leqslant f(\mathbf{x})\} \tag{3.1}$$

Equation 3.1 assumes that the set of netputs is such that the outputs $\mathbf{y}$ are not larger than the set of inputs after they have been transformed by the Production Function, $f(\mathbf{x})$, which describes how the inputs are used to produce the outputs. More formally, f is a mapping of the inputs onto the outputs:

$$f(\mathbf{x}) = f : \mathbb{R}^m \mapsto \mathbb{R}^s = \{\mathbf{y} \in \mathbb{R}^s \mid \mathbf{y} \text{ is maximal for -}\mathbf{x} \text{ in } Y\} \tag{3.2}$$



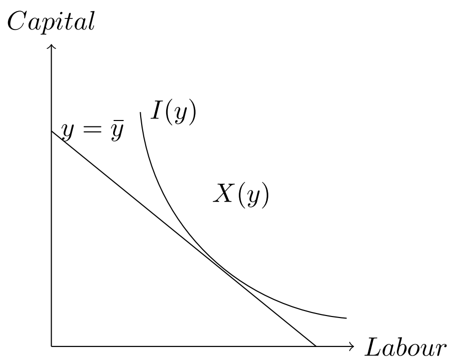
For the remainder of this part of the chapter, it will be assumed for simplicity that $s = 1$ - that there is only one output being produced, which allows for easier examples and illustrations later on.

In any case, the function f requires that the output y is the maximal amount for the inputs $-\mathbf{x}$, given that the combination is in the set Y. and is therefore producible. Further to this need for a maximal output, other sets can be defined for the inputs. The Inputs Requirement Set, $X(y)$, is the set of inputs that are used to at most produce their respective output:

$$X(y) = \{\mathbf{x} \in \mathbb{R}^m_+ \mid y \leqslant f(\mathbf{x})\} \tag{3.3}$$

In the optimal case of production, the output's value exactly equals the value produced from the inputs through the production function. A stricter set, called the Isoquant $I(y)$, can be defined as the set that only contains the inputs that equal the output y after transformation through f:

$$I(y) = \{\mathbf{x} \in \mathbb{R}^m_+ \mid y = f(\mathbf{x})\} \tag{3.4}$$



A small thing to note in both equations 3.3 and 3.4 is that the inputs are required to be non-negative, which is denoted by the stricter set of real numbers that they belong to.

A final useful set is the Transformation Function $T(y, \mathbf{x})$, which is defined as the function that chooses the netputs which maximise production:

$$T(y, \mathbf{x}) = y - f(\mathbf{x}) \leqslant 0 \tag{3.5}$$

Where production is considered Efficient if $T(y, \mathbf{x}) = 0$.

3.1.2 Production Set Properties

The production sets defined above are all useful for defining the requirements of the production process. However, additional properties are required of the inputs and outputs to ensure that they create well-behaved sets that allow for production to take place. These assumptions are:

1. Y is Non-Empty and Closed

2. There is 'No Free Lunch'; there is the Possibility of Inaction; there is Free Disposability

3. Production is Irreversible

4. Y is Additive, which allows for Free Entry

5. Y is a Convex Cone

6. Y can have Non-Increasing, Non-Decreasing, or Constant Returns-to-Scale

Assumption 1 ensures that there is some amount of the output produced, and that the amount producible has some definite maximum that is also producible. Assumption 2 states that, in order to produce something, some amount of inputs must be used - there is no 'free lunch', as it were. This assumption also states that producing nothing, and therefore being inactive, is a valid choice, and that firms can produce less of their output than is maximal with the same amount of inputs: more succinctly, if $y \in Y$ and $y' \leqslant y$, then $y' \in Y$.

Assumption 3 refers to the idea that, once some inputs have been used to produce some output, the production process cannot be undone - the act of producing is irreversible. Assumption 4 states that, for two production outputs y and y' in the set of producible outputs Y, their sum is also producible, $y+y' \in Y$. This extends to the idea of Free Entry into the production process - if y and y' are produced by two separate firms, then their sum being admissible implies that both firms are free to enter and produce.

Assumption 5 is a fundamental repeating property in microeconomics, whose definition is not too far from the previous Additivity assumption. For two outputs $y, y' \in Y$, Convexity states that some weighted average of the two must also be producible, $\alpha y + (1-\alpha)y' \in Y$, for all $\alpha \in [0,1]$. Furthermore, the assumption that production is a Convex Cone suggests that, for any $\alpha, \beta \geqslant 0$, $\alpha y + \beta y' \in Y$ - which is a more general version of additivity.

Assumption 6 pertains most practically to the forthcoming models. This assumption addresses the scaling of production:

- A production exhibits Non-Increasing Returns-to-Scale (NIRS) if, for $y \in Y$, $\alpha y \in Y$ for all $\alpha\,[0,1]$.

- A production exhibits Non-Decreasing Returns-to-Scale (NDRS) if $\alpha y \in Y$ for all $\alpha \geqslant 1$.

- A production exhibits Constant Returns-to-Scale (CRS) if $\alpha y \in Y$ for all $\alpha \geqslant 0$.

In effect, CRS combines the previous two assumptions: if scaling a production process yields an output that is proportionally no less than before (NDRS) and also proportionally no more than before it was scaled up (NIRS), then the results must be exactly equal to the previous production proportionally. As will be discussed later in the chapter, the reference of which returns-to-scale to use is important empirically, as it helps determine what the production frontier looks like, and therefore how it is calculated in practice.

3.1.3 Optimality Conditions

The aim of producer theory is to determine the optimal way in which companies produce their goods and services given their inputs. There are two approaches to solving the optimality problem - via a Profit Maximisation Problem (PMP), or by a Cost Minimisation Problem (CMP), which mathematically will yield the same answer through different information. For this thesis, we are more concerned with the Costs of firms, and how they can be best reduced, but both ways of solving will be addressed regardless.

Profit Maximisation

Arguably the more typically associated goal to firms and their shareholders is that of maximising Profit. Mathematically speaking, the problem can be written as:

$$\max_{y} p \cdot y,\ s.t.\ y \in Y \tag{3.6}$$

That is, firms try to produce as much as they can, with some Prices p, subject to that amount produced being actually feasible in the first place. When a firm only has one output, as has been assumed in this chapter, we can also write this problem with reference to its production function:

$$\max_{x>0} pf(x) - w \cdot x \tag{3.7}$$

where, for prices p and Factor Prices w, the firm maximises the difference between the outputted value - its Revenue - and the Costs to produce that output.

Cost Minimisation

Analogous to the PMP is cost minimising, which instead solves:

$$\min_{x\geqslant 0} w \cdot x,\ s.t.\ f(x) \geqslant \bar{y} \tag{3.8}$$

For factor price w, the firm minimises the value of the inputs, x, used in their production, such that they still produce at least some fixed level of output $\bar{y}$ through their production function f.

Quasi-Fixed Capital

One final issue not covered in the above theory, but pertinent to Chapter 6, is the notion of Quasi-Fixed Capital. All of the above analysis, and the resulting optimisation problems, have assumed that all factors in production are variable - they can, in the long run, be adjusted to their optimal values, as to maximise profits or minimise costs. However, this isn't necessarily the case in practice.

Consider, for example, the infrastructure of the water industry - its piping for instance, which lies underground and across all a companies catchment area. Suppose that the firm determines that they would be most optimal if those pipes were replaced and updated, as has been the case in the industry, which has seen a long-term effort to replace legacy pipes. The pragmatic question to ask this company is a simple one - how long will this take?

When we think of the 'long-run' economically, we tend to think of any time period greater than a year. Compare this stretch of time to the Herculean task above - can all of the piping be replaced within a year, for it to be considered an 'immediate' adjustment of the infrastructural capital? The answer, according to the industry in practice, is a steadfast "No". Indeed, in practice, though it could be reasoned that, eventually, all capital adjustments however large can be made so that a firm reached is optimal production, in the far more practical short-run, this is not always the case. Letting a producer's inputs be Capital k and Labour l, the short-run production function, with these unfinished capital adjustments, takes the following form:

$$y = f(l, k, \bar{k}) \tag{3.9}$$

Where k is the variable component of capital, and $\bar{k}$ is the fixed component. Thus, capital is defined as 'Quasi-Fixed', which means that, though in the short-run the capital must take fixed, unadjusted values, it becomes ultimately variable in the long-run. This creates interesting dynamics in a model of production, where the estimated costs of production require the consideration of capital which cannot be reduced in shorter periods of time, thereby potentially relatively higher costs in the short-run, owing to the need to invest into fixed amounts of certain inputs in production.

Notably, this change in definition of capital is not the same as incorporating new capital into the production process. Instead, the same total amount of capital employed in production is defined by the rate at which it is capable of being varied. One could similarly, for example, define labour as a total of short-term contracts and long-term employees - the measurable amount of labour in production is the same, but the differences in variability have been explicitly defined to better assess the behavioural differences in the production process.

3.2 Efficiency Analysis

To explore the models of the research literature whose results were cited in Chapter 2, the transition from the above theory to empirical models should first be covered. To first lead into the concept of Efficiency, the following question can be considered: To what extent, in reality, is it likely that every producer is perfectly efficient in producing their outputs given their inputs? The answer, bluntly, is that it is not at all likely.

So, there must then be some form of inefficiency, or equivalently some level of efficiency below a fully-efficient benchmark. The point of interest then becomes one of measuring such a value for companies in an industry. This further convolutes with the addition of another practical concern: To what extent, in reality, is the optimal production function known by any producer? The answer, bluntly, again, is that it is not at all likely.

3.2.1 Initial Efficiency Models

Before covering the initial methods by which these two questions were answered, it is worth distinguishing the different types of efficiency that are of interest in the research literature, with further definitions of other efficiencies to be defined as they arise. The 'core' definitions, as it were, are:

Technical Efficiency: A producer is Technically Efficient if they feasibly produce the maximum amount of Outputs possible with the minimum amount of Inputs possible under a given Technology.

Cost Efficiency: A producer is Cost Efficient if they feasibly produce the maximum amount of Outputs possible with the lowest Total Costs possible.

Allocative Efficiency: A producer is Allocatively Efficient if they feasibly produce the maximum amount of Outputs possible, such that increasing the production of one output must require decreasing the production of another.

Technical Efficiency can be broken down into smaller sub-categories: Input- and Output-Oriented Technical Efficiency, where the input-oriented efficiency requires that the producer produces a given level of their feasible outputs with minimal inputs given a production technology, and vice versa for output-oriented efficiency.

The initial definitions for Efficiency come from Debreu (1951)'s Coefficient or Resource Utilisation, which is analogous to Technical Efficiency, and Koopmans (1951)'s Efficient Points, which is analogous to Allocative Efficiency. First, Debreu (1951) defines a Coefficient of Resource Utilisation out of a general equilibrium model containing three resources: Physical Resources, $\mathbf{z}^0$, Production Possibilities, $\mathcal{Y}$, and Economic Organisation Possibilities, with the objective of achieving a given level of consumer Satisfaction, $\mathbf{s}^0$. The coefficient, ρ, is then defined as:

$$\rho\left(\mathbf{s}^0, \mathbf{z}^0, \mathcal{Y}\right) := \max_{\mathbf{z}\in\mathcal{Z}^{min}(\mathbf{s}^0)} \left(\frac{\mathbf{pz}}{\mathbf{pz}^0}\right) \tag{3.10}$$

Where ρ is the ratio of the value of the utilised physical resources, $\mathbf{z}$ to the value of the total available resources $\mathbf{z}^0$. In this model, the economic loss arising from technical efficiency is then defined, for an optimal vector $\mathbf{z}^*$:

$$TE(\mathbf{z}^0, \mathbf{z}^*, \rho) = \mathbf{z}^0 - \mathbf{z}^*, \ = \mathbf{z}^0(1-\rho) \tag{3.11}$$

Koopmans (1951) instead focuses on a Commodity Space, and uses Activity Analysis over General Equilibrium modelling as in Debreu (1951). A point in this space is considered (Allocatively) efficient whenever an increase in one of its dimensions, which is the value of a good's net output, can only be achieved at the cost of a decrease in another dimension - another good's net output. Mathematically, this is defined as:

$$AE : \mathbf{y} \in (A), \ \nexists \bar{\mathbf{y}} \in (A) \ s.t. \ \bar{\mathbf{y}} - \mathbf{y} \geqslant 0 \tag{3.12}$$

That is, a point, $\mathbf{y}$ is allocatively efficient if it is producible, and therefore within the cone available combinations of technologies (A), and there also does not exist any other point $\bar{\mathbf{y}}$ that is both producible within (A) and greater than $\mathbf{y}$.

One point of note about these definitions is the difference between their declaration in theory, and their application. As was mentioned earlier, in practice efficiency measures can be input- or output-oriented, which refers to the way in which the production process is optimised: input orientation implies a minimisation of inputs given a fixed level of outputs, as in cost minimisation, whereas output orientation follows profit maximisation, wherein outputs are maximised such that a certain level of inputs are used. In theory, analogous to the primal-dual relationship between the pair of optimisation problems, both orientations should be, by some measure, equivalent. This also holds in practice: Letting θ

be an input-oriented efficiency, and φ an output-oriented efficiency, it is found that:

$$\theta = \frac{1}{\varphi}$$

3.2.2 Cost Estimation

Equation 3.10 demonstrates the simplest way by which efficiency can be measured - the ratio of actual results to their optimal values. A firm would be optimal if the data-based actual results of, say, their costs were equal to the mathematically minimal costs calculated separately. Efficiency - in this case, Cost Efficiency - would then be measured by any gap between the two values:

$$CE = \frac{C^*(w,x)}{\hat{C}(w,x)} \tag{3.13}$$

One the questions that started this section arises again when looking at this measure - To what extent is the (optimal) production function known? The answer, as before, is that it is precisely unknown. However, what can be done is the estimation of some form of cost structure which, if backed sufficiently by theory, could be close enough to the true production function to be sufficient.

Where Cost Functions C are either optimal, C^*, or estimated, $\hat{C}$, and depend on factor prices w and inputs x, as defined in the aforementioned theory section. There are a suite of cost functions that could be applied, and a selection of them will be addressed here.

Linear Cost Functions

The first type of cost function are those that take a linear form - it is therefore assumed that there is only a linear relationship between costs, the production function, and its inputs and input prices. An example of this type of model would be:

$$C_i = \beta_0 + \alpha y_i + \beta_1 x_{1i} + \cdots + \delta_1 w_{1i} + \cdots + \varepsilon_i \tag{3.14}$$

This model has the advantage of being perhaps the most simple expression of costs, but assumes by definition that all inputs and their prices are additively separable, and that they and the resultant output neatly and linearly affect overall costs. In practice, this is unlikely. To remedy this, a Cobb-Douglas model can be used, and then log-linearised to yield a similarly easy-to-estimate functional form, while also possessing a more widely-accepted functional relationship between inputs:

$$\begin{aligned} C_i &= A y_i^{\alpha} x_{1i}^{\beta_1} \cdot \dots \cdot w_{1i}^{\delta_1} \cdot \dots \cdot \varepsilon_i, \\ \ln(C_i) &= \beta_0 + \alpha \ln(y_i) + \beta_1 \ln(x_{1i}) + \dots + \delta_1 \ln(w_{1i}) + \dots + \varepsilon_i \end{aligned} \tag{3.15}$$

Where A reflects some general state of technology, with $\ln(A) = \beta_0$. As popular as this form is in general, it too suffers from the assumed separability of its inputs and prices, and ultimately that all of the variables, once linearised, have strictly linear effects on costs.

Quadratic Cost Functions

The consequence of linear relationships impacting costs is more noticeable when addressing the marginal effects of the models above. Suppose we look at the marginal effect of increasing output on costs in equation 3.15:

$$\frac{\partial \ln(C_i)}{\partial \ln(y_i)} = \alpha \tag{3.16}$$

The effect, for all levels of output, is constant and takes the value of the parameter α, which can be estimated and interpreted as a percentage point change. Analysing this, is it realistic to assume that, be it the first unit of production or the thousandth, that the impacts on costs are the same?

The answer in practice is that, in all likelihood, these costs are different - it is known that, as the amount of production increases, Average Costs will decline up to a point, and may increase past that critical point if the company is experiencing diseconomies of scale in their production.

The simplest way to address this problem, for example with output, is to introduce a quadratic component to the model. Using equation 3.14 as the base model, we get:

$$\begin{aligned} C_i &= \beta_0 + \alpha_1 y_i + \boldsymbol{\alpha_2 y_i^2} + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \varepsilon_i, \\ \frac{\partial C_i}{\partial y_i} &= \alpha_1 + 2\alpha_2 y_i \end{aligned} \tag{3.17}$$

With this adjustment, the marginal effect of output on costs does now depend on the level of output produced - referring back to the behaviour of long-run average costs, we might expect $\alpha_1 < 0$ and $\alpha_2 > 0$, which would imply that there is cost reduction until y_i is sufficiently large, at which point the costs begin to increase.

This behaviour could analogously extend to any and all inputs, with the same respective marginal effects. However, this quadratic form still makes the claim that the inputs of production are separate - which is still not necessarily true. To address this issue, interactive terms between the inputs can be added in:

$$\begin{aligned} C_i &= \beta_0 + \alpha_1 y_i + \beta_1 x_{1i} + \beta_2 x_{2i} + \boldsymbol{\gamma_1 x_{1i} x_{2i}} + \cdots + \varepsilon_i, \\ \frac{\partial C_i}{\partial x_{1i}} &= \beta_1 + \gamma_1 x_{2i} \end{aligned} \tag{3.18}$$

The marginal effect looks similar to that of the quadratic form, but with one key difference: the effect of x_1 on costs now depends on x_2. Taking a practical example, suppose that x_1 is Capital, and x_2 is Labour. In theory, the variables are independent, as they represent infrastructural and operational costs respectively, but in practice, how might a machine produce its good without the labour to operate it? In this example, we might find that $\beta_1 > 0$ and $\gamma_1 < 0$ - costs do increase when using more capital, but decrease if more labour is used to operate it.

This notion of quadratic terms does not only apply to input combinations or outputs. For example, Arocena et. al. (2009) utilise a model of total that includes quadratic terms for all combinations of outputs, prices, and interactions between outputs and prices in the US electric power industry. Much like the highlighted terms of (3.17) and (3.18), the model is defined as:

$$CT_i = \alpha_0 + \sum_{g=1}^{n} \beta_g y_{gi} + \sum_{g=1}^{m} \delta_g w_{gi} + \frac{1}{2} \sum_{g=1}^{n} \sum_{j=1}^{n} \beta_{gj} y_{gi} y_{ji} + \frac{1}{2} \sum_{g=1}^{m} \sum_{j=1}^{m} \gamma_{gj} w_{gi} w_{ji} + \sum_{g=1}^{n} \sum_{j=1}^{m} \theta_{gj} y_{gj} y_{gi} w_{ji} + \sum_{g=1}^{h} \psi_g Z_{gi},\ \forall i = 1, \ldots, f$$

In this model, outputs y and input prices w are both linear and quadratic, inclusive of the combinations between them. Rather than using quadratic inputs in their total cost model, they instead derive the inputs via the partial derivatives of the total costs with respect to the relevant input price. This model, alongside the previous general model using inputs, demonstrates the flexibility of the quadratic form in defining costs, and deriving measures from costs.

The Translog Functional Form

The final model looked over here is the most general - the Transcendental Logarithm, or Translog, model. Rather than addressing costs via its inputs, this model instead estimates the relationship costs have with outputs and factor prices, much like Arocena et. al. (2009)'s model, in a general form that captures the desirable behaviours of costs from a theoretical point-of-view:

$$\ln(C_i) = \beta_0 + \beta_1 \ln(y_i) + \frac{1}{2}\delta_{yy} \left(\ln(y_i)\right)^2 + \sum_j \delta_{yj} \ln(y_i) \ln(w_{ji}) + \sum_j \beta_j \ln(w_{ji}) + \frac{1}{2} \sum_j \sum_k \delta_{jk} \ln(w_{ji}) \ln(w_{ki}) + \varepsilon_i \tag{3.19}$$

Where $\delta_{jk} = \delta_{kj}$. This model in effect is the result of all of the model changes described above in sum, with the inputs replaced by their prices, and further interactions between output and factor prices included. Further to this model, there a a few restrictions required on the parameters to properly adhere to the producer-theoretical properties of the cost function:

$$\sum_j \beta_j = 1$$
$$\sum_j \delta_{yj} = 0$$
$$\sum_j \sum_k \delta_{jk} = 0$$

Which ensure that the cost function is homogenous of degree 1 in all factor prices: proportional increases in input prices must increase costs by that same proportion. Interestingly, this type of function allows for scale effects of any kind, capturing both the decreasing and then increasing average costs.

3.2.3 Stochastic Frontier Analysis

Stochastic Frontier Analysis, interestingly, was first conceived of by two separate papers, Aigner et. al. (1977) and Meeusen & van den Broeck (1977). The principal difference between the two papers can be explained by defining both models used. First, Aigner et. al. (1977)'s model is defined as:

$$y_i = X_i'\beta + v_i + u_i,\ v_i \sim N(0, \sigma_v^2),\ u_i \sim N^-(0, \sigma_u^2) \tag{3.20}$$

The regression model differs from OLS models due to the decomposition of the error term into two parts v_i and u_i, the former representing the same white noise errors as OLS models require, and the latter representing errors due purely to inefficiency, assumed to be distributed on a Normal distribution truncated above zero. Another identical model instead truncates below zero, giving:

$$y_i = X_i'\beta + v_i - u_i,\ v_i \sim N(0, \sigma_v^2),\ u_i \sim N^+(0, \sigma_u^2)$$

Meeusen & van den Broeck (1977), on the other hand, provide a more general definition for a stochastic frontier model:

$$y_t = \phi(x_t)k_t u_t,\ k_t \sim (0,1),\ u_t \sim (0,\infty) \tag{3.21}$$

In this specification, k_t is the inefficiency measure, and u_t represents the white noise errors. They find that the exponential distribution is suitable for the model and, using an example Cobb-Douglas function, define a model as:

$$y_t = A\prod_j x_{t,j}^{\beta_j} e^{-z_t} e^{-v_t} \tag{3.22}$$

Where $k_t = e^{-z_t},\ z_t \sim (0,\infty)$, and $u_t = e^{-v_t}$ draws from random Gaussian errors v_t. These original models, though widely used, don't account for matters such as firm-effects, panel data, other types of efficiency, and so on. More contemporary models added to the depth of the original SFA models - for example, Lai & Kumbhakhar (2018) develop a Homoskedastic Four-Component Stochastic Frontier (H4CSF) model:

$$y_{i,t} = \beta_0 + x'_{i,t}\beta_1 + \tau_i - \eta_i + v_{i,t} - u_{i,t} \tag{3.23}$$

Where τ_i is Random Firm Effects, η_i is Persistent Inefficiency, and $u_{i,t}$ is Transient Inefficiency, the distinction between the latter two terms being one of whether the cause of inefficiency is independent across time t or not. This model is yet more general by allowing for determinants of both inefficiencies, and allowing those inefficiencies to be correlated with the regressors $x_{i,t}$. Another extended model comes from Lai & Kumbhakhar (2019), who derive technical and allocative efficiency scores from a panel stochastic frontier model using the model's First-Order Conditions (FOCs). The system that is then estimated is defined as:

$$
\begin{aligned}
&\ln(y_{i,t}) = \beta_0 + \sum_{j=1}^{J} \beta_j \ln(x_{j,i,t}) + \tau_i + v_{i,t} - u_{i,t} \\
&\ln(x_{1,i,t}) - \ln(x_{j,i,t}) = \ln\left(\frac{w_{j,i,t}}{w_{1,i,t}}\right) + \ln(\beta_1) - \ln(\beta_j) + \zeta_{j,i,t}
\end{aligned}
\tag{3.24}
$$

Where τ_i in this model instead represents Fixed Firm Effects, $w_{j,i,t}$ represents Input Prices, and $\zeta_{j,i,t} = \mu_{j,i} + \chi_{j,i,t}$ is Allocative Inefficiency, composed of a persistent component μ and time-varying component χ. This FOCs of this system, then, grant a measure of allocative inefficiency between two inputs, inputs 1 and j.

These kinds of model are certainly interesting considerations, but have the major restriction, by design, of requiring a functional form of the regression, such as linear, log-linear or translog, analogous to the forms of cost models discussed earlier in the chapter. In theory, semi- or non-parametric models could be achieved to remove this constraint, but would have the trade-off of much greater computational complexity.

3.2.4 Data Envelopment Analysis

From the definitions of Section 3.2.1, and in particular from Debreu (1951)'s definition of technical efficiency, the first description of what would become one of the conventional empirical models of efficiency - Data Envelopment Analysis - comes from Farrell (1957), who sought to identify an empirical specification for isoquants between multiple inputs.

Farrell (1957) wrote on the need to estimate the functions in producer without need for a specific functional form - indeed, as per the questions stated at the beginning of this section, he could not necessarily assume a functional form of, say, an isoquant curve, nor could he assume that all firms would reach such a curve in practice. To address these problems, he had suggested that a curve of piecewise linear equations could be used to 'envelope' the data of interest, such that the curve behaved according to producer theory. Figures 1 and 2 of

Farrell (1957) show an isoquant in theory and in practice.

For a firm to be technically efficient in Figure 1 on the curve SS', they would need to be at the point $\vec{OQ}$, and are allocatively efficient at the point Q' on that same curve. Were a firm instead at point P, their technical efficiency could be measured by:

$$TE = \frac{\vec{OQ}}{\vec{OP}} \tag{3.25}$$

In practice, Farrell (1957) determined that, in a set A of points containing the data observations, and the points $(0, \infty)$ and $(\infty, 0)$ necessary for the maths to work out, pairs of points are chosen such that the connecting line segment does not have a positive gradient, and that there are no points between the line and the origin. Algebraically, this becomes:

$$\begin{aligned} \lambda x_{i1} + \mu x_{j1} &= x_{k1} \\ \lambda x_{i2} + \mu x_{j2} &= x_{k2} \end{aligned} \tag{3.26}$$

For three points P_i, P_j and P_k of the form $P_i = (x_{i1}, x_{i2})$ in A, with λ_{ijk} and μ_{ijk} being the solutions to the above equations. The line segment joining P_i and P_j is a part of the empirical curve if and only if $\lambda_{ijk} + \mu_{ijk} \geqslant 1\ \forall P_k \in A$, and the technical efficiency can be defined as:

$$\hat{TE} = \frac{1}{\lambda_{ijk} + \mu_{ijk}} \tag{3.27}$$

This example can be extended to as many dimensions as is required, with the same consequent analytical estimate of technical efficiency, leading to Figure 2 graphically.

The next major advancement in representing this kind of mathematical problem was brought about be Charnes et. al (1978), who sought to turn the system of linear equations into a Fractional Programming problem. In doing so, it was also found that the problem could alternative be seen as a far more solvable Linear Programming problem, defining what remains the base model of DEA:

$$\begin{aligned} \min_{\theta,\lambda} \theta_i,\ s.t. \\ \theta_i X_{i_o} \geqslant \sum_{i=1}^{N} \lambda_i X_i,\ \forall i, \\ Y_{i_o} \leqslant \sum_{i=1}^{N} \lambda_i Y_i,\ \forall i, \\ \sum_{i=1}^{N} \lambda_i = 1;\ \lambda_i \geqslant 0\ \forall i \end{aligned} \tag{3.28}$$

$\theta_i \in (0, 1]$ is, for a firm i, their technical efficiency, more specifically their Input Technical Efficiency, as they are adjusting inputs. In this model, θ minimises all inputs at the same rate, subject to a collection of constraints:

Input Constraint: $\theta_i X_{i_o} \geqslant \sum_{i=1}^{N} \lambda_i X_i$, ensures that the inputs of one company i_o, scaled in this case by the efficiency score θ_i, cannot be less than a weighted sum of all inputs across all firms - there is some sort of lower bound for the inputs.

Output Constraint: $Y_{i_o} \leqslant \sum_{i=1}^{N} \lambda_i Y_i$, similarly ensures that, for the reference firm i_o, their outputs cannot exceed a weighted maximum of all firms' outputs - there is an upper bound for the outputs.

Returns-to-Scale Constraint: $\sum_{i=1}^{N} \lambda_i = 1$, determines the returns-to-scale of the model. As eluded to in previous theory section, this constraint can be changed to give the model restrictions to its shape due to to differing returns-to-scale. This constraint in particular allows for Variable Returns-to-Scale (VRS), which allows for their to be different scales at different production bundles.

Emphasis in the above model was placed on it being input-oriented, in that the efficiency score is determined through minimising the inputs. Analogous to this, an Output-Oriented model can also be defined as:

$$\begin{aligned} &\max_{\varphi,\lambda} \varphi_i, \ s.t. \\ X_{i_o} &\geqslant \sum_{i=1}^{N} \lambda_i X_i, \ \forall i, \\ \varphi_i Y_{i_o} &\leqslant \sum_{i=1}^{N} \lambda_i Y_i, \ \forall i, \\ \sum_{i=1}^{N} \lambda_i &= 1; \ \lambda_i \geqslant 0 \ \forall i \end{aligned} \tag{3.29}$$

Which has the same definition and constraints as the model before it, except in this case we instead maximise the outputs of the model, via a measure which increases all outputs in the same proportion, $\varphi_i \in [1, \infty)$. Finally, both models 3.28 and 3.29 have a link via their efficiency scores:

$$\theta_i = \frac{1}{\varphi_i} \tag{3.30}$$

In effect, then, the choice of orientation of the model is mostly based on the research topic, rather than mathematical differences in the results: in this thesis' case, and throughout the industry literature, input-oriented models are used, as cost minimisation is seen as a primary focus. A further comparison of DEA models, more so focused on the different types of DEA specification rather than the orientation of the model, is found in Kohl & Brunner (2020), who evaluate DEA models via a Monte-Carlo translog production function estimation.

3.3 Industry Models

Many of the models for this section mirror the research literature covered by the academic results discussed in Chapter 2. When referring to industry models, the thesis means to refer to those models employed in the research literature, which align with many of theoretical models discussed in the previous section. To begin, Lynk (1993) and Hunt & Lynk (1995)'s models can be be defined, respectively, as:

$$
\begin{aligned}
c_{i,t} = \alpha_0 + \sum_{j=1}^{3} \beta_j y_{j,i,t} + \gamma_{1,2} x_{1,2,i,t} + \gamma_{1,3} x_{1,3,i,t} + \gamma_{2,3} x_{2,3,i,t} \\
+ \gamma_{1,2,3} x_{1,2,3,i,t} + \delta_1 w_{i,t} + \sum_{k=1}^{10} \phi_k Z_k + \sum_{m=1}^{9} \sigma_m T_m + v_t + u_t
\end{aligned}
\tag{3.31}
$$

$$
\begin{aligned}
c_{i,t} = \alpha_0 + \alpha_1 c_{i,t-1} + \sum_{j=1}^{3} \beta_j y_{j,i,t} + \gamma_{1,2} x_{1,2,i,t} + \gamma_{1,3} x_{1,3,i,t} + \gamma_{2,3} x_{2,3,i,t} \\
+ \gamma_{1,2,3} x_{1,2,3,i,t} + \delta_1 w_{i,t} + \sum_{k=1}^{9} \phi_k Z_k + \sum_{m=1}^{7} \sigma_m T_m
\end{aligned}
\tag{3.32}
$$

For both models, $c_{i,t}$ is the log of Total Costs of production for company i at time t; $y_{j,,it}$ is the log of Outputs, which are Water Supplied, Trade Effluent, and Environmental Activities performed; x represent interactions between outputs, for example $x_{1,2,i,t} := y_{1,i,t} y_{2,i,t}$; $w_{i,t}$ is the Unit Labour Cost; Z_k is a set of dummy variables for all but on of the RWAs in the industry, and T_m is a set of dummies for all but one years used in the sample.

Lynk (1993) opt to use a SFA-type regression by having both two-sided idiosyncratic and one-sided inefficiency error terms, finding a mean difference in mean inefficiency of 9.75% between pre-privatisation Regional Water Authorities (RWAs) and Statutory Water Companies (SWCs), and a 9.61% difference in mean inefficiency in those companies after quality had been accounted for, via indices reflecting relative environmental and sewerage quality for y_3 and y_2, respectively. Hunt & Lynk (1995), though exploring similar interests, opt for a dynamic model with standard error composition in the model, finding significant and stationary behaviour by including previous costs as a factor affecting current costs.

As the research literature developed, cost estimation remained a fairly constant focus, with developments in derived measures from cost estimation, such as for economies of scale or scope, becoming more ubiquitous. A good example of models in this space comes from Bottasso & Conti (2009b), who estimate a translog variable cost function to assess the industry's price cap regulation, accounting for a general technology index $A(t)$:

$$
\begin{aligned}
\ln(VC_{it}) = \alpha_0 &+ \sum_{i=1}^{I} \lambda_i D_i + A(t) + \sum_{j=1}^{J} \beta_j \ln(p_{jit}) + \sum_{n=1}^{N} \gamma_n \ln(y_{nit}) \\
&+ \sum_{v=1}^{V} \delta_v \ln(k_{vit}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{s=1}^{J} \beta_{js} \ln(p_{jit}) \ln(p_{sit}) \\
&+ \frac{1}{2} \sum_{n=1}^{N} \sum_{p=1}^{N} \gamma_{np} \ln(y_{nit}) \ln(y_{pit}) + \frac{1}{2} \sum_{v=1}^{V} \sum_{x=1}^{V} \delta_{vx} \ln(k_{vit}) \ln(k_{xit}) \\
&+ \sum_{j=1}^{J} \sum_{n=1}^{N} \rho_{jn} \ln(p_{jit}) \ln(y_{nit}) + \sum_{j=1}^{J} \sum_{v=1}^{V} \psi_{jv} \ln(p_{jit}) \ln(k_{vit}) \\
&+ \sum_{n=1}^{N} \sum_{v=1}^{V} \chi_{nv} \ln(y_{nit}) \ln(k_{vit}) + \sum_{m=1}^{M} \zeta_m z_{mit} \\
&+ \sum_{j=1}^{J} \beta_{jA} \ln(p_{jit}) A(t) + \sum_{n=1}^{N} \gamma_{nA} \ln(y_{nit}) A(t) \\
&+ \sum_{v=1}^{V} \delta_{vA} \ln(k_{vit}) A(t) + u_{it}
\end{aligned} \tag{3.33}
$$

The model is then used to derive the shares of variable costs attributed to an input i, $S_i = \frac{\partial \ln(VC)}{\partial \ln(p_j)}$, as well as technical change. This model not only affirms the aforementioned flexibility of the translog model, but the wider paper also shows some of the applications of the modelling which provide wider insights to the industry's cost behaviour. Other calculations from cost models include economies of scale (Ashton (2003), Bottasso & Conti (2003, 2009a)) and scope (Saal & Parker (2000)) as mentioned, as well as economies of output and customer density (Bottasso & Conti (2003,2009a)), efficiency change (Saal et. al. (2007)), and so on.

A more contemporary trend in the literature modelling is the use of Meta-Frontiers, which first create subsets of the firms, and finds their sub-group efficiencies. Then, the whole sample is used for another set of efficiency scores, and the two are compared to create a Technology Gap Ratio, or something similar. Molinos-Senante & Maziotis (2019), for example, use the following cost function for a sub-group j:

$$C_{it}^{j} = \exp(X_{it}\beta^{j} + V_{it}^{j} + U_{it}^{j}) \tag{3.34}$$

The meta-cost function can be defined analogously, and is labelled as C_{it}^{*}. Defining Cost Efficiency for both types of model as the ratio of the estimated regression to total actual costs, the paper finds a Cost-Gap Ratio in cost efficiencies, defined as:

$$CE_{it}^{*} = CE_{it}^{j} CGR_{it}^{j} \tag{3.35}$$

That is, the cost gap ratio is found as the difference between whole-group and sub-group efficiencies; in effect, this approach accounts for technology differences between the sub-group and whole group samples, and so the gap can be defined as the difference (in cost efficiency) due to technological differences relative to the total frontier.

This is a very pertinent extension to many of the models in the previous section, in the context of the water industry, as it allows, to some extent, for both WaSCs and WOCs to be directly comparable in their efficiency scores, scaling for differences in technology. One gap not directly accounted for, however, is the totality of the operations in WaSCs compared to WOCs: by construction, the factors of production from which meta-frontier efficiency scores are derived must be constant across both types of water company - there cannot, for example, be an inclusion of wastewater outputs in the model, as WOCs cannot express production in wastewater. One could, as a solution, include wastewater but set all WOCs' outputs in that section of the industry to zero, but that may change the consequent efficiency calculations.

Another recent example of the application of meta-frontiers in the industry is Mocholi-Arce et. al. (2020)'s assessment of company performance including economic bads to reflect undesirable service quality measures. Though they use a Malmquist index approach to the modelling, which will be discussed later in the chapter, they operationalise the model via the following system, for pairs of time periods $r = t, t+1$:

$$
\begin{aligned}
&\overrightarrow{D^{df}}(X_{h',r}, Y_{h',r}, B_{h^i,r}) = \max\ \beta, s.t. \\
&\sum_{con} \lambda_{h,r} h Y^l_{h,r} \geqslant (1+\beta) Y^l_{h',r}, \forall l = 1, \dots, L \\
&\sum_{con} \lambda_{h,r} h B^n_{h,r} \geqslant (1-\beta) Y^n_{h',r}, \forall n = 1, \dots, N \\
&\sum_{con} \lambda_{h,r} h X^m_{h,r} \geqslant X^m_{h',r}, \forall m = 1, \dots, M \\
&\lambda_h \geqslant 0
\end{aligned} \tag{3.36}
$$

For L desirable outputs Y, N undesirable outputs B, and M inputs X. In this model, the meta-frontier comparisons are between groups h and h', and allow for heterogeneous groups of observations - WaSCs and WOCS in practice - to be compared, though as mentioned earlier in the chapter, they are compared only on water production factors: water connected properties and volume of water delivered, in this case. The use of meta-frontiers in this paper provide an interesting conclusion about the performances between WaSCs and WOCs, in that they find WOCs are more productive on average than WaSCs, when they are comparable to each other.

3.4 Other Empirical Models

Owing to the need of some restraint in the scope of the thesis, the previous section highlighted the most common models in the research literature, as the DEA models will be pertinent to the later contributing chapters, and indeed almost all of the literature uses these models. However, there exists a far greater expanse of models that aren't common to the literature or used in this project, but are worth going over in brief.

This section will go through what other models have been found in the search of empirical modelling literature, and will then be summarised in a table, alongside the previously mentioned models. This summary will hopefully serve as a point of reference for continuation in future research directions, but in this chapter also serves to summarise why DEA methods will be the primary type of models used in the thesis.

3.4.1 Other Non-Parametric Models

A good place to continue off from is the discussion of non-parametric models, which DEA is a member of. Though DEA models are themselves a class of model, there exist other similar non-parametric models that function closely to DEA, but have distinct differences that give them different nuances in estimation of efficiency scores. The succinct list of other non-parametric models in this chapter, and their key differences to DEA models, is as follows:

Multi-Dimensional Efficiency Analysis (MEA), which estimates a multi-dimensional Distance Function and uses a Circular Distribution to find not only overall inefficiency scores, but how each input contributes to that inefficiency.

Free Disposal Hull (FDH), which foregoes the Convexity assumption of producer theory to allow for the estimation of a non-convex production frontier.

(Inverse) Range-Directional Model ((I)RDM), which uses the Range measure to create a model of efficiency that can incorporate negative data, and can in the Inverse model be used to 'target set' goals for model factor improvements.

Multi-Dimensional Efficiency Analysis Models

Looking first at MEA models, the approach uses a two-stage procedure to optimise the distance between some ideal vector of netputs, and the vector of netputs gained from the data. The following is taken from Asmild et. al. (2016) as an example model of n DMUs, m Inputs and s Outputs:

$$\begin{aligned} z_i^{oI} &= \max(\delta_i),\ s.t. \\ &\sum_{j=1}^{n} \lambda^j z_i^j \geqslant \delta_i, \\ &\sum_{j=1}^{n} \lambda^j z_{-i}^j \geqslant z_{-i}^o,\ -i = 1, \ldots, i-1, i+1, \ldots, m+s, \\ &\lambda^j \geqslant 0,\ \forall j = 1, \ldots, n \end{aligned} \tag{3.37}$$

This first step considers individual improvements for each inputs and output, in order to create an estimated ideal netput vector, z^{oI}. Then, the second stage estimates the following:

$$\begin{aligned} \beta^o &= \max(\beta),\ s.t. \\ &\sum_{j=1}^{n} \lambda^j z_i^j \geqslant z_i^o + \beta(z_i^{oI} - z_i^o),\ \forall i = 1, \ldots, m+s, \\ &\lambda^j \geqslant 0,\ \forall j = 1, \ldots, n \end{aligned} \tag{3.38}$$

where β is the magnitude in the direction of z_i^{oI} required for z_i^o to be optimal. The benchmark vector for the model, then, can be given by $\mathbf{z}^{oB} = \mathbf{z}^o + \beta^o(\mathbf{z}^{oI} - \mathbf{z}^o)$, which allow for the calculations of Absolute and Relative Inefficiency Scores for each i, respectively:

$$AI_i = z_i^{oB} - z_i^o, \qquad RI_i = \frac{z_i^{oB} - z_i^o}{z_i^o} \tag{3.39}$$

which is analogous to measures of technical efficiency being defined by, for example, the difference between optimally utilised resources and actually used resources defined in (3.7), or other technical efficiency measures defined as the ratio of optimally weighted inputs to total inputs. The contributions of each factor i to the overall measures can be defined by:

$$\begin{aligned}\theta_i^o &= \cos^{-1}\left(\frac{\mathbf{z}_i^{oI} - \mathbf{z}_i^o}{||\mathbf{z}^{oI} - \mathbf{z}^o||}\right),\ \theta_i^o \in \left[0, \frac{\pi}{2}\right]\ \forall i,\\ \theta \sim g(\theta;\mu,\kappa) &= \frac{1}{2\pi I_0(\kappa)} e^{\kappa} \cos(\theta - \mu) \Big/ \int_0^{\frac{\pi}{2}} \frac{1}{2\pi I_0(\kappa)} e^{\kappa} \cos(\theta - \mu) dx\end{aligned} \tag{3.40}$$

Where θ_i^o is mathematically the angle found between the vectors $\mathbf{z}_i^{oI}$ and $\mathbf{z}_i^o$, the vector of which is distributed on the Truncated von-Mises Distribution $g(\theta;\mu,\kappa)$ for some Mean parameter μ and Concentration parameter, κ, which is restricted to only the quadrant where the netput vectors exist. [1]

Free Disposable Hull Models

The FDH model is similar to DEA models in design, with its only principal difference, as mentioned in the summary above, being the relaxation of the convexity assumption typical to production economics and consequently the estimation of DEA models. The estimation of efficiency scores, under an FDH specification, is taken from Kneip et. al. (2016), and is defined as:

$$\hat{\theta}_{FDH} = i \in \mathcal{I}(y) \min \left(\max_{j=1,\dots,p} \left(\frac{X_i^j}{x^j} \right) \right) \tag{3.41}$$

Where, for p inputs and n DMUs, the estimated efficiency score in the minimum of the maximum ratios of input vectors X_i^j and x^J, for the subset of i in $\mathcal{I}(y) = \{i \mid y_i \geqslant y,\ i = 1,\dots,n\}$.

The interest of FDH modelling in this thesis, though passing, is interesting. Kneip et. al. (2016) note that FDH models are, in effect, a more flexible version of VRS DEA models, as do not adhere to convexity as other production

[1] The distribution also uses the modified Bessel function of the first kind:

$$I_\nu(x) = \sum_{r=0}^{\infty} \frac{1}{r!\ \Gamma(r+\nu+1)} \left(\frac{x}{2}\right)^{2r+\nu},$$

where $\Gamma(x) = (x-1)!,\ x \in \mathbb{Z}_+$ is the Gamma Function.

function specifications do. Furthermore, their paper finds a system with which CRS DEA, VRS DEA and FDH models can be decided upon for empirical research, suggesting that future industry endeavours might find some use from selecting and comparing FDH models to the already established DEA models.

Range-Directional Models

Lastly, the RDMs instead forego the DEA assumption that data is non-negative. Portela et. al. (2004) base their estimation on the directional distance function of Chambers et. al. (1996,1998) and, for m inputs, s outputs and $J = \{1, \ldots, n\}$ DMUs, first define the 'Range of Possible Improvement of unit o' as:

$$\begin{aligned} R_{ro} &= \max_J(y_{rj}) - y_{ro},\ \forall r = 1, \ldots, s \\ R_{io} &= x_{io} - \min_J(x_{ij}),\ \forall i = 1, \ldots, m \end{aligned} \tag{3.42}$$

The RDM is, then, defined as:

$$\begin{aligned} &\max(\beta_o),\ s.t. \\ &\sum_{j=1}^{n} \lambda_j y_{rj} \geqslant y_{ro} + \beta_o R_{ro},\ \forall r = 1, \ldots, s, \\ &\sum_{j=1}^{n} \lambda_j x_{ij} \leqslant x_{io} - \beta_o R_{io},\ \forall i = 1, \ldots, m, \\ &\sum_{j=1}^{n} \lambda_j = 1,\ \lambda_j \geqslant 0 \end{aligned} \tag{3.43}$$

The resultant efficiency score is defined as $1 - \beta$, which can alternatively be defined according to whether either the input constraint or output constraint is binding, which is required for optimal solutions by definition:

$$1 - \beta = \begin{cases} \frac{\max_J(y_{rj}) - y_r^*}{\max_J(y_{rj}) - y_{ro}}, \text{if Output constraint is binding,} \\ \frac{x_i^* - \min_J(x_{ij})}{x_{io} - \min_J(x_{ij})}, \text{if Input constraint is binding.} \end{cases} \tag{3.44}$$

The IRDM specification arises from the mechanical bias of the RDM procedure, as it prioritises factors for which a DMU has scope for the most improvement which, while beneficial in the long-term as targets, does not provide much use for short-term improvements. The IRDM is defined as:

$$
\begin{aligned}
&\max(\beta_o),\ s.t. \\
&\sum_{j=1}^{n} \lambda_j y_{rj} \geqslant y_{ro} + \beta_o \frac{1}{R_{ro}},\ \forall r = 1, \ldots, s, \\
&\sum_{j=1}^{n} \lambda_j x_{ij} \leqslant x_{io} - \beta_o \frac{1}{R_{io}},\ \forall i = 1, \ldots, m, \\
&\sum_{j=1}^{n} \lambda_j = 1,\ \lambda_j \geqslant 0
\end{aligned}
\tag{3.45}
$$

The interpretation of $1-\beta$ from the IRDM is more opaque, as the model sets reference points for each observation, creating issues with comparing efficiencies between DMUs or the RDM efficiency scores. In practice, this could be a useful tool from a regulatory perspective, as it could allow companies to understand and be better set targets by Ofwat in their price reviews.

3.4.2 Semi-Parametric Models and Model Conversion

Tangential by definition to non-parametric models are Semi-Parametric Models, which consist partly of known functional forms, and partly of unknown but estimate-able functions. A general regression form of this type of model can be expressed as:

$$y_i = x'_{1i}\beta_1 + f(\beta_2; x_{2i}) + \varepsilon_i \tag{3.46}$$

Where the outcome y for observation i is regressed on variables $x = (x_1, x_2)$, with variables x_1 possessing a known linear functional form in this example, and the remaining variables x_2 possessing an unknown functional form f. A further generalisation would also be to suggest that the functional form for x_1 is known, and not necessarily linear in x_1, and x_2 is still in some unknown form f.

There isn't too much of semi-parametric modelling in efficiency analysis or public utilities, but one interesting example is by Kuosmanen & Kortelainen (2012), who create a semi-parametric model that uses both the encompassing frontier of DEA models with the composite error term of SFA models. This is achieved via a two-stage procedure known as the Stochastic Non-Smooth Envelopment of Data (StoNED) approach for a model $y_i = f(x_i) + v_i - u_i$:

1. To estimate the shape of the function f, a Convex Non-Parametric Least Squares (CNLS) regression is used.

2. With extra assumptions on the error distributions, σ_v^2 and σ_u^2 are estimated, and the conditional expected values of inefficiency can then be computed.

To best manage the estimation of the first stage of the StoNED approach, Kusomanen (2008) finds the following Quadratic Programming problem:

$$
\begin{aligned}
&\min_{v,\alpha,\beta} \sum_{i=1}^{N} v_i^2,\ s.t. \\
&y_i = \alpha_i + \beta_i' x_i + v_i, \\
&\alpha_i + \beta_i' x_i \leqslant \alpha_h + \beta_h' x_i,\ \forall h, i = 1, \ldots, N, \\
&\beta \geqslant 0\ \forall i
\end{aligned}
\tag{3.47}
$$

The problem minimises the squared residuals accounted for in the first constraint, which is a linear regression. The other constraint represents a system of Afriat Inequalities (Afriat (1967,1972)), whose total satisfaction allows for Afriat's Theorem, which states that there must exist a monotonically increasing, concave function $\hat{f}$ such that $y_i = \hat{f}(x_i) + v_i$.

The second stage of the process follows the derivation of the composite error distribution in Aigner et. al. (1977), which is used to find estimates for σ_v and

σ_u:

$$\hat{\sigma}_v = \sqrt{\hat{M}_2 - \left(\frac{\pi - 2}{\pi}\right)\hat{\sigma}_u^2},$$
$$\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\left(\sqrt{\frac{2}{\pi}}\right)\left(1 - \frac{4}{\pi}\right)}} \tag{3.48}$$

Where $\hat{M}_2$ and $\hat{M}_3$ are the estimated second and third moments of the error distribution, respectively. Using the estimates of the composite error term, $\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u\sqrt{2/\pi}$, the conditional inefficiency scores can be found as:

$$\hat{\mathbb{E}}(u_i|\hat{\varepsilon}_i) = -\frac{\hat{\varepsilon}_i\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} + \frac{\hat{\sigma}_u^2\hat{\sigma}_v^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}\left(\frac{\phi(\hat{\varepsilon}_i/\hat{\sigma}_v^2)}{1 - \Phi(\hat{\varepsilon}_i/\hat{\sigma}_v^2)}\right) \tag{3.49}$$

Which is, in essence, calculated as the entire error less the estimated composite error, which only contains the random error component v_i, plus a further adjustment utilising the standard Normal PDF and CDF functions, ϕ and Φ, of the composite error estimates.

Another interesting idea, coming from Kuosmanen & Johnson (2010), is the idea of model conversion - that is, being able to convert a DEA model into a regression model. One the key disadvantages to DEA in terms of interpretability is that it lacks marginal effects, which are estimated as the model parameters of a regression model. To that end, Though DEA and SFA models may both measure efficiency, they cannot necessarily be compared to each other. This paper, however, defines a Corrected Concave Non-Parametric Least Squares (C^2NLS) method of estimation, which estimates a concave quadratic program, and then uses the same corrections seen in COLS to reach the final efficiency estimations. The quadratic program is defined as:

$$
\begin{aligned}
&\min_{\alpha,\beta,\varepsilon} \sum_{i=1}^{N} \varepsilon_i^2, \ \ s.t. \\
&\varepsilon \leqslant 0, \\
&y_i = \alpha_i + \beta_i' x_i + \varepsilon_i, \\
&\alpha_i + \beta_i' x_i \leqslant \alpha_h + \beta_h' x_i \ \forall h, i = 1, \ldots, N, \\
&\beta_i \geqslant 0
\end{aligned}
\tag{3.50}
$$

Which is much like the previous quadratic program, with the addition of non-positive value for ε_i. From this program, all parameters and model residuals can be estimated, which matches the outcomes of standard regression models. Taking the estimated errors, ε^{DEA}, and given that the model is a least-squares formulation, a goodness of fit statistic can be derived, which does not exist inside the scope of DEA modelling specifications:

$$
R^2 = 1 - \frac{\sum_i (\varepsilon^{DEA})^2}{\sum_i (y_i - \bar{y})^2}, = 1 - \frac{\sum_i ((1-\theta_i) y_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3.51}
$$

The numerator of the measure uses model residuals and, since regressions for efficiency account for inefficiency in the error term, the numerator effectively becomes the sum-square of y_i, scaled by the inefficiency of the observation. Assumedly, $\bar{R}^2$ can also be calculated analogously. Lastly, the COLS-style corrections are used to shift the residual and intercept values:

$$
\begin{aligned}
\hat{\varepsilon}_i^{C^2NLS} &= \varepsilon_i^{CNLS} - \max_h \varepsilon_h^{CNLS} \\
\hat{\alpha}_i^{C^2NLS} &= \alpha_i^{CNLS} + \max_h \varepsilon_h^{CNLS}
\end{aligned}
\tag{3.52}
$$

3.4.3 Index Decompositions

One major area of the literature that has been mentioned previously, but is not utilised later in the thesis, is that of Index Models of factors of the industry, such as Cost or Profit. An index, in its broadest sense, looks to compare two entries to each other, to get a measure of one observation to the other. To better get a look at this idea in practice, there exists a wealth of research in the water industry that prefer index approaches over the previously described models.

Take, for example, Maziotis et. al. (2015)'s introductory Profit Change Index:

$$\pi_{i,t} = \frac{\Pi_{i,t}}{\Pi_{i,1}} \tag{3.53}$$

Profits of a firm i in time t are compared relative to the profits made in the first time period, though in general any base period can be used. The same principle applies generally to any factor of importance, such as Costs or Productivity. Furthermore, to better understand how these changes have happened, and what causes the changes, various decompositions have been considered. A primary paper on this kind of decomposition is Diewert & Fox (2017) who find, axiomatically, a decomposition into Technical Progress, the difference in technology between two periods, Technical Efficiency Change, and Returns-to-Scale, all of which could be explanatory factors in context for changes in, say industry profits per the Maziotis et. al. (2015) paper.

3.5 Conclusion

There are far more models that could be used for efficiency than were discussed in this chapter. However, referring to the models that appear conventionally in the literature, this chapter has looked at the details surrounding modelling ideas that will be carried forward to the subsequent research chapters.

This thesis intends to use DEA as its primary modelling structure, owing to its lack of a functional form allowing the most flexibility of the models described. This choice also adheres to arguably the most conventional models in the literature, which means that comparison to old results in the research chapters could be used, in theory, to best highlight how the new research of the thesis changes relative to preceding results.

Chapter 4

Data Review

A short chapter should be devoted to the Data used in the literature, now that it has been detailed, so that, in the later contributing chapters, there can be a good understanding as to why particular choices about the data used in this thesis were made, and how those choices are either conventional or relatively new or novel in the research area.

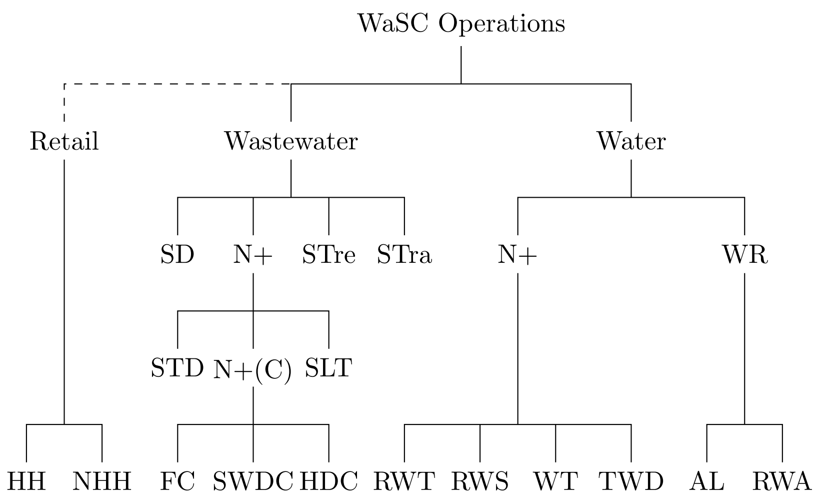
Broadly speaking, the data can be split into a small set of categories: Inputs, Outputs, Quality Variables, Non-Discretionary Variables, Prices, and Other Variables, which encompasses any variables that don't fit neatly into the other preceding categories. To motivate these categories somewhat, as well as to define what the industry structure is viewed as for the purposes of modelling, there is first a discussion of how the water and wastewater process is specified. Finally, before covering the categories of variables, there will also be a discussion on the pertinent matter of dimensionality, what problems that can cause for the forthcoming models and their results, and how those problems might be addressed.

4.1 Industry Specification

Table 4.1: Figure 4.1 Key of WaSC Sub-Operations

Name:	Meaning:
SD	Sludge Disposal
N+	Network Plus
STre	Sludge Treatment
STra	Sludge Transport
WR	Water Resources
STD	Sewage Treatment/Disposal
N+(C)	Network Plus (Collection)
SLT	Sewage Liquor Treatment
HH	Household Retail
NHH	Non-Household Retail
FC	Foul Collection
SWDC	Surface Water Drainage Collection
HDC	Highway Drainage Collection
RWT	Raw Water Transport
RWS	Raw Water Storage
WT	Water Treatment
TWD	Treated Water Distribution
AL	Abstraction Licences
RWA	Raw Water Abstraction

Figure 4.1: RAG WaSC Operations Breakdown, Ofwat (2021)



First and foremost, it is useful to illustrate how it is the industry is organised. Figure 1 of Thanassoulis (2000) illustrates a breakdown of WaSC functions into Clean and Dirty Water Operations, which are themselves broken down into various, more-specific activities. However, the description that will be followed hereafter is Ofwat's more recent 'RAG' model structure, which is illustrated by

Figure 4.1[1] and its related Key, Table 4.1, so that an idea of how the industry is considered partitioned, for the purpose of determine output variables later in the chapter, for example, can be made.

The industry is broadly split into two main sub-sections: Water and Wastewater, with a third Retail sector that was re-introduced in 2017. Given the need in any econometric model to appropriately choose Inputs and Outputs for the models used, Figure 4.1 can be used to determine to some extent what parts of the industry are most useful to represent in the forthcoming chapters.

So, what parts of the industry stand out as the most representative, from an empirical perspective? In both the Water and Wastewater sub-sectors, there are various categories of the overall process that describe facets of the total services provided: for example, the Water Resources section of Water services describes how the resultant provided water is first drawn from its sources, and from that the Network Plus sub-section covers its treatment and eventual distribution to the customer base. Analogously, Wastewater is split into the operations surrounding industrial Sludge, which is distinct from the remaining operations that deal with the collection, treatment, and disposal of Sewage.

The most natural choice, then, might be to represent each sub-sector's services by measures of some final output. As will be described shortly, there are also further considerations around the customer that also have to be accounted for, namely that this final measure of output has to include Household and Non-Household customers, unless otherwise specified.

And what of the Retail section? Upon its re-introduction, the retail markets allowed for customers to take their water and wastewater services from any available company in the industry, promoting a degree of competition into the industry due to the capacity for customers to now choose who will provide their water utilities. It is therefore somewhat pertinent to have some form of measure related to this part of the industry, even if it remains auxiliary to the primary

[1]Figure 4.1 is an exact reproduction of Figure 1 of Thanassoulis (2000), with an added Retail branch.

services.

This section only serves to provide a brief context to the industry, before other section formally identify what data is useful for modelling purposes. In doing so, what has been identified is the need to properly choose data that best represents the overall scope of what services the industry provides. This extends beyond the Outputs discussed above; Inputs, similarly must also be chosen to best cover as much of the production requirements of the industry as possible, and any other specific issues such as the incorporation of Quality, Price and other Non-Discretionary Variables must also be addressed. In doing so exhaustively, however, there can be an issue relating to Dimensionality, which itself leads to a more complicated decision around the choice of data than is initially conceived.

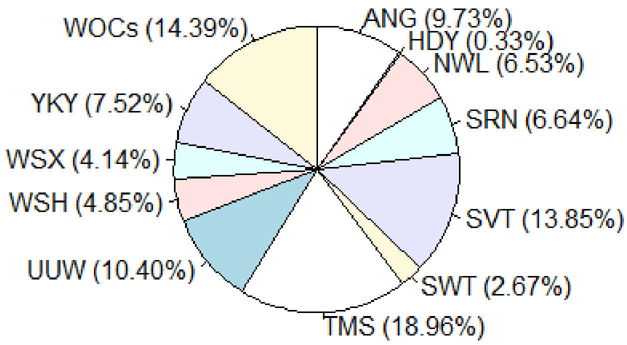
4.2 Sampling Frame

Prior to discussing what data is fit to represent the processes of the industry, the sample for this project should be defined. There are two key choices in the selection of data for this industry: whether both WaSCs and WOCs, or just WaSCs, should be used; and for how many years should the industry be observed?

The decision of whether or not to include WOCs into the sample is a significant one, owing to the the fundamental operating differences of those companies who, by definition, do not have wastewater processes, compared to the WaSCs that do. Since this project wishes to look at a representative whole of the industry's services and their production, it therefore concerns itself with both water and wastewater processes, and so will exclude the WOC companies from the sample.

There are trade-offs for both sides of this decision. On the one hand, includ-

Figure 4.1: Shares of Measured and Unmeasured Properties, 2019/20



ing all companies results in far more data, to the point where the foreshadowed dimensionality issues are far less likely to be an issue. On the other hand, because of the differences in production behaviour, the inclusion of WOCs into models that also look at wastewater production can lead to unrepresentative results, as the WOCs would be biased in some fashion by their 'zero' wastewater-specific outputs and Totex contributions.

Though one solution to this problem could be to use Meta-Frontier modelling, as Molinos-Senante et. al. (2015a) did for their model of the water sector, the more fundamental exclusion of sewerage operations from WOCs does not permit such a solution, since the model would be trying to compensate for what is effectively two different production processes.

Closing the choice of DMUs in the sample is a small caveat, which concerns the additional omission of Hafren Dyfrdwy, which is a company operating in Wales that, though initially a WOC, has been operating as a WaSC as of 2017, after it was purchased by Severn Trent, as separate WaSC to its parents operating in North Wales. The exclusion of this company is due to the fact that, although it does meet the definition of a WaSC by definition, it does not have the scale of operations afforded to the other WaSCs who have operated as such since at least their 1989 privatisation - effectively, the company has the size of a WOC, which can results in biased results compared to its WaSC counterparts.

To remedy this exclusion, in those years where the company operates as a DMU, its data is added to Severn Trent's. Justifying this decision further is Figure 4.1, which shows that Hafren Dyfrdwy has a customer share of just 0.33% in the industry, about eight to nine times smaller than the smallest WaSC, SWT, who has a share of 2.67%.

The other choice in the sample selection is the length of the data window. The literature has a variety of windows for their respective models, and this project will choose to opt for a longer window, beginning in the reported year 2002/03, and ending in the 2019/20 year, providing an eighteen-year span of time in which all WaSCs operate.

The reasoning for this longer window is so that the models in the later chapters can cover an amount of time that includes multiple regulatory periods: assuming that the year that a Price Review is set to begin is the start of a regulatory period - PR04 beginning in 2004/05 and ending in 2008/09, for example - this window contains three complete regulatory periods, as well some of both PR99 and PR19. This yields interesting points of observation from a policy point-of-view, as if the forthcoming results see significant differences in certain regulatory sub-samples, then there is an implication that those changes could be due to regulatory changes. This is more apparent in parametric regressions that use regulatory dummy variables, such as in Erbetta & Cave (2007) and Bottasso & Conti (2003), wherein the marginal effects of the changes in regulation can be estimated, but is nonetheless interesting in DEA models.

There is a downside to the use of a longer time window, however. As regulation changes, so too does the measurement of data, and this provides a significant difficulty in the selection of representative variables, as they may have had changes in their measurement, or simply had not existed for particular lengths of time. This issue is primarily a problem for the Common Performance Commitments that are addressed in the selection of new quality data, as despite their broad coverage of important factors of quality have only existed in their current state since PR14, with only a fraction of these variables consistently identified in the entire data window.

So, in choosing to work with a longer data window, this project prioritises the interest of potential differences over regulatory periods for most of the models of the thesis, with the later section on quality data highlighting the belief that there is still sufficiently representative quality data throughout the longer time period.

The data are sourced primarily for all years from annual public company reports of performance and costs: from 2002/03 until 2011/12, these reports are the June Annual Returns (JARs); from 2012/13, the Annual Performance Reviews (APRs) were used instead, although both publications have the same fundamental purpose of yearly records of each company. An additional source is used for the collection of old Water Quality data, the compliance rates for which are collected from web archives of the Drinking Water Inspectorate's website. This data is collected for all but 2019/20, whereby the measures from the DWI were discontinued - since this data is historically very consistently similar across companies over the whole time period chosen, it is assumed that the last year's observations of this data are the same as those of the year prior.

4.3 Dimensionality Issues

One might think that, given the appropriate facets of the industry are chosen, the best way to address the behaviours of these areas is to incorporate as much relevant data as possible, conditional on those choice not leading to any confounding results, such as Spurious Regressions due to using multiple variables that explain the same thing. While this could, in theory, be doable in Parametric regression models, it is far less sensible in the models that this thesis will use - Non-Parametric models.

As Cooper et. al. (2001) describe, there is in all empirical models a need to manage its Degrees of Freedom, which allow for variability in the models that lead to believable results. In regression models, this is often not much of an issue, as the degrees of freedom lost from the inclusion of independent variables is often insignificant compared to the total number of observations. Even in this project, with its relatively small number of WaSCs being considered, the total observations in the data window are likely to be sufficient in parametric models.

This sufficiency is far less so in DEA models, however. As Cooper et. al. (2001) states, the problem of degrees of freedom is compounded in those models which use relative measures, which DEA models do by definition, in addition to the issues of losing degrees of freedom as variables are added to the model. With that in mind, the so-called Cooper's Rule is defined as:

$$N \geqslant \max\{ms, 3(m+s)\} \tag{4.1}$$

Where N, the number of DMUs, must exceed the maximum of two values related to the number of Inputs and Outputs, m and s respectively, in the model. This rule-of-thumb is a conventional measure of sufficient sample size in the literature, and if this condition is met in the model, then the model can be loosely considered to be satisfactory with respect to the available degrees of freedom.

This poses a somewhat immediate problem in the context of this industry, however. The DMUs used in this project - the ten WaSCs - therefore set $N = 10$, which means that, ideally, the total number of variables in the models cannot exceed three: (4.1) permits maximally either one or two Inputs, and then two or one Outputs, respectively. This can be doable, in theory, but for the purposes of this project, which wishes to incorporate quality as an additional output variable, this guideline cannot be met. This issue persists in the literature for most models, though panel or pooled regression models may abate small-sample issues with a sufficient time period. For non-parametric models, as discussed, however, this issue persists and requires some degree of solution or adjustment to remove any consequent biases.

So, what can be done? One solution to remedy this problem could be to reduce the variables into linear combinations of factors via Principal Component Analysis (PCA), or some such equivalent. Assuming additivity and linearity of the components this, theoretically, is a very useful method to address dimensionality, and is illustrated in Cordero-Ferrera et. al. (2010)'s two-stage model that utilises these components to provide non-discretionary adjustments to their models of Technical Efficiency. However, this method has the downside of convoluting the results of the model if, say, Inputs and/or Outputs were subject to PCA, such that the final number of components used satisfied Cooper's Rule. One of the more important things to draw from these models, in the context of this industry at least, is how the relative efficiencies of the companies are calculated from the actual industry variables, rather than a safer, but mechanically obscured, linear combination of various factors, and so PCA methods are not chosen for this thesis on the grounds of a lack of economic interpretation of the outcomes and consequent principal components.

Another way that this has been addressed is in Charles et. al. (2019), wherein a collection of dimension-reducing ideas were trialled. These methods are functionally similar to the outcome of PCA techniques, in that they propose various combinations if Input and/or Output merging, as to reduce the total number of variables in the model. This method seems slightly more pliable, given that the assumption is instead that the choices for merging variables is driven by context or expert opinion, and not an optimal weighting of factors as PCA methods produce.

If it seems unlikely that this rule-of-thumb for the amount of included variables cannot be satisfied, what then is a work-around? If it is believed hereafter in this thesis that the potentially mechanically advantageous, but contextually difficult, PCA-type procedures are not usable, another way to account for these problems is to consider Bootstrapping methods.

Rather than accounting for dimensionality issues by removing variables or creating combinations of the variables, bootstrapping procedures hold no restrictions with respect to the choice of variables, and instead attempt to address the resulting problems through adjusting the resulting Efficiency Scores. Two methods will be discussed a short appendix of the thesis: Simar & Wilson (1998)'s one-stage DEA procedure, and Simar & Wilson (2007)'s three-stage procedure, which is also used in Cordero-Ferrera et. al. (2010). But, in short, what these methods intend to achieve is the removal of any biases caused by the small sample - in context, this reduces the tendency for small-sample efficiency scores to be higher than expected, as they tend to their maximum value of 1 as more variables are added.

In summary, this thesis will opt for the risk of not using PCA-reduced variables or anything similar, and will attempt to remedy any sample-biases via merging Inputs, where possible, and the bootstrapping methods referenced above, which are shown in Appendix A.

4.4 Data Categories

The data that the literature broadly considers in its models can be split into a collection of major sub-categories, which can then themselves be further separated into more precise facets of the industry. Broadly, there are Inputs, Outputs, and Prices, as well as Quality Variables and Non-Discretionary Variables, which are categories tailored more towards this project. Finally, there is a category for any other interesting variables not categorised into any of the aforementioned types of data.

4.4.1 Outputs

Outputs are broadly defined as the product of the production process of a company, whatever that may be. Relating this back to Figure 4.1 and the previous discussion of the industry's design, Water Services and Wastewater Services can be considered the two main branches of Outputs in the literature.

A question arises for these outputs: what constitutes as a unit for the outputs? One interpretation is that the measurement should be a measure of the actual output - a Volume of Water delivered, for example - but another approach is to instead consider the Population that are served by these outputted services - in this case, the unit would be a Household that receives, say, any of the deliverable water. In each discussion of the main service outputs, attention will be drawn to both sides of this argument, and then to what the choice is for this thesis.

Water Outputs

The first water output to look at is the Volume of Water Delivered, measured in Ml per Day or per Year; this output reflects the actual activity of the company in delivering water to its customers, and includes both Potable Water and Non-Potable Water Delivered, reflecting some use of water for industrial purposes. This outputs appears to be the most conventional in the literature, with some of the earliest papers in the field, such as Lynk (1993) and Hunt & Lynk (1995) choosing to use this data as their water output.

An alternative, as eluded to in the prior section, is instead to consider the Total Connected Properties, which is the combination of Household and Non-Household Connected Properties. Rather than the physical output, which Water Delivered represents, Connected Properties reflects the quantity of customers who are affected by the water process.

Some papers, such as Garcia & Thomas (2001) and Stone & Webster (2004) suggest that, ideally, both variables should be used as outputs to best describe the water service, based on the findings that there are differences in the marginal costs and short-run and long-run Elasticities of Production and Customer Density, with the latter falling into diseconomy in the long-run.

Other more recent papers incorporate both measures of the water service, such as in Molinos-Senante & Mazioitis (2019, 2020) and Mocholi-Arce et. al. (2020, 2021). However, a critical difference between these papers' models and this thesis' is the exclusion of the wastewater section of the industry, which not only allows for more degrees of freedom in the model from including less outputs, but also allows for the WOCs to be included as further observations, allowing with ease the inclusion of multiple outputs describing the water process. As this project seeks to address the whole industry, and so necessarily requires both water and wastewater services, only one of these outputs can be used to describe water services.

Water Delivered is chosen to best measure the water services hereafter, owing to its representation of the overall process at the activity level, rather than at the household level, aligning with the conventions of the literature that also covers the overall industry in its analysis. This is supported further by Erbetta & Cave (2007)'s defence of the variable, noting that water delivered reflects the production of water via abstraction and treatment, which is the type of production outputs that this thesis wishes to employ.

Wastewater Outputs

Analogous to the water services, there are two main measurements that are used to denote Wastewater outputs: Physical Wastewater/Effluent, and Equivalent Population Served.

As with the water service representations respectively, Physical Wastewater is used as a measure of the overall activity of this part of the industry; Equivalent Population Served, then, reflects the corresponding customer base that are affected by the wastewater services provided. As before, there is merit to incorporating both facets into models where possible, to best reflect the total process effectively. Saal et. al. (2007) and Erbetta & Cave (2007) use both wastewater outputs in their models, and further use all four of the discussed outputs.

Interestingly, the conventional wastewater output used in the literature, and the one used in this thesis hereafter, is Equivalent Population Served, defined as a measure of sewage load based on assumptions of per capita capital required, rather than the corresponding Physical Wastewater. The choice of this output, and the water output of the previous sub-section, can be elucidated somewhat by the descriptions of the data in Erbetta & Cave (2007): Water Delivered reflects the production of water, via its abstraction and treatment; Water Properties reflects that water's distribution; Wastewater Properties reflects the collection of sewage; Physical Wastewater then finally reflects the treatment and disposal of that sewage.

In much the same way that the previous literature and this thesis suppose that, by choosing Water Delivered as the desired output, the collection and treatment of water is more conventionally considered more important than its distribution, Equivalent Properties Served is chosen as the wastewater output to reflect the conventional notion that the collection of sewage from customers is more important to model than its consequent treatment and disposal, and that it correctly measures sewage load for customers, although in both cases, as papers such as Erbetta & Cave (2007) suggest, inclusion of all outputs best describes the complete production of the WaSCs.

Other Outputs

Other papers in the literature have chosen different outputs, or outcome variables, though in most cases these choices are for empirical models that differ from what this thesis is concerned with: Mazioitis et. al. (2014) model a Profit

Change Index, and in doing so use Economic Profits as an output; in Productivity Growth and Meta-Frontier Productivity indices respectively, Mazioitis et. al. (2016) and Mocholi-Arce et. al. (2020) employ Total Factor Productivity and Total Expenditures as outcomes of their models; in those parametric models of the industry, as in Ashton (2003), Bottasso & Conti (2009a) and Lynk (1993), Total Costs are used as the dependent variable, with the production outputs above being used in some combination as explanatory factors instead.

There also exists other research that includes another type of output - Undesirable Outputs; as the name suggests, these are modelled such that they are minimised when possible: Sala-Garrido et. al. (2021), in their Stochastic Frontier model that assesses the marginal costs of reducing Greenhouse Gases, uses those emissions as an additional, undesirable output; in some recent papers that have addressed the addition of the Retail sub-section of the industry (Molinos-Senante et. al. (2015b,2016,2017b)), measures such Total Complaints - which will later in this chapter be treated as a measure of retail Quality - as well as Total Unplanned Water Supply Interruptions and Properties below a Reference Water Pressure, are also treated as undesirable outputs. Brea-Solis et. al. (2017) utilise Water Losses in a similar fashion.

4.4.2 Inputs

The Inputs common to research in the industry are such that they best reflect what goes into companies' production processes. Practically, these inputs should be chosen to best reflect Total Costs of companies in a given year. Ofwat's definitions used in the industry handily categorise the parts of total costs in a way that also bears similarity to the expected inputs of production in theory: Total costs are split into Capital Expenditures (Capex), Operational Expenditures (Opex), and Other Costs, which are similar to the conventional Capital and Labour inputs of producer theory.

Many models in the literature adhere to this decomposition, and this thesis follows this description of the inputs, but with a slight distinction; in more recent years, Ofwat has shifted the definitions of Capex and Opex slightly, into Base Capex and Base Opex, which together form Base Total Costs, or Botex, which is not equivalent to Total Costs, as it relegates the enhancement costs of capital and operations to the miscellany of Other Costs. This thesis prefers the use of the Base inputs when the choice arises. These definitions are meant to be proxies of the definitions required for Total Costs in theory: $TC = rK + wL$, the sum of the values of capital and labour respectively. As will be noted later in the section, expenditure costs may not reflect the exact definitions used in this theoretical relation, but will be considered on the basis of their regulatory relevance.

A final note on the inputs is the need for the deflation of monetary inputs. Some research uses input data in terms of costs, rather than physical quantities. In these cases, a price deflator is typically used: historically, the Retail Price Index (RPI) was used, but the industry has recently moved to using the Household Consumer Price Index (CPI-H), which better reflects the customer base.

Capital

Capital is often measured in one of two ways: through Monetary or Physical units. Each method is estimated with a different value, with the monetary capital stock best represented by the Regulatory Capital Value (RCV), which measures the the amount of financial capital employed by companies, and the physical capital stock best represented by the Modern Equivalent Asset (MEA), which measures the replacement cost of tangible fixed assets for companies. Saal & Reid (2005) disaggregate this stock into Water Capital Stock and Sewerage Capital Stock, as to see if the different capital required for the water and sewerage processes affect Opex productivity growth in WaSCs.

Though the MEA method of estimating capital stock is no longer recorded in industry annual reports, it is the most prevalent method in the research literature. Taking Erbetta & Cave (2007)'s definition as an example, Capex is defined to be proportional to the stock of capital, which is determined by the annual consumption of capital defined as the product of the monetary value of capital and a depreciation rate. The consumption of capital is defined by the estimation of the MEA as defined above.

Other interpretations of Capital can be found through looking at the monetary values of Capex or Base Capex, which account for the reported expenditures on short-term capital projects, excluding expenses for enhancing that capital in the base Capex case. A further definition of capital, though notably much rougher in its definition, is to define physical capital as the sum of company infrastructures - i.e. the sum length of Water Mains and Sewers (Bottasso et. al. (2011)). This last definition can work as a proxy for capital, but has the shortfall of not accounting for the heterogeneities in capital stock between companies, which would be addressed, for example, by differences in Base Capex costs and therefore any Enhancement Capex costs.

In this thesis, where base capex is not used to proxy the total value of capital spending by a company, capital stock will be estimated via the less traditional Bottasso et. al. (2011) approach, primarily because of the decommissioning of the MEA method in more recent regulatory cycles.

Labour

The most typical Labour Input employed in the literature is the Labour Costs; this variable represents longer-term investments into operations, such as the cost of Full-Time Employees, and sees more use in the earlier literature, such as in Erbetta & Cave (2007), Hunt & Lynk (1995) and Lynk (1993). The physical equivalent for this cost is the Number of Full-Time Employees, which is used in papers such as Saal et. al. (2007) and Sala-Garrido et. al. (2021).

As with Capital, alternative monetary representations of Labour come in the form of Opex and Base Opex, which represent a companies operational costs, excluding enhancements to those factors is the base Opex case. Opex also sees some use in the literature, such as in Mocholi-Arce et. al. (2021), Molinos-Senante et. al. (2015a, 2019), and Saal & Parker (2000, 2005), to name a few. It is important to note however, that labour costs and operating expenditures are not equal - opex does not only refer to manpower costs, as labour costs does. In this thesis, as in the literature, opex is chosen because it can proxy as a labour input, and is well-defined as a part of total expenditure, as with capex.

Other Inputs

Other Costs are typically defined as the remainder of Total Costs less Capital and Labour Costs. This variable includes the miscellaneous costs of a company, such as Third Party Rates, the cost of Energy or Materials, and in the case of the use of Base Capex/Opex, also includes Enhancement Costs. In this thesis, Other Costs is a catch-all input, found as the remainder of total expenditure after subtracting the other model inputs, base capex and base opex, from the total. Hence, it capture all of the miscellaneous costs of companies, and enhancement costs which have, by definition, been taken away from both capex and opex.

This variable isn't too standout as an input, and is used widely in those papers which apply models using the separate inputs of Capital and Labour, such as in Bottassi & Conti (2009), Erbetta & Cave (2007), Mazioitis et. al. (2015), or Saal & Parker (2000), to name but a few. One other input recently considered is Water Losses, as in Brea-Solis et. al. (2017). Therein, they identify that the losses caused by leakages in the water system can be used as an input factor for water delivered.

This section refers not only to miscellaneous inputs, however; one other factor to consider here relates back to this chapter's discussion on Dimensionality. As was concluded, the objective of the choice of data should be such that the remaining variables are as few as possible, but also are as representative of the industry process as possible. Where this relates to the model inputs is the use of separate or merged inputs - whether or not Total Costs is a sensible input on its own.

Recent econometric models by Ofwat have moved from separate Capex and Opex models to unified Totex models, with yet more recent models instead looking at Botex, the Base Total Costs. Were either of the latter cost specifications used, this would reflect the recent attention that Ofwat has given to cost modelling, and Totex in particular has already seen some use in parametric Cost Functions, but as an output that depends on the separated inputs (Ashton(2003), Bottasso & Conti (2009), Lynk (1993)).

The question that remains is whether a merged, singular total cost input, in the context of the forthcoming parametric models, is sufficiently representative as the industry, compared to the more detailed, separate production factors. There is merit in assessing models with a singular, combined input, primarily because of the consequent ability to increase discriminatory power by abating some of the small-sample problems that inhibit DEA-type models. As a form of robustness test, models containing both merged and separate inputs can be built, to see if the results are significantly different, although this is not carried out in this thesis.

One other consideration around the employment of Totex as a representation of total costs is the difference between this measure of costs, and the theoretical definition of costs. Total economic costs and total expenditures may not necessarily equate, and so it is important to note that, though totex is chosen as it better reflects the industry's consideration of total costs, it may not be exactly theoretically accurate.

4.4.3 Quality Variables

Critical to the contributions of this thesis to the research area are the changes in perception surrounding the use of Quality data. The previous literature hasn't avoided the incorporation of quality into its model, as the adjustment to model outputs with quality was introduced and subsequently adopted as a typical adjustment in the literature by Saal & Parker (2000), but this section will discuss the use of newer quality data, in both more recent research and the forthcoming contributing chapters.

Previous Quality Measures

The initial measures of quality referred to indices that measured the average compliance for River Water Quality and Bathing Water Quality for wastewater quality, defined as:

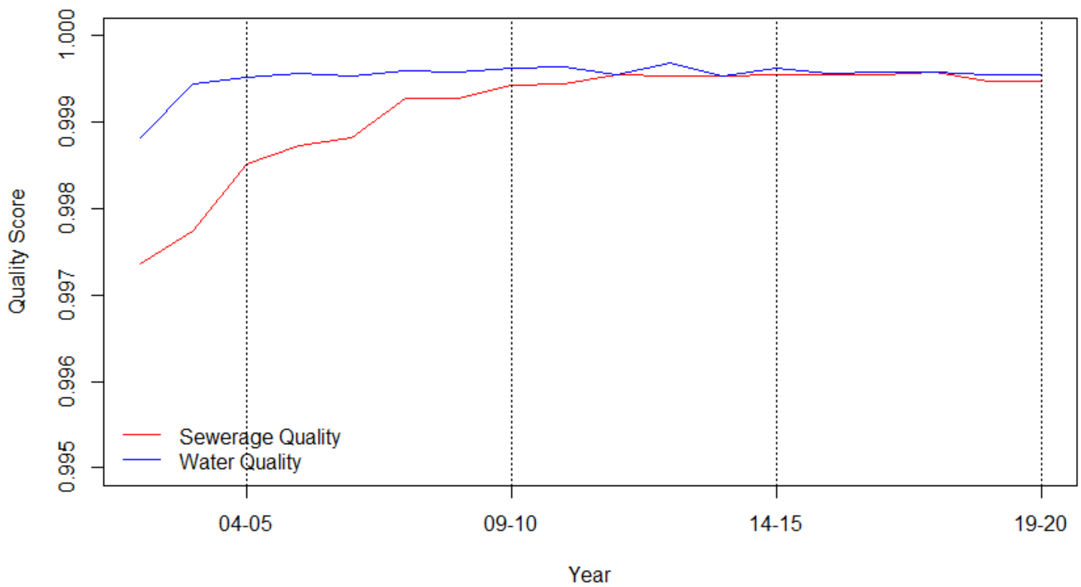
$$Q_S = \left(\frac{S_{River}}{S_{River} + S_{Bathing}}\right) River_Q + \left(\frac{S_{Bathing}}{S_{River} + S_{Bathing}}\right) Bathing_Q \quad (4.2)$$

Where each facet of the index is weighted by the Shares, S of a company's identified River Lengths or volume of Bathing Water relative to the industry total. Water Quality defined by the average rate of compliance of a selection of appropriate measures of water quality, such as various chemical compliances.

Currently, the water quality, Q_W can be collected from archived Drinking Water Inspectorate websites as an average compliance of various quality measures, as before, with the final 2019/20 entry equalling that of the year prior, owing to the DWIs discontinuation of the measurements; wastewater quality, Q_S, can be defined as the the percentage of total sewerage that has not received Secondary Treatment, or equivalently the percentage that has only received Primary Treatment. Figure 4.2 illustrates the average compliances for both measures.

As can be seen, these measures are almost uniformly met, with average compliance rates of 99.9% for both measures from around 2009/10, with that same level of compliance met on average for water quality for all but the first year of the data window. As was mentioned as a precedent for this project, these older measures are, graphically at least, 'stagnant', as adjustments of the industry outputs would be minimal, as their values on average approach 1. The following sub-section, then, looks into alternative measures of quality, so that

Figure 4.2: WaSC Average Quality Compliances, 2002/03 - 2019/20



more scope for quality improvement can be introduced.

Outcome Delivery Incentives

More recent regulatory periods, starting with PR14, have introduced to the industry a collection of incentivised commitments for the purpose of improving various facets of quality throughout industry operations, known as Outcome Delivery Incentives (ODIs); these commitments either provide a monetary reward for their satisfaction by a WaSC, or far more commonly give a penalty should a company fail to reach these targets.

Many of these incentives are unique to WaSCs, who can determine targets related to specific quality issues in their operations, but there are also a subset of these targets, the Common Performance Commitments (CPCs), which are required of all companies in the industry. Table 4.2 lists these commitments as of PR19, and their selection as so-called ‘new’ quality data in the forthcoming research.

Of the 14 CPCs, very few are carried forward as chosen quality data, for various reasons: C-Mex and D-MeX, which reflect measures of household and non-

Table 4.2: Quality Data Selection - Common Performance Commitments

CPC Name:	Used as Data?
Customer Experience Measure (C-MeX)	No
Developer Services Experience Measure (D-MeX)	No
Water Quality Compliance	No
Customer Water Supply Interruptions*	No
Leakage	**Yes**
Per Capita Consumption (of Water)	No
Customer Property Internal Sewer Flooding*	No
Wastewater Pollution Incidents	**Yes**
Risk of Severe Restrictions in a Drought	No
Risk of Sewer Flooding in a Storm	No
Mains Bursts	No
Unplanned Outages	No
Sewer Collapses	No
Treatment Works Compliance	No

* Commitments were considered for addition into the Composite Indicator, but were removed partly due to small-sample issues associated with the Weighting methods.

household experiences with companies, are new as of PR14, and so only cover a minority of the data window periods chosen for the project; Severe Drought Risk and Flooding Risk both cover investing against extreme circumstances due to the weather, and so aren't representative of typical quality deficiencies - similarly, Mains Bursts, Unplanned Outages and Sewer Collapses, though related to the water and wastewater operations of the industry, also reflect extreme cases of quality deficiency, and so don't appropriately reflect more consistent measure of quality over time; per Capita Consumption doesn't reflect either environmental or service quality, and instead reflects something akin to an investment into the customers, by education as to reduce the amount of water used, which is not an aspect of quality for this thesis. That so many CPCs are available from 2014 is also representative of a lack of monitored long-term investment.

Of the remaining commitments, both Water Quality Compliance and Treatment Quality Compliance could be used as data, as they would appropriately reflect environmental quality for both water and wastewater services respectively, but suffer the same issues as the older measures of quality, in that they are likely to be near-uniformly met by all companies in all years of the data window.

What remains of the CPCs are Leakage, Pollution Incidents, Supply Interruptions and Internal Flooding Incidents, all of which appear to be viable in theory. However, the latter two variables are also dropped: Supply Interruptions suffers an issue that comes as a consequence of frequently updated regulation, as its measurement has significantly changed over time, from a number of incidents to an average time of interruption - rather than approximate one measure through converting the other, which would lead to worst-case over-estimations of the interruptions, this variable not used; Internal Flooding Incidents, despite being consistently measured, reflects issues at the household level, which is not a perspective being used in this project. Furthermore, given the forthcoming inclusion of a service quality variable, the restriction owed to the small sample of DMUs each year means that, were either of these variables still sensible, they would not be included to avoid the aforementioned issues related to dimensionality.

So, the CPCs being used as quality variables are Leakage, which reflects losses of water in the delivery of water services, and Pollution Incidents, which similarly reflect instances of pollutions caused by wastewater operations. This data is believed to sufficiently reflect environmental quality issues for both major sections of the industry, and so are 'newer' analogues for water-specific and wastewater-specific quality variables, as the older quality variables were previously.

One thing to note about the use of Leakage as a quality variable is its relation to Water Delivered, which is a production output in the forthcoming DEA models. These variables have a direct negative relationship, in that water delivered is net of water lost due to leakages. This thesis uses both variables in their models, due to the nature of the forthcoming use of leakage as part of a composite indicator of quality - in doing so, the thesis creates an index composed in part of leakage data, which has been aggregated with other quality data, and so has much of its direct relationship removed.

Service Quality

Lastly, this project wants to incorporate some representation of the Retail part of the industry, by way of addressing investments into Service Quality, which reflects improvements in customer services. Some recent papers have been adding such features, such as Molinos-Senante et. al. (2015b, 2016, 2017b), who treat Total Written Complaints, Total Unplanned Interruptions greater than 12 hours, and the number of Properties under a Reference Water Pressure as detrimental outputs, in conjunction with the typical, good industry outputs as described above. Saal & Reid (2005) also consider the reference pressure data in their model of Opex productivity growth, which is used as a hedonic variable to control for a particular opex cost related to improving water pressure.

As appealing as all of these measure are, there is once again an issue of having a small sample in each year, which is circumvented in the above papers by either only considering water services, and so including WOCs, or using a regression model of costs that doesn't the same dimensionality issues as non-parametric DEA models, respectively. So, this project chooses as its last quality variable Total Written Complaints, as it best reflects a general measure of customer service. Reference pressure is useful in a cost function model, as Saal & Reid (2005) estimate, and in the case of a water-only model, as in the other papers, but is not reflective of the entire industry in those models that look only at WaSCs; Unplanned Interruptions faces similar problems to the aforementioned CPC, and has issues of changing measurements throughout this project's data window, which were not an issue in Molinos-Senante et. al. (2015b, 2016, 2017b)'s models, as their data covers the 2001 - 2008 period which uses a singular measurement of this factor.

One final alternative that was considered for service quality was the SIM Score measurement, which is a composite score that combines quarterly customer survey scores with other quantitative measures of service quality to form a reflection of customer service quality from multiple points of view. As representative as this measure would be for service quality, it too suffers from both changes in measurement over time and an incomplete sample within this project's data window, and so it too is not used hereafter.

Further quality variables could be employed if, for example, this project's time period was significantly shortened to a window where various quality factors had consistent definitions and measurement. However, as will be discussed in Chapter 5, it is also the case that, because of the empirical sensitivities to using too many variables in DEA models, the amount of quality variables used hereafter is relatively strict - other models, such as parametric specifications, might benefit from more freedom in this respect.

4.4.4 Non-Discretionary Variables

Many of the models used in the literature consider a variety of factors used as controls. This section covers a particular subset of these controls, which account primarily for heterogeneities between the WaSCs due to difference in their operational environments. Other variables that don't fall into this category will be looked into in a later section, as to cover more of what the literature has used as data beyond what is being chosen for this thesis.

This thesis has a few ways to categorise its set of models in the forthcoming contributing chapters, and one way to stratify these is to split them between One-Stage and Three-Stage models, the latter of which containing a second stage in which the models' Inputs are adjusted by a set of Non-Discretionary variables, which are the very set of environmental variables that were just introduced. This project doesn't innovate especially much in this regard, and so all of the non-discretionary variables covered in this section are chosen because of the precedent set by Pointon & Matthews (2016), who use the same three-stage procedure.

This project has four non-discretionary variables: Water Density, Wastewater Density, the Proportion of Distribution Input collected from Rivers, and the Proportion of Trade Effluent. This selection comes directly from Pointon & Matthews (2016), who employ the same non-discretionary variables, less Leakage, which has been defined in this thesis as a quality variable. The first two variables refer to Population Density for both the water and wastewater services, as differences in populations will lead to different expenses owing to the ease by which services can reach all of their customers. By incorporating these variables as company heterogeneities, the companies can be compared in the efficiency models with an adjustment that effectively makes all DMUs produce in an environment with the same population densities, removing any advantages or disadvantages related to difference in the populations of the areas covered.

The second pair of variables account for other facets of the operational environments of companies. The Proportion of Distribution Input taken from Rivers reflects the fact that each method of abstracting water from sources offers different costs, and by using one such measure of abstraction, reflects the geographical make-up of the WaSC operating environments by accounting for the presence of Rivers, as opposed to the use of Boreholes or Reservoirs for water abstraction. The Proportion of Trade Effluent reflects the difference in the levels of industrial waste services by companies - it might, for example, be expected that regions of England and Wales with a greater industrial presence produce more Trade Effluent, compared to those regions that are primarily agricultural or service-industrial instead.

As with the population density variables, the four variables used hereafter should, in effect, lead to each company's inputs being treated as though they are drawn from the same operating environment, once the adjustment has been made. Since the method by which these adjustments are made is a regression model and are part of the three-stage bootstrapping methodology, there are no issues with dimensionality as with the other, non-parametric models. As a final point of order, Pointon & Matthews (2016) also incorporate Leakage as a non-discretionary variable, to reflect differences in infrastructure and consequent water loss; this is not the case in this thesis, as it is instead used as a Quality variable to represent the same issues as a measure of water service quality.

4.4.5 Prices

Prices are useful to incorporate in models of Costs, as well as in models that use physical representations of factor inputs, and dynamic models that can look at price contributions to productivity growth or efficiency.

Many papers have used factor prices for these types of model: Ashton (2000, 2003), Bottassi & Conti (2003, 2009), Molinos-Senante & Mazioitis (2018), Saal & Reid (2005) and Sala-Garrido et. al. (2021) are some of the examples of Cost Function estimations that incorporate factor prices, which then allow for price elasticity estimates and estimates of the marginal impacts prices have on Total Costs; Mazioitis et. al. (2014, 2015) and Molinos-Senante et. al. (2019) use prices in their models of Profit Change and Profit Growth Indices; Mazioitis et. al. (2016) use prices in their Productivity Growth Index model, and Molinos-Senante & Mazioitis (2020) use prices as a facet of their Total Factor Productivity decomposition model.

Typically, the prices of inputs are calculated from the costs and estimated physical stocks of the input. The Price of Labour, for instance, is found as the Total Costs of Labour divided by the Total Number of Employees. Other Costs are found analogously, regardless of the choice by which its physical stock was determined: the Price of Other Costs is defined as the Total Other Costs divided by its physical stock. Capital, despite its stock estimate often requiring the use of MEA estimations and depreciation, is calculated much the same: the Price of Capital is the Total Capital Costs divided its physical stock estimate. As with the monetary costs when they are used as inputs in empirical modelling, prices too are deflated by a relevant price index - the current choice, as with monetary Opex, Capex, Totex and Botex, is to deflate the prices by the CPI-H.

This thesis will opt to use prices according the CPI-H deflation used in the industry, and will calculate prices via input estimates and total input expenditures. Therein, the prices of the various labour and capital types that aggregate to the final factors of production, are not considered separately.

4.4.6 Other Variables

Though all of the data expected to have use in the following contributing chapters have been defined throughout the discussions of the previous sections, it is worth briefly covering some of the other options used within the literature that were not defined into the categories above. Many of the following variables could certainly provide elucidation towards facets of industry behaviour that this thesis doesn't cover, but one common theme that dictates their exclusion is that they are used in regression-type models, which are not what the forthcoming research is choosing to use as its empirical models.

Time Trends were used in Ashton (2003), Erbetta & Cave (2007), Hunt & Lynk (1995) and Lynk (1993), to account for changes in the estimated Cost behaviour due to dynamic effects, with use in one-stage translog cost functions, or a two-stage adjustment of DEA scores in Erbetta & Cave (2007); Regulatory Dummies function similarly, as seen in Bottassi & Conti (2003) and Erbetta & Cave (2007), and measure changes in the outcome variable captured specifically by shifts between Price Review periods; Hunt & Lynk (1995) and Lynk (1993) also consider DMU Dummies, which capture effects to the outcome variables due to company heterogeneities.

Saal et. al. (2007) use, in their model of Productivity Growth, controls such as the Ratio of Trade Effluent to Resident Population, Bathing Water Intensity, and the Proportion of Metered Properties, which are used as exogenous factors that represent operating characteristics of the firms; in similar fashion, factors such as Access to Drinking Water (Molinos-Senante & Mazioitis (2018, 2019)), Access to Sewerage Services (Molinos-Senante & Mazioitis (2019)), Length of

Mains (Bottasso & Conti (2003), Cubbin & Tzanidakis (1998), Pointon & Matthews (2016), Thanassoulis (2000)) and Statutory Area Size (Bottasso & Conti (2009), Brea-Solis et. al. (2017), Molinos-Senante & Mazioitis (2018)) can also be used to account for exogenous company-level heterogeneities, with the Length of Mains used in Pointon & Matthews (2016) as a component of the second-stage water population density variable.

Lastly, Brea-Solis et. al. (2017) considered a selection of other control variables in their Bayesian model for water loss reduction, such as the Proportion of DI sourced from Underground Sources, the Proportion of DI sourced from Reservoirs, Average Pumping Head and Mains Bursts, as well as Leakage and Proportion of DI sourced from Rivers, which have been referred to elsewhere as a quality and non-discretionary variable, respectively. The sourced distribution inputs cover near-exhaustively the abstraction stage of the water production process, while the Average Pumping Head and Mains Bursts are used as operating characteristics.

4.5 Declaration of Software

A brief section of the thesis must, for posterity's sake, address what software will be used throughout the thesis, specifically as it pertains to the coding of the research in the following chapters.

Data was compiled using Microsoft Excel, and then inputted into RStudio, a software based in R, which serves as the coding software in this thesis. The thesis is written using TeXStudio, a version of LaTeX writing software, which then outputs a PDF file of the document. A list of the coding packages used in RStudio, as well as their uses, are found in Appendix B.

4.6 Conclusion

All in all, there are many considerations to take care with when deciding exactly what data is most appropriate in this area of research, though the sentiment surely extends to any area. In this industry, not only is the best choice of variables a primary concern, but so is the importance of remedying any consequences of choosing more variables than can be managed empirically, a matter most pressing in non-parametric models, which this thesis will employ. To refer to the analogous thesis chapter of Pointon (2014), the issue in sum is one of 'Garbage In, Garbage Out': if data is inappropriate, insufficient, or causes inaccuracies, then the resulting research outcomes will be damaged, perhaps to the point of irrelevance.

This chapter has aimed to cover what choices were made in the literature, and then defends what choice this thesis deems as best for the project, as well as any techniques used to mitigate the near-certain issues of small-sample bias. Going forward into the contributing chapters, any chapter-specific data concerns will be discussed when relevant, though the for the most part the choices discussed here remain constant throughout all three of the forthcoming research contributions.

Part II

Contributing Chapters

This part of the thesis contains the original research for the purpose of answering Chapter 1's research questions.

Chapter 5 developed the crux of the thesis' contributions to the field, by developing a novel Composite Indicator of Quality. The chapter then explores how the indicator, used as an output, affects technical efficiency. Chapter 6 continues this application by examining whether the indicator affects Capex bias, via allocative efficiency in dynamic models that use quasi-fixed capital. Finally, Chapter 7 explores other extensions of the indicator, looking at its dynamic properties, and how the indicator interacts with measures of extreme weather.

Chapter 5

Topic 1: A Composite Quality Indicator

Quality, as discussed, has in recent years converged to a near-uniform standard, as measured by the water and sewerage quality indices most commonly used in the literature. Though recent papers have begun to steer away from this convention, and others have also decried the older quality measurements as stagnant, there is yet to be an industry-wide consideration of quality as a positive factor of production - a point of investment, and therefore a point of concern.

Ofwat is very much interested in the adherence to new quality standards. As Chapter 4 describes, many of the new quality objectives are useful to promote improvements in industry standard and fight against the so-called stagnation in the industry's quality improvements. Yet many of these aspects are under-utilised if used at all in analysis. Most of the newer research around service quality, for example, uses retail-oriented factors such as complaints data (Molinos-Senante et. al. (2015b, 2017b)), and though other research happens to incorporate some of the factors elsewhere, such as using Leakage as a non-discretionary variable to account for operating environment difference, there has yet to be an overall consideration of the newer quality measures that have, in Ofwat's view at least, become points of concern for the companies in the industry.

However, as discussed in Chapter 3, from a methodological point-of-view there are also concerns with trying to use too many factors in the evaluation. DEA requires a reasonable amount of observations for each input or output included, less it yield results that lose comparability or any significant differences at all. The previously defined Cooper's Rule is one such benchmark that wards against using too many factors, but in strictly adhering to this rule, the idea of incorporating multiple quality factors becomes more far-fetched. Bootstrapping procedures, as described earlier, could do away with these concerns, but in addition to having their own resultant biases after use, they too become less impactful if the loss of discriminatory power is great enough.

So, what can be done? Much like previous research, such as Molinos-Senante et. al. (2015b), specific parts of the industry could be used to address particular facets of quality, as their paper did in analysing service quality measures in all companies by only considering the water production sub-sector of the industry. While this produces interesting results without doubt, it does fly against the objective of finding measures for the industry's WaSCs.

One option, novel to the industry and a good amount of microeconomics more generally, is the application of a Composite Indicator. As will be discussed shortly, the advantages of this indicator are multiple: in collecting various inputs to produce a singular result, many facets of quality could be used to give a consequent 'Aggregate' or 'Overall' Quality measurement; in using this indicator as a single factor of production - an output, as will be decided - the issue of the loss of discriminatory power is curtailed somewhat, as there is only an increase in factors by one, rather than by excessively many. Furthermore, where research such as Molinos-Senante et. al. (2015b) was restricted to an approximate half of the total industry, the use of a composite indicator that aggregates factors of quality from over the entire industry means that the following DEA analysis can also be for the whole industry.

The application of a more prominent quality-related production factor also has a basis within current industry regulation. Price caps, as discussed throughout the thesis, are defined via an $RPI + K$ model, which contains a quality-specific improvement factor, in that companies will be set a higher price cap throughout a price review period, if they have in the previous period invested in quality improvement within their services. This, along with the aforementioned CPCs, create a collection of reasons for firms to invest in quality and, therefore, could change total costs, by way of changes in opex and capex investment based on additional needs for quality improvements. In this thesis, this is reflected by the economic trade-off between water and wastewater outputs, and quality improvement: firms will choose to invest their inputs into service production, or into (composite) quality improvement.

Given that the Composite Indicator is to be used hereafter in this thesis, the rest of chapter is laid out as follows: First and foremost, the definition, previous uses in economics and the design of a composite indicator are described, as to show how it is built and where it has already found use. Then, in treating the indicator as a production output, DEA models are drawn to compare a model without quality, with the old measures of quality, and with this new novel indicator of quality. Various other tests and analyses are also performed to best compare this new measurement against its predecessors, to determine if there scope to utilise it an interpretation of quality that is not stagnant nor stagnating any time soon.

The application of this indicator in DEA also presents a modelling choice - in all models hereafter, quality will be considered an Output. This conforms with recent literature such as Molinos-Senante et. al. (2015b), and presents the choice of quality as an economic choice in the firms' production of industry services. Rather than scaling existing outputs as the older quality measures did, the application of quality as an output hopes to demonstrate the degree to which firms are willing to trade-off between the further production of water and wastewater services, and investment into improvements in 'overall' quality.

5.1 Composite Indicators

The most fundamental question to first ask in this chapter would be the question of what exactly a composite indicator is. Formally, a composite indicator is defined as:

$$CI_i = f_{r,s}\left(\{I_{q,i}\}_{q=1}^{Q}, \{w_{s,q}\}_{q=1}^{Q}\right) \tag{5.1}$$

In words, a composite indicator for DMU i is a function $f_{r,s}$, which follows a given Aggregation method r and Weighting method s, of the set of Q Normalised Input Factors $\{I_{q,i}\}_{q=1}^{Q}$ and the set of Q Weights $\{w_{s,q}\}_{q=1}^{Q}$. How each of the normalisation, weighting and aggregation methods are chosen, as well as what factors are used in the indicator at all, are left to the choice of the designers, and evaluation thereafter can test the change in results following specification changes in the indicator, as to see if the design choices are robustly chosen or not.

Another key question to consider is why a composite indicator is preferred to individual quality variables. As has been discussed before, relating to dimensionality issues discussed in Chapter 3 and the previous section, a composite indicator of quality has the advantage of taking multiple facets of quality, and outputting a single measure. Given the shift in research to using additional quality outputs, in this industry where there is a notably small sample (10 WaSCs), minimising small sample issues with a composite indicator is pragmatic, and still yields an interpretable measure of quality, subject to design choices therein.

Much of the discussion in design choices is left to a succeeding section, but the what can be observed before that is how previous indicators have been constructed, as to see what design choices are common, if any, and also where such indicators are employed.

5.1.1 Preceding Literature

Bandura (2008) provides a large survey of pre-existing indicators at the time of their writing, and summarises 178 different composite indicators in their report. One immediate significant point of interest is that all of these indicators are macro-economic in scale, in that all of the indicators are specifically designed to compare either a particular set of countries, or all possible countries in the world with available data.

This highlights the most typical, and indeed the majority use of indicators like these - to compare between a set of entities. Though at a country scale in this survey and therefore in all of the indices surveyed, in principle the indicator is useful to compare an overall measure of some economic factor, and so has a good place in the water and sewerage industry given its historically comparative nature in its regulation and subsequent benchmarking.

To further explore how these indices are built and used, let us take some examples from Bandura (2008)'s survey[1]. The first example is the Corruption Perception Index (CPI) by Transparency International, which is defined in Lambsdorff (2005), which evaluates a selection of measurements of corruption and other aspects of political stability, compiled from a sample of multiple surveys over as many countries as is feasible, and then standardises the results before taking their average to form a final composite indicator.

Translated into the procedural steps that will be used in the design of this chapter's indicator, the Weighting and Aggregation methods are fairly uninteresting, in that the final result is an average of the standardised factors, which therefore implies that the factors are equally weighted. What is interesting in this case is the way the indicator normalises the data from the various inputted surveys, as to scale it to the indicator's historical standard $[0, 10]$ interval scale.

[1]Chapter 3 also covers the theoretical basis for indices in brief.

The transformation used is:

$$I = 10 \int_0^1 \left(\frac{X}{10}\right)^{\alpha-1} \left(\frac{1-X}{10}\right)^{\beta-1} dX \tag{5.2}$$

Where, for each value X, the resulting value I is a standardisation using the Beta Transformation, which uses the initial data and two parameters α and β to best adjust the data to within the scale of the indicator, such that the mean and standard deviation of the final indicator are equal to that of the previous year's indicator for a joint subsample of countries.

A second example is The Economist's Big Mac Index (The Economist (2021)) which calculates a measure of purchasing power parity through the use of the prices of the titular Big Macs in a selection of countries. Referring to the Adjusted Big-Mac Index, which while measuring purchasing power parity also accounts for GDP of the countries observed, the index first normalises the Prices of the Big Macs in each country to a uniform currency - the US Dollar - by dividing each country' price by the US Dollar Exchange Rate. Then, to account for local GDP, a Linear Regression estimates an Adjusted Dollar Price against GDP measured in US Dollars:

$$\hat{P}_i = \hat{\beta}_0 + \hat{\beta}_1 GDP_i$$

Using the new price estimates, a subset of Base Countries is chosen. Though by extension any measurable country on this index could be used to compare parities, the typical bases used are $c = \{USD, EUR, GBP, JPY, CNY\}$. The Adjusted Big-Mac Index is then defined as:

$$\hat{I}_{c,i} = \left(\frac{P_i}{\hat{P}_i} / \frac{P_c}{\hat{P}_c}\right) - 1 \tag{5.3}$$

The Indicator can take any real value, with $\hat{I}_{c,i} \lessgtr 0$ reflecting Under- or Over-Evaluation of the Big Mac relative to the Base Country's currency respectively, and $\hat{I}_{c,i} = 0$ reflecting parity between the observed and base country. Compared to the previous example, this indicator instead focuses on one factor and one particular data set, rather than a collection of surveys, and re-scales the measurement according to a broad measure of country-based heterogeneity via GDP in the chosen uniform currency.

To conclude with the examples of indicators, there is the World Economic Forum's Innovation Capacity Index (Porter & Stern (2001)), which is itself comprised of a collection of other sub-indices measuring aspects of countries' innovation and their consequent capacity for innovation. This index is an equally-weighted sum of its sub-indices, but most interestingly in this example is the method by which each sub-index is chosen, and how each sub-index is then designed.

To find out what factors impact innovation capacity internationally, a Baseline Regression of International Patents is drawn against a collection of other factors and then analysed, with significant factors then contributing as sub-indices. Then, each sub-index is itself a weighted sum of further factors, which are determined via their statistical significance in a appended baseline regression model that further includes factors related to the sub-index in question. The weights of the sum of the sub-index is then determined by the estimated coefficients, which can be illustrated broadly as:

$$I_i^s = \sum_{k=1}^{K} w_{k,i}^s \left(\hat{\delta}_k^s \right) X_{k,i}^s \tag{5.4}$$

Where the sub-index I^s for country i is composed of weights $w_{k,i}^s$ dependent on the regression coefficients $\hat{\delta}_k^s$ of the variable $X_{k,i}^s$. This investigative method to determine the contributors to the final index is an interesting and thorough way of designing the index, and in principle would be a more robust standard on which to build composite indicators, but is contingent on factors commonly available for all observations, and doesn't necessarily hold superiority over other methods, such as expert opinion. All in all, this example serves to show a more data-driven way of building an indicator, but as with other data-based methods might not fully reflect all useful components due to the tools used.

In summary, it seems as though there are already a variety of methods employed for the design of a further variety of indicators. Though the chosen examples here, and by extension all of those indicators surveyed, are not relevant in terms of application to this thesis, the principle takeaway is the scope of the ability to design the a given indicator. One thing to draw from this, in my opinion, is the notion that much of the indicator's design is very much up to the designer - as was mentioned before, but with the necessary caveat the the design choices are suitable at at least face value, and from an academic standpoint are yet more justifiable after robustness checks and other further analysis.

The attraction of this method is in its wide variability in design, though interestingly very recent papers regarding various European utilities (Henriques et. al. (2020), D'Inverno et. al. (2021), Yakymova et. al. (2022)) have common structure to their microeconomic indicators. Indeed, as with much of the thesis overall, the aim of this indicator is to be a first exploratory design in this industry, with the intention to see if its use in DEA modelling leads to significant differences in resulting efficiency scores. The indicator used should certainly meet the criteria of being sensible and justifiable from both a qualitative and analytic sense, but ultimately as will be discussed is not the same as the above examples, in that it is not a definitive final quality assessment for the industry - more exploration and, more importantly, discussion would be required for a satisfactory indicator of that nature, in my opinion at least.

5.1.2 Designing a Composite Indicator

So, the next step in the process is to design the indicator to be used as the novel contribution of this chapter. Much of the point of the previous discussion of indicators was to demonstrate the range of choices available in the design stage, with the idea that choices are done as to adhere to consistency or sensibility in the context of where or how they are used.

To begin with, there ought to be a general guideline to follow, and one such material is the OECD's Handbook on Constructing Composite Indicators (OECD (2008)), which covers a broadly applicable methodology for designing and evaluating composite indicators. Much of the way that this project's indicator is built follows these steps in principle.

Given the general form of the indicator defined by 5.1, there are three broad ordered steps in construction: Normalisation of the data, calculation of the Weights of the factors, and the final Aggregation of the results in to the final indicator score.

Normalisation

Normalisation is first used to convert the raw inputted data into comparable and similarly scaled data, which benefits some weighting methods as will be addressed later, but also can then be used in a comparative exercise of the data's behaviour without being concerned about any differences in scale of measurement in collection - Lambsdorff (2005)' Beta transformation on their data from various surveys is a good example of this part of the process, as it standardised the data of various disparate surveys into values that were all within the same scale. This process allows for the various input factors to be compared, and organises the data for each variable to create a relative scale of quality. Differences in the magnitude of quality are accounted for by the scale conversion, where better-performing firms will have higher normalised values.

A few choice of normalisation method are considered here: Min-Max, Z-Score, or No Normalisation. The first, the Min-Max method, re-scales the data into a the unit interval $[0, 1]$, using the following adjustment:

$$I_{q,i} = \frac{X_{q,i} - \min(X_{q,i})}{\max(X_{q,i}) - \min(X_{q,i})} \tag{5.5}$$

Where the inputted data $X_{q,i}$ is adjusted by the minimum and range of the values of the factor q over all companies i. As a result, the highest performing company would achieve the maximum of the values, and so would normalise to a value of one, and similarly to a value of zero for the worst-performing company. This assumes that the quality data is increasing - i.e., that the maximal value for a factor q is the best outcome.

Another choice is the Z-Score Normalisation, which normalises the data to fit a Standard Normal Distribution, as is common in standard statistical tests of regression models, amongst other things. This transformation is defined as:

$$I_{q,i} = \frac{X_{q,i} - \mu_q}{\sigma_q}, \ \sim N(0, 1) \tag{5.6}$$

Where μ_q and σ_q are the Mean and Standard Deviation of the factor q respectively. Much like the Min-Max method, this method also standardises the data into a specific interval (approximately [-4,4]), but does so by directly using the moments of the data, rather than its extreme values. In a sense, the data is fitted to better match its observed sample distribution, rather than to a uniform scale.

In both cases, however, there needs to be a slight adjustment. As discussed in Chapter 3 and Chapter 4, many DEA methods are Translation Variant, in that changes in the data that still retain the ordinal rank of the data might still give differing results, and the quality data are specifically transformed to all be increasing data, as to assume that increasing each factor is desirable. Where this becomes relevant again is in the results of these normalisation methods. One of the restrictions of the basic DEA models used in this chapter is that they require strictly positive data - which is no longer achieved after either normalisation method discussed is used. To address this, the transformations are adjusted as follows:

$$I_{q,i} = \frac{X_{q,i} - \min(X_{q,i})}{\max(X_{q,i}) - \min(X_{q,i})} + \varepsilon$$

$$I_{q,i} = \frac{X_{q,i} - \mu_q}{\sigma_q},\ I'_{q,i} = I_{q,i} + \max(|I_{q,i}|) + \varepsilon$$

Both methods are adjusted by an arbitrarily small positive value $\varepsilon > 0$, which in the case of the Min-Max method is sufficient to overcome the zero-valued data that the worst-performing DMUs are transformed into. For the Z-Score method, since the scale can be negative, the largest absolute value of the data $I_{q,i}$ is also used with ε to ensure that no normalised data can be negative or zero.

A final method to be considered is the lack thereof - No Normalisation. In this case, the data is simply:

$$I_{q,i} = X_{q,i} \tag{5.7}$$

Since the raw inputted quality data is all positive, no further transformations are required to ensure strict positivity.

For the baseline composite indicator, the Min-Max method is chosen. This method is chosen over the others on the basis of it giving comparable results between the factors, which is not necessarily achieved by the No Normalisation method, while also not requiring particular distributional assumptions, as the Z-Score method does, and only requiring minimal technical adjustments to prevent future difficulties in the indicator's construction. One issue to consider going forward in this design, in relation to normalisation, is that the transformations of data may make nominally irrelevant variables become more significant than they ought to be in practice, because of the re-scaling of all data to a single interval of values, such as the $[0, 1]$ interval

Weighting

The next step is to take the normalised data, and decide upon the weights used to aggregate the data into the final indicator. This step could be considered the most contentious, if weighting methods that require only expert opinion or conscious decisions on the weights are used. To circumvent this problem, minimal executive decisions are made on the requirements of the weights, as there exists methods by which the weights can be determined mathematically and without the designer's interference.

The main family of weighting methods used here are the Benefit of the Doubt (BoD) methods (Cherchye et. al. (2007)), which are non-parametric and happen to be equivalent to DEA models with a singular dummy input. The choice to use these methods is not only to avoid arbitrary weighting choices, but to also provide similarity between this indicator and the forthcoming DEA models. As a consequence, however, the data had to be adjusted in normalisation as was defined above, and in maintaining this design choice modern composite indicator research (Cherchye et. al. (2007a,b), Zhou et. al. (2010), Bernini et. al. (2013)) appear to support the notion that the advantages given by the autonomous weight determination of the BoD methods is greater than the consequential potential changes in results that come from BoD models being DEA models that are translation variant.

The first method uses the Lower Bound BoD method, which contains some additional information about the lower bounds of the weights, based off of the current CPCs, of which a subset is used as quality factors. Since Ofwat requires that all of the CPCs are met by the companies, and that they must also strive to improve other factors of quality, it will be assumed in this method that $w_q, i > w_L = 1/15$, where the lower bound is chosen to mean that the companies in the worst case provide interest in each factor equally, with the total assumed industry factors being all of the CPCs and an additional Retail Service factor.

As the indicator only uses a subset of these quality targets, this lower-bound notion assumes investment in all factors at least exogenously. The weights are then found via:

$$\max_{w}\left(\sum_{q} w_{q,i} I_{q,i}\right),\ s.t.$$
$$\sum_{q} w_{q,i} I_{q,i} \leqslant 1, \forall i, \tag{5.8}$$
$$w_{q,i} \geqslant w_l > 0,\ \forall q, i$$

Where the method seeks to maximise the weighted sum of the normalised data, such that the sum is no greater than one, and all of weights are at least the value of the lower bound, which is strictly positive. In this specification, maximisation is used as the optimising function as it assumed that firms wish to maximise their level of overall quality, defined by the objective function. The weights, then, are the mathematically optimal solutions to this problem, subject to the aforementioned lower bound.

A similar method is the Unbound BoD method, which could be considered the standard model, but relative to the previous method is the same BoD method without the additional lower bound information:

$$\max_{w}\left(\sum_{q} w_{q,i} I_{q,i}\right),\ s.t.$$
$$\sum_{q} w_{q,i} I_{q,i} \leqslant 1, \forall i, \tag{5.9}$$
$$w_{q,i} \geqslant 0,\ \forall q, i$$

Both methods determine the weights of the indicator for each company from the data alone, with first method also using the deliberate lower bound information. In the other extreme, another method considered in this step is that all of the factors are Equally Weighted, for all companies. This gives the following:

$$w_{q,i} = \frac{1}{Q}, \forall i \tag{5.10}$$

Rather than data-driven determination of the weights, this method disallows biasing weight choices by naively assuming that they are equally well-considered by each company.

In a similar sort of vein as the equal weights, but also in line with the other BoD methods, the last weighting method considered in a Common Weights BoD model. This method is considered as it somewhat reflects the fact that, at least in the case of the quality measures in the industry, the targets are regulated and incentivised, and hence there is an interest by all companies to weight each factor significantly. Translating into this weighting model, it is therefore assumed here that the weights of each factor can be commonly determined for the industry from an assessment of each company's preferences according to the BoD model.

To achieve this in a technically similar fashion to the other models, the method of Zohrehbandian et. al. (2010) is used to convert what would have to be Multi-Objective Linear Program (MOLP), as in Kao & Hung (2005), into a single Linear Program is the other methods are by definition. Furthermore, in line with papers such as Bernini et. al. (2013), adjustments are also made to account for the differences between ordinary DEA models - which the other papers are centred around - and the BoD method used for weighting.

Kao & Hung (2005) define the Common Weights DEA model as a minimisation problem of one of a family of Distance Measures:

$$D_p(E(u,v)) = \left(\sum_{i=1}^{N} (E_i^* - E_i(u,v))^p \right)^{\frac{1}{p}} \ , \ p \geqslant 1$$

Which measures the difference between the Ideal Efficiency Score E_i^* and the score determined by the DEA Weights u and v, $E_i(u,v)$ for a DMU i, scaled by a finite parameter p. For the purposes of this indicator, only $p = 1$ is considered, as it reduces the problem to a linear program, and avoids sensitivities in the results due to outlier values.

Translating the problem into the paper's final MOLP, the Common Weights DEA is:

$$\max \left(\sum_{i=1}^{N} \left(\frac{\sum_{r=1}^{s} y_{r,i} u_r}{\sum_{j=1}^{m} x_{j,i} v_i} \right) \right), \; s.t.$$
$$\sum_{r=1}^{s} y_{r,i} u_r - \sum_{j=1}^{m} x_{j,i} v_i \leqslant 0, \; \forall i = 1, \dots, N \tag{5.11}$$
$$u_r, v_j \geqslant \varepsilon > 0, \; \forall r, j$$

Zohrehbandian et. al. (2010) propose to translate this problem further into a single Linear Program, much much like other DEA models. To achieve this, a two-stage procedure is used, wherein a CCR DEA model is first run with the data to produce efficiency scores $\varphi_{CCR,i}$. Then, for the second stage, the data used is adjusted by the CCR scores based on the corresponding orientation of the model. Combining this with the model alterations into a BoD model, the distance measure to be optimised is defined as:

$$D_p(w) = \left(\sum_{i=1}^{N} \left(1 - \sum_{q=1}^{Q} w_q \hat{I}_{q,i} \right)^p \right)^{\frac{1}{p}}, \; p = 1$$

Where $\hat{I}_{q,i} = I_{q,i} / \varphi_{CCR,i}$ are model factors adjusted with efficiencies from an output-oriented CCR DEA model. Ordinarily, the measure minimises the distance between optimal scores and estimated scores. Since the ideal composite indicator score is one, and the aggregated indicators are the objective of the problem, this distance measure is the difference between the two. Therefore, the final Common Weights BoD Method is defined as follows, with $d_i = 1 - \sum_{q=1}^{Q} w_q \hat{I}_{q,i}$:

$$\min \left(\sum_{i=1}^{N} \left(1 - \sum_{q=1}^{Q} w_q \hat{I}_{q,i} \right) \right), \; s.t.$$
$$\sum_{q=1}^{Q} w_q \hat{I}_{q,i} + d_i \leqslant 1, \; \forall i = 1, \dots, N \tag{5.12}$$
$$w_q \geqslant 0, \; \forall q$$

The Weighting method chosen for the base indicator is the so-called Lower

Bound BoD method, owing to it being the most apparently suitable model out of the set, in that the lower bound weights are most easily acceptable, compared to other models such as the assumption of completely equal weights or commonly determined weights. The common weights model is particularly interesting however, as its determination unlike the other methods in the selection are determined over the whole industry, and it will be interesting to see what would as a result be the 'Sector determined' best choice of weights, and by extension quality factors to improve.

Aggregation

The last step in the construction process is to aggregate all of the data with their associated weights from the previous step. Though it could already be a useful tool to assess all of the equally-scaled data or the weights as a way to assess what each company sees as its most important priority, aggregating each factor allows for a singular comparative measure that instead compares the overall average performance of each company over all of the inputted factors.

Broadly, two main types of aggregation are used: Linear and Geometric, since linear aggregations are easily interpretable weighted sums of the factors, and geometric sums can be log-linearised to also create linear sums. The first two methods are exactly these aggregations, with the Linear method defined as:

$$CI_i = \sum_q w^*_{q,i} I_{q,i} \tag{5.13}$$

Where $w^*_{q,i}$ is the qth factor's optimal weight for the ith DMU, as derived by the previous weighting method. Interestingly, by the mechanical virtue of the BoD methods optimising the weights of a weighted sum of the input factors, it must necessarily be that the composite indicator is linearly aggregated.

The Geometric method is defined as:

$$CI_i = \prod_q I_{q,i}^{w^*_{q,i}} \tag{5.14}$$

As Zhou et. al. (2010) explores, and as was mentioned beforehand, more generally BoD weighting methods that optimise with a geometric sum can be considered, by instead optimising the logs of the model as to convert the problem into a linear optimisation problem. Since the results are functionally the same, the geometric indicator can be also be represented by:

$$CI'_i = \sum_q w^*_{q,i} I'_{q,i}$$

With $CI'_i = \ln(CI_i)$ and $I'_{q,i} = \ln(I_{q,i})$ respectively.

A final aggregation method explored in the design of this indicator is also linear, but first compares the data to a pre-determined benchmark, and is the Linear Threshold method:

$$CI_i = \sum_q w^*_{q,i} \text{sgn}\left(\frac{I_{q,i}}{\bar{I}_q} - 1\right) \tag{5.15}$$

This method takes the Sign of the ratio of each factor compared to its sector average $\bar{I}_q$, adjusted to yield negative values of the factor was below average, and positive if it is above average. These signs are then summed with their weights to give the final indicator, as to provide a sense of importance to each factor's relative under- or over-performance for each company. In theory, any useful benchmark could be chosen as a point of comparison: for example, in this indicator any one company could instead be chosen as the comparative benchmark, if there was interest in seeing how companies performed in comparison to that company.

The method chosen for the baseline model is in effect pre-determined by the baseline choice of weights: Linear Aggregation is chosen in part due to its simplicity in interpretation, but also in part by necessity since the weighting model used is a BoD model, and therefore requires the same aggregation as was optimised.

5.1.3 Composite Indicator Analysis

With the baseline indicator decided upon, it stands next to test the robustness of the choices argued in these sections, as to better quantify the justifiability of the choices made. To do so, all of the mentioned methods here will be employed in Uncertainty Analysis, which measures the shift in Ranks of companies' indicator scores due to randomly drawn changes in method in each step of construction, as well as the potential exclusion of one of the input factors. Other models can also be employed, to better understand the behaviour of the composite indicator over time by pooling the data across the industry in each year.

Uncertainty Analysis

As was just described, the purpose of Uncertainty Analysis is to create a measure of how the choices in the design of the indicator, which are at best arbitrarily chosen and somewhat justified, might affect the outcomes of the indicator. It would be relatively easy to, for example, create an indicator such that certain outcomes are more likely or completely pre-determined, or some companies or observations are ranked comparatively better than others by design. As the previous sections aimed to do, the choice of this indicator are hopefully such that the results is relatively easy to understand, and in the case of the most abusable step - the weights - arguably unbiased given the data-driven determinations of the weights over other opinion-based weight selections.

To find a measure of uncertainty, various Input Factors of the analysis are chosen, which aim to test different aspects of the indicator's design that could be potentially biased by the designer. Alongside all of the steps detailed above, the Exclusion of Variables, Inputation of Missing Data, and various other factors could be considered. For this indicator, Table 5.1 lists which factors are used, as well as the different options in each factor that will be considered.

Table 5.1: Uncertainty Analysis Input Factors and Factor Choices

Input Factor:	Factor Choices:
1: Normalisation	Min-Max Normalisation, Z-Score Normalisation, No Normalisation
2: Weighting	BoD Weighting with Lower Bound, BoD Weighting with No Bounds, Equal Weighting, Common Weights
3: Aggregation	Linear Aggregation, Geometric Aggregation, Linear Threshold Aggregation
4: Exclusion of Independent Factors	No Exclusion, Variable 1 Excluded, Variable 2 Excluded, Variable 3 Excluded

To accomplish this analysis, the indicator is Bootstrapped - in this case, $B = 2000$ bootstraps are used. In each iteration, rather than repeating the default indicator's estimation, at each point in the construction where an input factor in the analysis occurs, one of the options are randomly drawn to be used in the indicator instead. In the simplest case, and the case that is used here, the random draw are taken from Uniform distributions, and are mapped to options for the indicator design steps:

$$X_4 = \begin{cases} 1, & if \text{ No Exclusion}, \ with \ \zeta_4 \sim U\left[0, \frac{1}{4}\right) \\ 2, & if \ TotalComplaints \text{ excluded}, \ with \ \zeta_4 \sim U\left[\frac{1}{4}, \frac{1}{2}\right) \\ 3, & if \ PollutionIncidents \text{ excluded}, \ with \ \zeta_4 \sim U\left[\frac{1}{2}, \frac{3}{4}\right) \\ 4, & if \ Leakage \text{ excluded}, \ with \ \zeta_4 \sim \left[\frac{3}{4}, 1\right] \end{cases} \tag{5.16}$$

$$X_1 = \begin{cases} 1, & if \text{ Min-Max Method}, \ with \ \zeta_1 \sim U\left[0, \frac{1}{3}\right) \\ 2, & if \text{ Z-Score Method}, \ with \ \zeta_1 \sim U\left[\frac{1}{3}, \frac{2}{3}\right) \\ 3, & if \text{ No Normalisation}, \ with \ \zeta_1 \sim \left[\frac{2}{3}, 1\right] \end{cases} \tag{5.17}$$

$$X_2 = \begin{cases} 1, & if \text{ BoD Weights with Lower Bound}, \; with \; \zeta_2 \sim U\left[0, \frac{1}{4}\right) \\ 2, & if \text{ Unbound BoD Weights}, \; with \; \zeta_2 \sim U\left[\frac{1}{4}, \frac{1}{2}\right) \\ 3, & if \text{ Equal Weights}, \; with \; \zeta_2 \sim U\left[\frac{1}{2}, \frac{3}{4}\right) \\ 4, & if \text{ Common BoD Weights}, \; with \; \zeta_2 \sim \left[\frac{3}{4}, 1\right] \end{cases} \tag{5.18}$$

$$X_3 = \begin{cases} 1, & if \text{ Linear Aggregation}, \; with \; \zeta_3 \sim U\left[0, \frac{1}{3}\right) \\ 2, & if \text{ Geometric Aggregation}, \; with \; \zeta_3 \sim U\left[\frac{1}{3}, \frac{2}{3}\right) \\ 3, & if \text{ Linear Threshold Aggregation}, \; with \; \zeta_3 \sim \left[\frac{2}{3}, 1\right] \end{cases} \tag{5.19}$$

The Exclusion of Independent Variables, X_4, is defined first due to its placement in the design procedure - prior to the other steps.

There is a small caveat in this selection, owed to the decision to employ BoD weighting methods. As mentioned above, the use of these methods necessitates particular aggregation methods that match what was optimised in the BoD procedure. So, if the Weighting method chosen is not the Equal Weights method - i.e., one of the BoD methods - and the Aggregation method chosen is not the Linear Aggregation, then the aggregation method is re-drawn to be the Linear Method, as to preserve the requirements of the BoD weights.

Once each input factor is drawn, the indicator in each bootstrap iteration is generated, and the Ranks of each company's indicator are taken, and the sum of the difference in ranks between the iteration's ranks and the baseline indicator's ranks are taken over each DMU and averaged. At the end of the bootstrap, the average of all of the average rank changes are taken, to give the final measure of uncertainty defined by the average sum of changes in rank due to methodological differences in the indicator:

$$
\begin{aligned}
Y_b &= \bar{R}_S^b = \frac{1}{N} \sum_{i=1}^{N} \left| Rank_0(CI_i^b) - Rank(CI_i^b) \right| \\
\bar{Y}_b &= \frac{1}{B} \sum_{b=1}^{B} Y_b
\end{aligned}
\tag{5.20}
$$

In effect, the uncertainty measure estimates the average absolute shifts in the indicator ranks due to changes in the indicators specifications. Since the absolute changes are used, a high value of the uncertainty measure would suggest that many companies on average shift ranks due to modelling changes when compared to the initial benchmark indicator.

However, this measure is a little obtuse in clear interpretation. One method of evaluation could be to assume that the outcome is normally distributed, and then a t-Test could be performed with $H_0 : \mu_{Y_b} = 0$, with a test statistic of:

$$
Z = \frac{\bar{Y}_b - \mu_{Y_b}}{\sigma_{Y_b}/\sqrt{N}}
$$

With a statistically significant result suggesting the presence of significant uncertainty in the indicator, which in turn suggests that the outcomes of the indicator have some degree of dependence on how the indicator is constructed. On the other hand, the idea that there is difference in outcomes from design choices is likely to be expected, and so there is difficulty in deciding what arbitrary value of $\bar{Y}_b$ would be considered an acceptable level of uncertainty, and therefore what could be declared as 'too uncertain'.

Similarly, a more qualitative measure could be to consider $\frac{1}{N}\bar{Y}_b$, the average uncertainty for each company. One metric would be to see if this value is less than, say, one, with the implication being that an acceptable level of quality is such that each company on average moves less than one rank due to changes in indicator design. Again, however, the evaluation of the uncertainty in this way is also arbitrary, as there is no particular benchmark past which the indicator is considered to be detrimentally uncertain.

All in all, the uncertainty analysis does serve an interesting purpose in determining whether random changes in design lead to changes in the resultant indicator values, and by extension their ranks. However, there is little in the way of clearly defined benchmarks in this evaluation, and the method only shows if there is in fact uncertainty.

Pooled Composite Indicator Model

Owing to the way in which the indicator is defined, and the DEA models that it is applied to, though the indicator does cover a window of time, it is not Dynamic, is each year's composite indicator scores are found independent of other time periods, and therefore there are no inter-temporal factors in the model. Though this is perfectly fine for this chapter's investigations, further analysis of the indicator could be done by pooling the sector and then assessing how the so-called Sector-wide composite indicator develops over time.

The actual construction of the indicator is the same methodologically speaking, but in each time period the data of all companies is summed. So, the indicator is built with the following steps, with the pooled data defined as $X_q^p = \sum_{i=1}^{N} X_{q,i}$:

Normalisation:

$$I_q^p = \frac{X_q^p - \min(X_q^p)}{\max(X_q^p) - \min(X_q^p)} + \varepsilon,\ \varepsilon > 0$$

Weighting:

$$\max_w \left(\sum_q w_q^p I_q^p \right),\ s.t.\ \sum_q w_q^p I_q^p \leqslant 1,\ w_q^p \geqslant w_l > 0$$

Aggregation:

$$CI^p = \sum_{q=1}^{Q} w_q^{p*} I_q^p$$

By extension, the forthcoming DEA models could also be pooled in this manner, but since the emphasis of this model is specific to the composite indicator, it will not be carried out.

5.2 Data Envelopment Analysis Modelling

With the composite indicator built and somewhat analysed, the second primary goal of this chapter is to see how the use of the indicator as the measurement of quality in a DEA model of the industry compares to previous versions of the models that employ older forms of quality. The objective of these models is to primarily see if there are significant differences between a DEA model with the old quality-adjusted outputs and a model with the composite as an additional quality output alongside the other unadjusted water and sewerage outputs. Much of the explanation of the DEA models used in the applications of the following chapters have been discussed in Chapter 3, and so only a brief explanation will be provided on the mechanics of the models here - the primary points of elaboration are on how the models are used and how they compare to the other models.

5.2.1 DEA Models

To begin, in general all of the models here are Input-Oriented CCR-DEA models (Charnes et. al. (1978)) with Variables Returns-to-Scale, and are estimated for T time periods independently - they are Static DEA models within the data window. The first model, with no quality adjustments or additions whatsoever, is defined as:

$$\begin{aligned}
&\min_{\lambda,\theta} \theta_{i,t},\ s.t. \\
&\theta_{i,t} X_{i_0,t} \geqslant \sum_{i=1}^{N} \lambda_{i,t} X_{i,t},\ \forall i,t \\
&Y_{i_0,t} \leqslant \sum_{i=1}^{N} \lambda_{i,t} Y_{i,t},\ \forall i,t \\
&\sum_{i=1}^{N} \lambda_{i,t} = 1;\ \lambda_{i,t} \geqslant 0,\ \forall i,t
\end{aligned} \tag{5.21}$$

Where Technical Efficiency $\theta_{i,t}$ of DMU i at time t is minimised such that the efficiency-adjusted Input Vector of Reference DMU i_0, $X_{i_0,t}$, exceeds a weighted sum of the inputs of all DMUs in the industry, the Output Vector of reference i_0, $Y_{i_0,t}$, is no greater than the weighted sum of the industry's outputs, and the weights $\lambda_{i,t}$ for each DMU sum to one by the VRS assumption and are non-negative.

The next model is very similar to (5.21), but contains quality-adjusted outputs instead:

$$
\begin{aligned}
& \min_{\lambda,\theta} \theta^*_{i,t}, \ s.t. \\
& \theta^*_{i,t} X_{i_0,t} \geqslant \sum_{i=1}^{N} \lambda_{i,t} X_{i,t}, \ \forall i,t \\
& Y^*_{i_0,t} \leqslant \sum_{i=1}^{N} \lambda_{i,t} Y^*_{i,t}, \ \forall i,t \\
& \sum_{i=1}^{N} \lambda_i = 1; \ \lambda \geqslant 0, \ \forall i,t \\
& y^*_{i,t} = y_{i,t} \cdot Q_{y,t}, \ Y^*_{i,t} = \left[y^*_{i,t} \right], \ \forall i,t
\end{aligned}
\tag{5.22}
$$

The only major difference in the model is the final re-definition of the outputs. For each output $y_{i,t}$, there is an adjustment by the corresponding Quality measurement $Q_{y,t}$ to give $y^*_{i,t}$. The Adjusted Output Vector, then, is a vector of those outputs $y^*_{i,t}$.

The final model incorporates the new composite indicator measurement of quality as an additional output, and is defined as:

$$
\begin{aligned}
& \min_{\lambda,\theta} \hat{\theta}_{i,t}, \ s.t. \\
& \hat{\theta}_{i,t} X_{i_0,t} \geqslant \sum_{i=1}^{N} \lambda_{i,t} X_{i,t}, \ \forall i,t \\
& \hat{Y}_{i_0,t} \leqslant \sum_{i=1}^{N} \lambda_{i,t} \hat{Y}_{i,t}, \ \forall i,t \\
& \sum_{i=1}^{N} \lambda_i = 1; \ \lambda \geqslant 0, \ \forall i,t \\
& \hat{Y}_{i,t} = \left[Y_{i,t}, CI_{i,t} \right], \ \forall i,t
\end{aligned}
\tag{5.23}
$$

The new output vector, $\hat{Y}_{i,t}$, contains the old outputs $Y_{i,t}$ and the indicator $CI_{i,t}$, which is also in this chapter estimated in each time period t independently.

To further develop the comparisons between the models, extension to Three-Stage DEA models (Blank & Valdmanis (2005), Pointon & Matthews (2016)) can also be looked into to further assess the differences in efficiency due to the definition of quality.

The three-stage model used for this additional modelling procedure follows Blank & Valdmanis (2005), who adjust their data by Slack Adjustments estimated from a Stochastic Frontier regression containing operating environment characteristics which, in this industry at least, are quite heterogenous between the companies. This model's second-stage more closely follows Fried et. al. (2002)'s approach to the second-stage adjustment, rather than the more frequently applied OLS regression from Fried et. al. (1999), which has been used in the other applications of the three-stage model aforementioned.

$$\begin{aligned} S_{m,i} &= f^m(z_{j,i};\gamma_{j,i}) + V_{m,i} - U_{m,i},\ \forall m,i, \\ S_{m,i} &= x_{m,i} - x^*_{m,i},\ V_{m,i} \sim N(0,\sigma_V^2),\ U_{m,i} \sim N^+(0,\sigma_U^2) \end{aligned} \tag{5.24}$$

Where f^m is some function - in this case a linear function - and the errors $v_{m,i} \sim N(0,\sigma_v^2)$ and $u_{m,i} \sim N^+(0,\sigma_u^2)$ are an independent White Noise Error and the Technical Inefficiency of the regression measured on the strictly positive Truncated Normal distribution, respectively. The estimated weights, $\hat{x}_{m,i} = \sum_j z_{j,i}\hat{\beta}_{j,i} + \hat{v}_{m,i} - \hat{u}_{m,i}$, are then used to adjust the inputs as follows:

$$\hat{x}^{A*}_{m,i} = \frac{\max_i(x_{m,i}) - \min_i(x_{m,i})}{\max_i(\hat{x}^A_{m,i}) - \min_i(\hat{x}^A_{m,i})}\left(\hat{x}^A_{m,i} - \min_i(\hat{x}^A_{m,i})\right) - \min_i(x_{m,i}),\ \forall m,i$$

With $\hat{w}^A_{m,i} = w_{m,i} - \hat{w}_{m,i}, = w_{m,i} - \sum_j z_{j,i}\hat{\beta}_{j,i} - \hat{v}_{m,i} + \hat{u}_{m,i}$. Finally, the Three-Stage DEA model uses these new adjustments to produce a set of weights which should account for company environmental heterogeneities:

$$\begin{aligned}
&\min_{\lambda,\theta} \hat{\theta}_{i,t}, \; s.t. \\
&\hat{\theta}_{i,t}\hat{X}_{i_0,t} \geqslant \sum_{i=1}^{N} \lambda_{i,t}\hat{X}_{i,t}, \; \forall i,t \\
&\hat{Y}_{i_0,t} \leqslant \sum_{i=1}^{N} \lambda_{i,t}\hat{Y}_{i,t}, \; \forall i,t \\
&\sum_{i=1}^{N} \lambda_i = 1; \; \lambda \geqslant 0, \; \forall i,t
\end{aligned} \tag{5.25}$$

Where $\hat{Y}_{i,t}$ is as before the vector of three outputs containing water output, wastewater output and the composite indicator, and the vector $\hat{X}_{i,t}$ contains the adjusted elements $\hat{x}^{A}_{m,i}$.

It is worth briefly mentioning that, though the BoD method of the composite indicator looks like it can facilitate the same three-stage adjustment as the DEA models, it cannot. The notion of an adjustment in this manner is certainly appealing, and future endeavours might serve to find a new adjustment procedure that accounts for this, but as the BoD model does not have true inputs from any data - 'dummy' inputs are used for mechanical reasons - there cannot be appropriate slacks generated for the second-stage of the procedure. Since slacks are, by definition, deviations from the optimal values of their constituent variables, creating slacks from the weights, which would be the target of adjustment, isn't manageable with the specification above because these weights are the objective of the model in the first place.

5.2.2 Dimensionality Issues and Bootstrapping

As was discussed in Chapter 4, the number of DMUs in a given time period in this application is sufficiently low to cause a lack of discriminatory power between the companies, as many of them may be considered efficient, to the point where the majority of the companies can't be adequately compared to each other.

To address this issue methodologically, bootstrapping techniques are used in the models in much the similar fashion to how they were used in the further exploration of the composite indicator. In the one-stage models (5.21) - (5.23), the Simar & Wilson (1998) algorithm is used to generate a collection of efficiency scores $\hat{\theta}^b_{i,t}$ for $B = 2000$ bootstraps, and the final scores are defined as the average:

$$\bar{\hat{\theta}}_{i,t} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{i,t} \tag{5.26}$$

Where the efficiency scores of the composite indicator DEA are used as an example. In the three-stage models, the first algorithm of Simar & Wilson (2007) is used in the second-stage stochastic frontier regression to produce bootstrapped estimates of the slacks of each input which then adjust the initial input data to produce environment-adjusted efficiency scores in the final stage of the model:

$$\begin{aligned} S^b_{m,i} &= f^q(z_{j,i};\gamma^b_{j,i}) + V^b_{m,i} - U^b_{m,i}, \\ \hat{x}^A_{m,i} &= x_{m,i} - \bar{\hat{S}}_{m,i}, , \\ \bar{\hat{S}}_{m,i} &= \frac{1}{B}\sum_{b=1}^{B}\hat{S}^b_{m,i}, \ \forall m, i \end{aligned} \tag{5.27}$$

Where the third stage of the respective DEA models then follow the same specifications as the un-bootstrapped models above. Much like the one-stage bootstrap, the second-stage bootstrapping algorithm draws random errors in each of the B iterations, which are then used to draw bootstrap estimates of $\hat{\gamma}_{j,i}$ and then $\hat{S}_{m,i}$. The algorithm used for the three-stage model differs, however, and is instead drawn from Simar & Wilson (2007) and Cordero-Ferrera et. al. (2010).

5.3 Data

Much of the discussion about the particulars of the data is in Chapter 4, so with that in mind this section aims only to define what is used in the models of the chapter, rather than repeat the reasoning behind the choice of data. Some descriptive statistics of the data employed in this chapter are listed in Table 5.2. The data window used is from 2002/03 up to 2019/20, and data is collected from the industry's Annual Performance Reviews (APRs), which in the earlier portion of the window were known as the June Annual Returns (JARs). One exception to this is the old Water Quality measure, which was collected external to the companies' reports and was found in archived annual reports from the Drinking Water Inspectorate.

Only the ten largest WaSCs are used in the analysis, as all WOCs by definition do not have sewerage services, and since they face significant difference in technology (Saal & Parker (2005)), they are not comparable. Hafren Dyfrdwy is also not used as an observation here, since its recent conversion to a WaSCs means that it both has a far smaller scale than its other counterparts, but would also be incomparably different in the model because of its relative immaturity in its wastewater services. So, in those years where it is recorded to have sewerage services, those service are merged with its parent company, Severn Trent, as to at least reflect the company's contribution to the industry.

The two Outputs for the DEA models, which excludes the Composite Indicator which acts as a further output, are Water Delivered and Equivalent Population Served for sewerage. Water delivered is defined as the sum of Potable and Non-Potable Water delivered, and the equivalent population served is defined as the Sum of Residential and Non-Residential Equivalent Population Served. Other measures of water and sewerage populations can also be considered for outputs (Erbetta & Cave (2007)), but due to dimensionality issues from the small available sample per year, only two outputs are used.

There are three Inputs in the models also: Base Opex, Base Capex, and Other Costs. Base Opex and Base Capex are respectively the Operating and Capital Expenditures required for to cover all maintenance costs in the industry, and so other costs are calculated as Total Costs less Base Opex and Base Capex, and account for all other miscellaneous costs, such as third-party expenditures,

Table 5.2: Data Descriptive Statistics - 2002/03 - 2019/20

Variable (Units)	Mean	Std. Dev	Min.	Max.
Outputs:				
WaterDelivered (Ml/Day)	1023.61	550.55	270.96	2179.44
EqvPopulation (Pop.(000s))	5591.60	3625.42	1511.62	15771.01
Inputs:				
BaseOpex (£m)	328.13	171.35	90.48	860.68
BaseCapex (£m)	350.68	215.69	86.24	1121.32
OtherCosts (£m)	39.93	46.32	0.69	219.26
Quality Factors:				
Leakage (Ml/Day)	280.21	205.40	61.35	946.04
PollutionIncidents (Nr./1000km)	62.92	47.02	12	289
TotalComplaints (Nr.)	13610.15	11838.08	1467	68874
WaterQuality (%)	0.99953	0.00028	0.9973	0.9999
WastewaterQuality (%)	0.99913	0.00142	0.99186	0.99999
Non-Discretionary Variables:				
PropDIRivers (%)	0.3597	0.2177	0	0.781
WDensity (Pop.(000s)/km)	0.1571	0.0522	0.1032	0.3198
WWDensity (Pop.(000s)/km)	0.1721	0.0211	0.1321	0.2283
PropTradeEffluent (%)	0.0445	0.0299	0.0092	0.1332

as well as the costs declared for the purposes of industry enhancement. This approach aligns with more recent definitions of industry inputs (e.g. Saal & Parker (2006)), but is slightly different in that it considers Base Total Expenditure - Botex - as the sum of Base Capex and Opex, where other recent papers instead use Total Capex and Opex as the representative capital and labour inputs.

One thing to note about this data, as it pertains to the forthcoming results of the chapter, is the difference between its specification, and that of the theoretical costs, inputs and outputs. Base opex and capex, by definition, differ from the respective theoretical definitions of labour and capital, and so they, and other costs, may imply slightly different results and efficiencies for firms, compared to the purely economic definition.

To create the composite indicator, three quality variables are used: Leakage, Pollution Incidents, and Total Complaints. Leakage reflects the quality of water service infrastructure, and similarly the number of pollution incidents reflects wastewater infrastructural quality. Total complaints measures the amount of Written Complaints, and is used to reflect general service quality (Molinos-Senante et. al. (2015)). As with the DEA model data restrictions, there are various other measures of the types of quality considered in the indicator, but many of those factors are prone to changes in measurement throughout the price review periods, and in general only three factors are used as to prevent the indicator from also suffering from dimensionality issues.

For the older measures of quality also used in some of the modelling, Water Quality and Wastewater Quality are used as defined in previous research (Saal & Parker (2000)). Water Quality is defined as an average compliance of a collection of measures drawn by the Drinking Water Inspectorate, with the exception of the 2019/20 entry which, due to the discontinuation of the measure, is extrapolated to be the same value as the year before. Wastewater Quality is defined as the percentage of sewerage load that has not received at least Secondary Treatment, i.e. the percentage of total sewerage that has only received Primary Treatment.

A final note on the use of the quality variables in the composite indicator is that the data is transformed to be increasing. By default, for example, an increase in Total Complaints is assumed to be worse than a decrease, but since DEA-type models assume that increases are preferred, the data must be adjusted accordingly. This takes little effort, as in practice the data's reciprocals are taken for each quality factor which, despite the potential for varying results due to the model's lack of translation variance (Sarkis (2010)), is enough to allow the models above to be used properly. This need for adherence to modelling assumptions is also echoed in the strict positivity of the normalisation methods discussed in the previous section.

In the three-stage models, the second-stage stochastic frontier model uses four Non-Discretionary variables which reflect operating environment conditions: the Proportion of Distribution Input generated by Rivers, Water Density, Wastewater Density, and the Proportion of Trade Effluent. Water and Wastewater Densities are used to reflect differences between urban and rural population density. The proportion of water abstracted from rivers reflects the different costs associated with the method compared to other abstractions such as Boreholes or Reservoirs. Finally, the proportion of trade effluent reflects the costs associated with industrial effluent as a proportion of the total wastewater returned. In previous uses of this model (Pointon & Matthews (2016)), Leakage was also a non-discretionary variable, but since it is already a quality variable in this case, it is not accounted for again in this second-stage regression.

5.4 Results and Discussion

Following the order of the previous methodology section, the first set of results concern the resultant ranks of the base composite indicator specification, listed in Table **??**. Then, the various analyses and additional models and specifications are also illustrated, as in Tables 5.4 and 5.5. Finally, the ranks of the efficiency scores from the DEA models also aforementioned are also listed, with comparisons between the models and the additional three-stage changes.

5.4.1 Composite Indicator Results

Table 5.3: Ranks of DMU CI Scores, 2002/03 - 2019/20 (1=Best, 10=Worst)

DMU	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
ANG:	8	6	6	7	8	9	8	8	10	10	9	10	2	6	7	7	10	10
NWL:	9	9	9	9	7	8	6	6	5	9	6	5	10	8	2	2	3	5
SRN:	1	2	2	1	4	5	4	5	4	3	3	4	5	3	3	3	5	4
SVT:	7	7	8	10	10	10	10	10	9	5	8	8	6	5	8	10	6	9
SWT:	4	4	4	4	3	4	5	3	2	2	4	3	8	4	5	4	2	2
TMS:	3	3	3	3	2	3	2	2	3	6	10	9	3	7	6	8	8	7
UUW:	5	5	5	5	6	6	7	7	7	7	2	2	4	2	4	6	7	6
WSH:	10	10	10	8	9	7	9	9	8	4	5	6	7	10	9	5	4	3
WSX:	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
YKY:	6	8	7	6	5	2	3	4	6	8	7	7	9	9	10	9	9	8

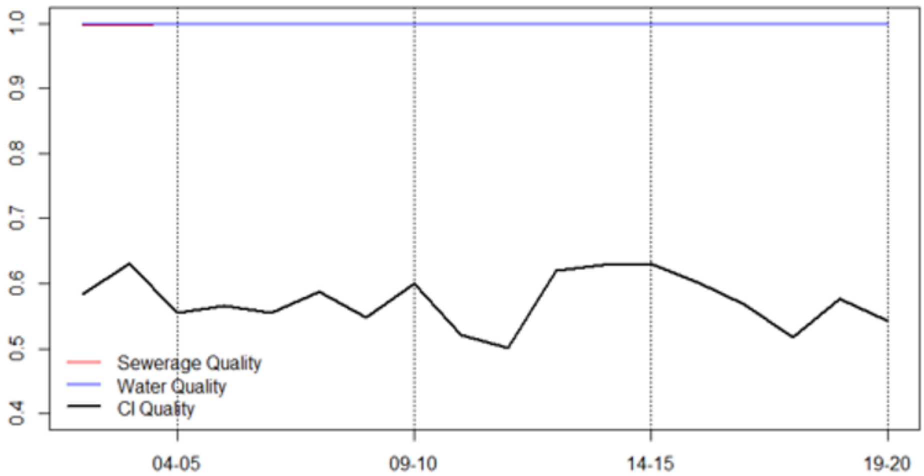
Dashed lines indicate Price Review periods.

DMU CI Scores, 2002/03 - 2019/20

DMU	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
ANG:	0.40	0.66	0.37	0.39	0.28	0.28	0.30	0.31	0.22	0.25	0.40	0.29	0.89	0.55	0.45	0.36	0.32	0.32
NWL:	0.28	0.29	0.25	0.24	0.36	0.36	0.35	0.63	0.48	0.28	0.49	0.68	0.29	0.45	0.90	0.91	0.90	0.52
SRN:	0.94	0.93	0.93	0.93	0.77	0.75	0.72	0.68	0.66	0.72	0.73	0.70	0.74	0.72	0.69	0.67	0.56	0.55
SVT:	0.43	0.50	0.30	0.20	0.05	0.21	0.26	0.24	0.29	0.40	0.46	0.49	0.60	0.59	0.43	0.27	0.45	0.33
SWT:	0.78	0.78	0.76	0.77	0.79	0.76	0.67	0.83	0.77	0.76	0.72	0.72	0.45	0.71	0.58	0.52	0.91	0.88
TMS:	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.76	0.39	0.37	0.37	0.78	0.51	0.46	0.32	0.36	0.37
UUW:	0.59	0.67	0.60	0.67	0.52	0.41	0.34	0.41	0.33	0.35	0.88	0.75	0.74	0.87	0.61	0.39	0.42	0.48
WSH:	0.23	0.21	0.21	0.25	0.27	0.40	0.29	0.30	0.33	0.53	0.66	0.68	0.52	0.29	0.31	0.48	0.58	0.62
WSX:	0.87	0.93	0.93	0.92	0.94	0.92	0.94	0.96	1.00	1.00	1.00	1.00	0.98	1.00	0.98	0.96	0.93	1.00
YKY:	0.45	0.46	0.33	0.43	0.69	0.92	0.75	0.77	0.39	0.33	0.48	0.59	0.32	0.33	0.26	0.30	0.35	0.35

Dashed lines indicate Price Review periods.

Figure 5.1: Comparative Graph of Old and Composite Quality Measures, 2002/03 - 2019/20



The ranks of the composite indicator have various interesting observations that can be taken from them. First, it can be noted by comparing the first and last year's ranks for each company that no WaSC in the industry retains their position overall. However, there are some companies that display some consistent general ranks over time. WSX, for example, are almost always ranked the best in the industry in each year - with only two exceptions - which in turn suggests a consistently high relative level of quality as measured by the indicator.

Other examples tell a similar tale: both ANG and SVT appear to stay relatively low in the ranks, and so suggests relatively low levels of quality. Interestingly, a lot of these results seem to be indicative of differences in scope, which are factors not yet corrected by the three-stage composite indicator model. Indeed, it appears that the smaller companies, in the end, rank higher than their larger counterparts. TMS, YKY, SVT and ANG cover larger parts of England and Wales, and it seems as though they happen to rank the worst in the industry as of 2019/20, compared to the smaller companies of WSX, SWT and WSH, which ultimately rank in the higher end of the industry.

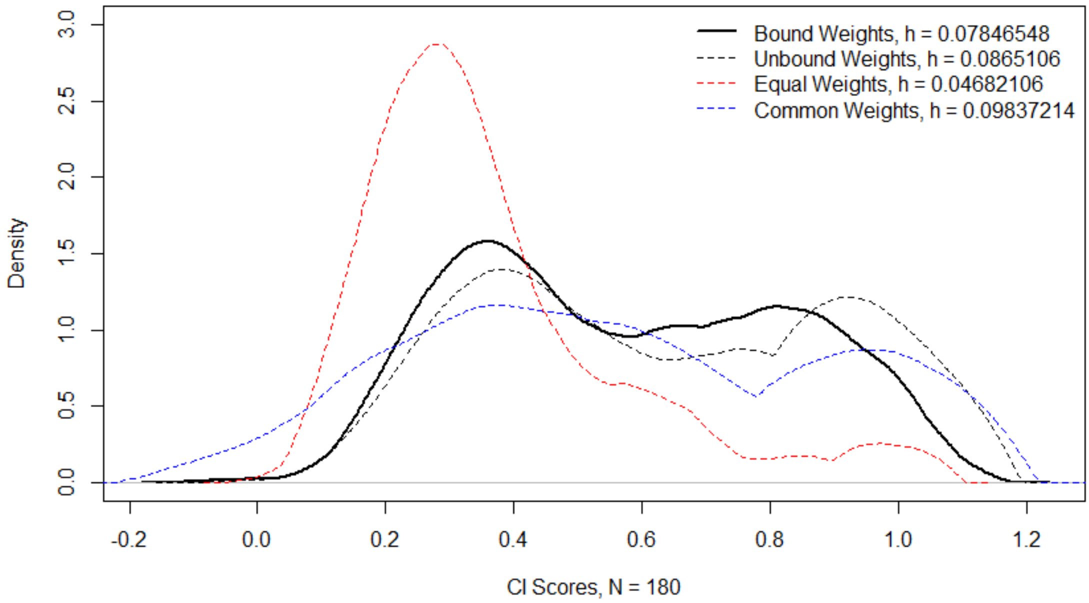
This isn't to say that this is necessarily a consistent behaviour. TMS, for instance, was consistently high in the rankings for the first half of the time window covered. and only seems to face demotion in the latter half of the data, though this isn't always the case, despite their large scope. A similar case is SVT, which is consistently average or below average, but not entirely consistent as WSX is in the upper ranks.

To evaluate one of the auxiliary research questions, WSH provides an interesting point of focus, given their uniquely different structure as a not-for-profit organisation. On this basis, there is reason to expect that the company, given their focus on improving customer experience and service quality, would perform consistently well. However, this is not always the case, according to these results. There is a significant difference from start to finish, but the company seems to experience improvement over time, becoming more consistently better in terms of quality in the latter half of the time frame, with some exceptions. Another case in the opposite direction is YKY, who appears to consistently worsen is time goes on.

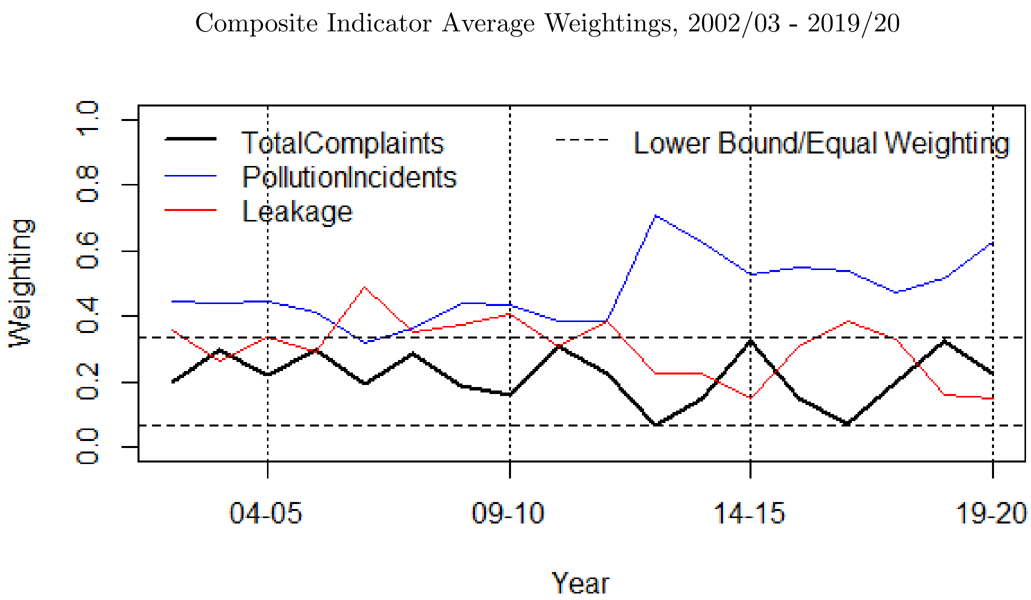
A closing point to consider for these results is to assess the trends in these ranks in each Price Review period, as indicated by the vertical separations on the table. In WSH's case, for example, PR09 and PR14 show evidence of the aforementioned quality improvements, climbing three ranks in both review periods. YKY, on the other hand, notably falls to its ultimate low ranks throughout PR09 specifically, despite some relative improvement in PR04 previously.

Looking at Figure 5.1, the difference in average compliance can be observed for the older measures of quality and the new composite indicator measure. The old quality measures overlay almost exactly in Figure 4.2, but the difference in compliance in terms of scale demonstrates quite how uniform the old measure have become. Compared to the average compliance rates of 99.9% for both the old water and wastewater quality measures, the new composite measure has an average of 57.4%, and demonstrates significant volatility compared to the older measures. While this looks like a significant and stark decrease in quality, from a regulatory standpoint this result shows the scope of the new composite measure as a regulatory target - there is evidence not only that the older measures are stagnant, but that this new measure has a vast scope for industry-wide improvement.

Figure 5.2: Composite Indicator KDE Distributions, 2002/03 - 2019/20



More analysis can be done by looking at the Kernel Density Estimates of the indicator scores. Figure 5.2 plots the estimated distributions for each weight specification of the composite indicator, for all observations in the sample. All KDEs in this figure, and the thesis hereafter, are estimated with no



pre-determined bandwidth and with the Epanechnikov distribution as the basis of the estimates.

The KDEs provide some interesting conclusions: it appears as though, in most specifications, the indicator scores follow a sort of bi-modal distribution, and in all cases, the most likely mode of the distribution is a sub-par score, with most models likely to score around 0.4, and the Equal Weights model's singular mode at around 0.3. Focusing primarily on the Bound Weights specification, as it is the model used in later modelling, it is interesting to observe that this data suggests that companies, in all of the years sampled, are most likely to have below-average overall quality, and are next-most likely to perform quite strongly at around 0.8. These results seem to suggest that there is a somewhat distinct split between observations that perform well, and those that perform poorly, but such an evaluation is spurious given that there is also significant density for the values in-between, and there is no way in this figure to deduce when these modal values occur, and to whom they are attributed.

Accompanying Figure 5.2 is a graph of the industry average weights for each quality factor over time. Therein, we can also observe how, under the lower bound weighting method, the average weights for each quality factor appear to remain broadly the same until 2011-12, after which pollution incidents gains more importance throughout the industry on average. This suggests that, thought total complaints and leakage have remained pertinent industry issues, changes in the industry during PR09 led to more focus on prioritising pollution incidents. In the context of the composite indicator, this also suggests that, mathematically, more firms have relatively good quality scores for reducing pollution incidents.

In summary, at the very least Table 5.1 demonstrates that there is significant variance in quality ranks over time according to this measure. Though the behaviours differ for some, no company retains its position over the whole time period. A lot of the conclusions of these results can only really afford to be given scepticism, however, as it is likely that, much like the efficiency scores of the one-stage models of previous research, the weightings of the indicator and therefore the ranks are influenced by the difference in company operating environments, such as the scale of the companies. Furthermore, as was discussed to justify some of the following analysis, there is no measure yet of how these ranks have been affected by the design of the indicator itself which, despite efforts to best remove personal biases in the design choice, might still lead to results dependent on the indicator model employed. The next results will cover the extent to which this is true.

Composite Indicator Analysis

First in the further analysis is Uncertainty Analysis, whose outcomes and sample statistics are described in Table 5.4. The outcome measure of uncertainty - the Rank Shift - is the sum total amount of ranks shifted by all companies in the industry averaged over the bootstrap iterations.

Looking first at the summary statistics, there appears to be relatively low variance about the ranks shift average, which suggests that there is a consistent amount of uncertainty over time, according to this analysis. Using the

Table 5.4: Uncertainty Analysis Results, B=2000 Bootstrap Simulations

Year:	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
Rank Shift:	4.22	4.14	4.30	4.32	4.29	4.37	4.32	4.40	4.31
Year:	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
Rank Shift:	4.35	4.19	4.18	4.22	4.15	4.41	4.38	4.07	4.20
Mean Rank:	4.26763								
Rank S.D.:	0.09911								

Z-Score metric under the null hypothesis that there is no shift in rank, the score, $Z = (4.26763 - 0)/0.9911 = 43.0595$, is most certainly significant at any sensible significance level, suggesting at the very least that, by this metric, there is significant ranks shifts and therefore significant uncertainty. However, it is worth noting that a Kolmogorov-Smirnov test for Normality in the rank-shift data rejects strongly the notion that this data is normally distributed, which suggests that this statistic, though an interesting thought exercise, is not reliable in the first place; some test statistic against the observed empirical distribution of the data, or a non-parametric equivalent, would be similarly useful for this quantitative point-of-view.

Taking the more qualitative approach of the average shift per company, $4.26763/10 = 0.42676$, and supposing that a rudimentary benchmark is to assume that a value greater than one is a measure of significant uncertainty, it can also be suggested that, though there is some uncertainty, it is perhaps at a reasonable level, given that the average rank shift per company per in each year is about 0.4 - each company on average moves less than one rank in each year.

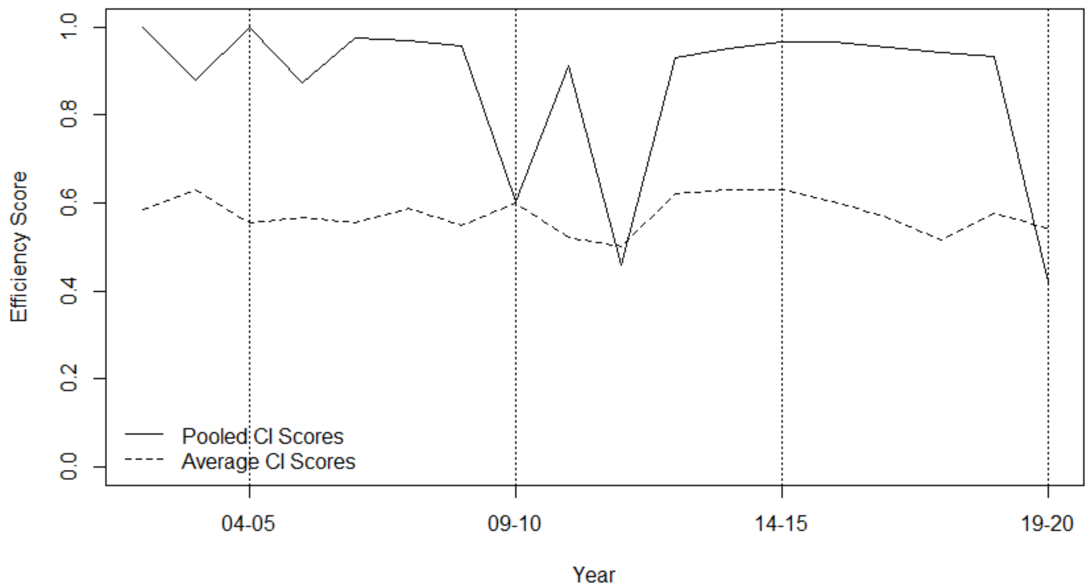
The difference in results doesn't necessarily lead to separate and conflicting conclusions, however. Though the Z-Score interpretation of uncertainty suggests a significant amount of uncertainty, the descriptive measure of average rank shift per company suggests that that amount, though significant statistically, might be below the point of concern - however arbitrary that point is. A value greater than one for this measure, for example, would suggest that in each year on average of all bootstraps, all companies change rank, or certain companies have larger changes in rank. Either case suggests that the ranks would be quite volatile, which is not a factor explained elsewhere in the results.

Pooled CI Results

Table 5.5: Pooled Composite Indicator Model Ranks, 2002/03 - 2019/20

Year:	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
Rank:	1	14	2	15	3	4	7	16	13
Year:	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
Rank:	17	12	9	5	6	8	10	11	18

Figure 5.3: Pooled Composite Indicator Scores, 2002/03 - 2019/20



The Pooled Composite Indicator Ranks of Table 5.5, which are ranked for the pooled scores of each year in the time period, appear to very quite significantly between time periods, with no consistent trend. Observing also Figure 5.3, which compares the plots of each pooled indicator and the average composite indicator score over time, there is clear evidence that the pooled model is far more inconsistent than the default indicator design.

Interpreting the pooled indicator values as the composite indicator, and so the aggregated quality, of the industry in total in a given year, it appears that there is a slightly negative trend in the pooled scores, though this appears to also be insignificant. This seems to be comparable behaviour to the average indicator scores, but they too have an insignificant negative trend, if any at all. Looking in particular at each price review period, it appears that all of the complete review periods - PR04, PR09, and PR14 - follow a more consistent local trend, before changing near the end of the review time window. This could reflect various things: on one hand, it could reflect an anticipation for the next review period which is drafted and deliberated years prior to its implementation; on the other hand, it could instead reflect a sort of 'phase-in' process, wherein the actual consequences of the current review - the quality changes in this case - take multiple years to see tangibly.

Other issues could also persist in this type of model. Compared to the forthcoming adjustments for environmental difference between companies, the pooled model might not require such concerns, given that the entire industry is itself one observation in the model. However, in a similar sense, because the whole industry is being compared over each year, there could well be significant external events in particular years that have impacted how the industry operated, and by extension how the industry addressed or improved quality as a whole. This issue is also one that would also affect results at a company level, but the impact of these effects are likely to be accounted for by the heterogeneties that will are being accounted for already.

5.4.2 DEA Modelling Results

As Figure 5.4 illustrates somewhat, there are notable difference in efficiency scores between the composite indicator DEA model and the other DEA speci-

Figure 5.4: Average Efficiency Scores across DEA Models, 2002/03 - 2019/20

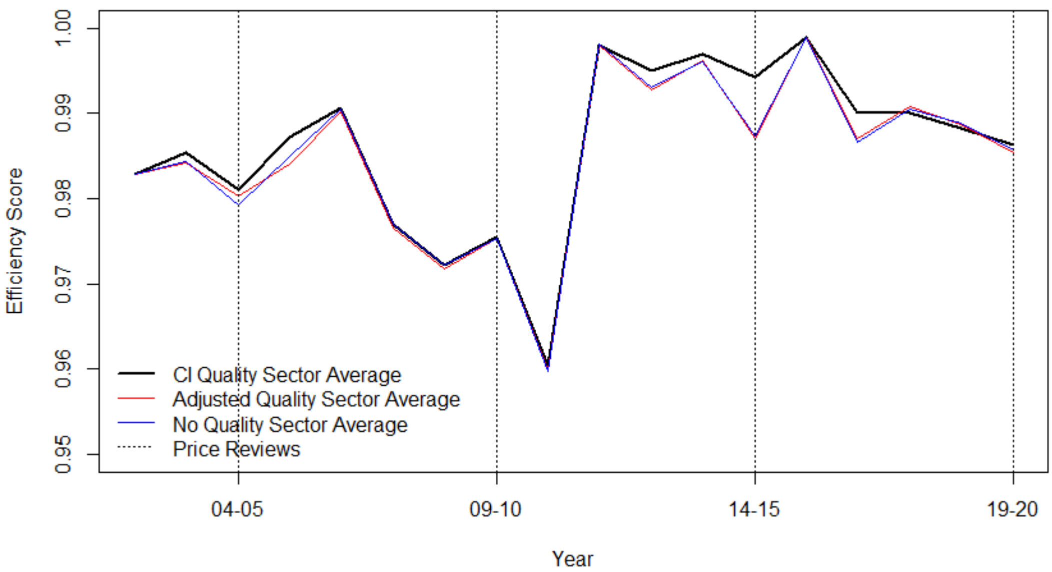


Table 5.6: One-Stage CI DEA Efficiency Scores, 2002/03 - 2019/20

DMU	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
ANG:	0.957	0.962	0.960	0.972	0.979	0.956	0.944	0.958	0.929	0.988	0.982	0.991	0.993	1	0.972	0.975	0.981	0.966
NWL:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SRN:	1	1	1	1	0.997	0.964	0.978	0.967	0.948	0.996	0.999	1	1	1	0.989	0.978	0.980	0.969
SVT:	0.958	0.966	0.961	0.976	0.982	0.964	0.951	0.959	0.944	1	0.981	0.990	0.982	1	0.989	0.987	0.980	0.967
SWT:	0.996	0.996	0.962	0.976	0.981	0.962	0.957	0.961	0.971	1	1	1	0.992	1	0.989	1	1	1
TMS:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
UUW:	0.960	0.966	0.961	0.979	0.983	0.962	0.955	0.960	0.943	0.995	1	1	1	1	0.990	1	1	1
WSH:	0.960	0.962	0.963	0.975	0.983	0.962	0.950	0.958	0.943	1	0.984	0.990	0.977	0.989	0.973	0.976	0.977	0.965
WSX:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
YKY:	1	1	1	1	1	1	0.991	0.997	0.918	1	1	1	1	1	1	0.991	0.973	0.995

Dashed lines indicate Price Review periods.

fications. For the most part, it seems that, on average, the indicator DEA has in most year better or insignificantly different efficiency scores than the other models, with minor exceptions near the end of the time period.

There are multiple ways that this could be explained: In the most optimistic case, it could be that the inclusion of a quality indicator variable into the model causes more relative efficiency improvements than not throughout the sector, suggesting perhaps that the companies are more efficient on average when their investments into quality improvements are considered as a distinguished output of production rather than as an adjustment to other outputs. On the other hand, owing to the already known issues of dimensionality[2], it could instead be that these models, though bootstrapped to try and remove some of the small sample bias, have lost some discriminatory power when the additional variable is added into the composite indicator model. In that case, the general average increase in scores might instead reflect only a loss in discrimination, rather than the revelation of quality-based efficiency improvements. Furthermore, owing to the mechanics of the bootstrapping method, there may well be an upwards bias in the scores because of the need to bootstrap them, which gives way to a sort-of-replacement of the small sample issues via the solution to that problem.

This latter thought could arguably be a weaker case than the former, however, given that not only are the changes in average efficiency inconsistent in magnitude over time, but are also inconsistent in direction - it might be expected that, if there is in fact a loss in discriminatory power, then all scores are uniformly closer to the efficient value of one in each time period, which is evidently not the case. Table 5.6, however, shows the actual scores over time for all companies for the composite indicator model, and appears to add more credence to this latter argument of dimensionality issues. A significant amount of the scores are fully efficient, with companies such as NWL, TMS and WSX showing complete efficiency throughout the entire sample. At this stage in the modelling, it could well be argued that an issue with these results comes from not yet accounting for environmental conditions, but even with that consideration there is a clearly large amount of observations that are efficient - a sign that is more likely to be because of a loss in discriminatory power. An apt example for this argument is the 2015/16 observations of these scores, where only WSH appears to be inefficient.

Regardless of quite how the differences might be interpreted, the statistical significance of these modelling differences should be tested to determine if the

[2]As both Chapter 4 and Appendix B discuss.

Table 5.7: Wilcoxon Signed-Rank Tests for One-Stage DEA Model Efficiency Scores

Model Pair:	Significance Level:
CI, No Quality	0.0004***
CI, Adjusted Quality	0.0012***
No Qual., Adj. Qual.	0.6594

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

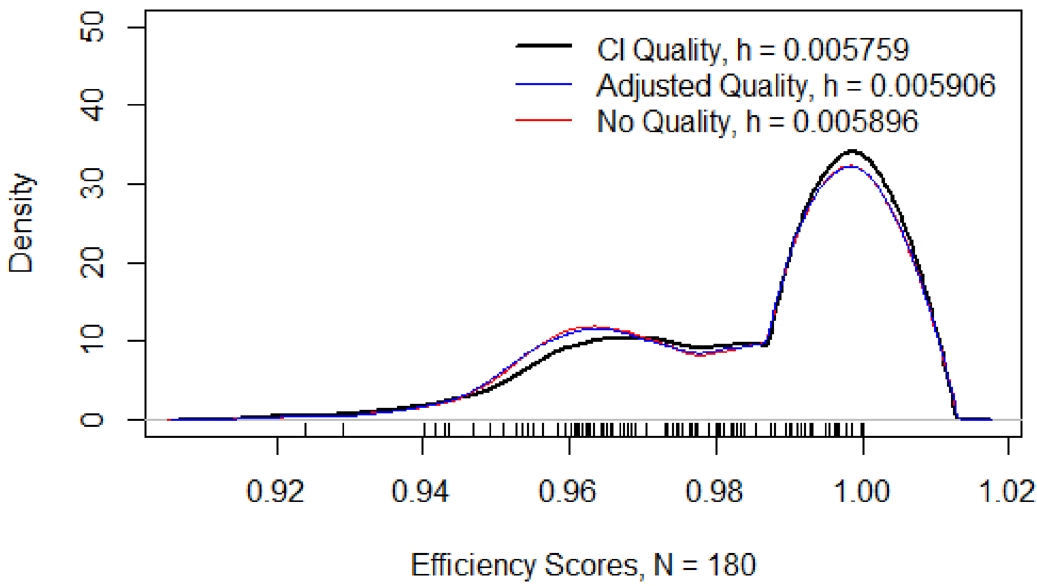
case for the novel DEA model is upheld as sufficiently useful. Table 5.7 displays the paired Wilcoxon Signed-Rank Tests between each pair of models, and finds that in all pairs containing the new composite indicator model, there is a highly significant difference in the efficiency of that model and the other model being compared. Interestingly, and based on the premise for this research, perhaps unsurprisingly, there is no significant difference between the DEA model with no quality measurement and the model with the older output-adjusted quality measures.

So, based on the conclusion of these test statistics, it appears that not only was Saal et. al. (2017) entirely right in suggesting a so-called stagnation in the measures of quality in the industry - as indeed the old quality measures appear to have no functional difference on the efficiency scores - but also that this new model containing the composite indicator measurement of quality is sufficiently different to previous models which, on top of the previous analysis finding much variability in the composite indicator scores and ranks of companies over time, might well suggest that this new method appeals to the recent need for newer and more non-uniform quality measurements - providing perhaps some sort of use for these new methods as targets for further quality improvements.

Further analysis can be given by looking at the KDEs of the efficiency scores for each model, and is illustrated by Figure 5.5:

The results are similar to the efficiency score graphs: the distributions are quite heavily negatively skewed, with the modal efficiency score being around

Figure 5.5: One-Stage Efficiency Score KDE Distributions, 2002/03 - 2019/20



1, suggesting that many observations are fully efficient; an observation that is demonstrated by Table 5.6. Though there is a second increase in density at around 0.96 in all models, the fact that the distribution appears to exclusively be for efficiency of 0.9 or higher suggests that all companies are well-performing in each model, with some differences in distribution due to the inclusion of the composite indicator. As before, however, these results also suggest that other issues might be present: there might well be over-estimations due to not accounting for environmental factors; the bootstrapping method might have biased the scores upwards as a consequence of its use; or the models, and in particular the indicator model, have insufficient discriminatory power relative to the small sample size.

Figure 5.6: Average Efficiency Scores across 3SDEA Models, 2002/03 - 2019/20

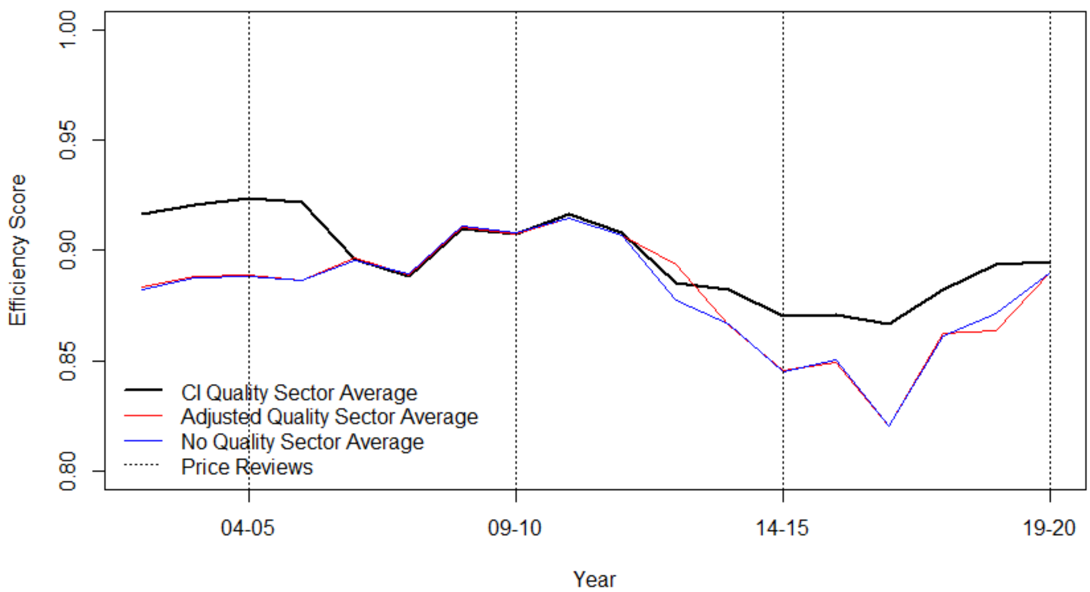


Table 5.8: Three-Stage CI DEA Efficiency Scores, 2002/03 - 2019/20

DMU	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
ANG:	0.596	0.585	0.622	0.628	0.596	0.643	0.652	0.655	0.685	0.628	0.557	0.576	0.616	0.528	0.517	0.527	0.552	0.545
NWL:	0.915	0.925	0.892	0.944	1	1	1	1	0.782	0.872	0.850	0.873	0.772	0.800	1	1	1	0.874
SRN:	1	1	1	1	0.664	0.684	0.832	0.783	0.981	0.834	0.732	0.822	0.860	0.773	0.758	0.613	0.698	0.642
SVT:	0.785	0.806	0.805	0.810	0.780	0.821	0.790	0.788	0.827	0.775	0.716	0.724	0.714	0.763	0.714	0.782	0.795	1
SWT:	1	1	1	0.962	0.959	0.846	0.896	0.851	0.899	1	1	0.913	0.898	0.964	0.872	0.974	1	1
TMS:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
UUW:	0.986	1	1	0.981	1	0.934	0.960	1	0.988	0.971	1	1	1	1	0.931	1	1	1
WSH:	0.884	0.890	0.918	0.891	0.963	0.956	0.966	1	1	1	1	0.916	0.842	0.879	0.874	0.925	0.889	0.888
WSX:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
YKY:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Dashed lines indicate Price Review periods.

Three-Stage DEA Results

Analogous to the one-stage results, Figure 5.6 and Table 5.8 show the average three-stage DEA efficiency scores over the sample time period and the full table of scores for all companies for the composite indicator DEA model, respectively. Figure 5.6 shows that, like Figure 5.4, the indicator model seems distinct from the other two models, which follow very similar trends over time. The three-stage results, perhaps as a consequence of the removal of environmental heterogeneities, seem more pronounced, in that the efficiencies for all models are lower overall, and the indicator model yields more noticeable improved efficiency scores on average in a majority of the years in the sample.

Table 5.8 sees a desired reduction in fully-efficient scores, though there is still a quite significant amount of them, with TMS, WSX and YKY being fully efficient in all time periods despite the adjustment for environmental factors. These results show more promise than the first-stage results in this respect, and match the results of Pointon & Matthews (2016) in that there are less fully-efficient scores and lower scores overall, but demonstrate more robustly that there are likely to be issues related specifically to discriminatory power, especially since the bootstrapping method of the three-stage procedure does not have the same upward biasing problems of the one-stage models.

Table 5.9: Wilcoxon Signed-Rank Tests for Three-Stage DEA Model Efficiency Scores

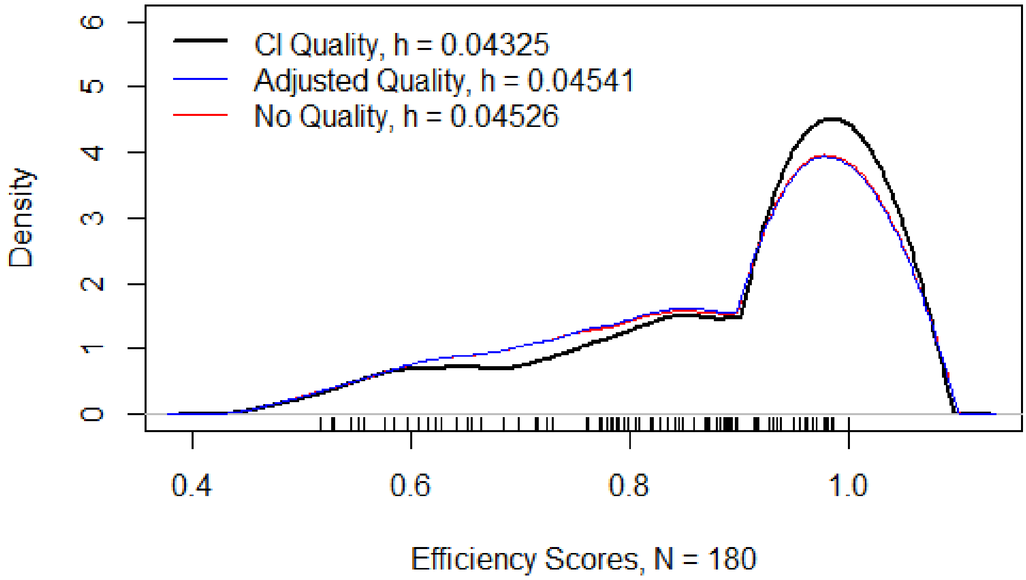
Model Pair:	Significance Level:
CI, No Quality	0.0015***
CI, Adjusted Quality	0.0375**
No Qual., Adj. Qual.	0.3432

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

As before, Wilcoxon Signed-Rank tests are performed for each pair of model specification and, as before, the composite indicator model is significantly different from the older output-adjusting quality model and the model without quality. Furthermore, both of the latter two models appear to not be significantly different from each other. These results mirror the one-stage test results and show that, accounting for the operating environment heterogeneities, the older definition of quality used in modelling is no longer sufficient in explaining differences in quality improvements, and that the novel composite indicator approach still shows promise by providing significant changes in the results.

Figure 5.7 shows the KDE distributions of the efficiency scores for each three-stage model, and the results drawn from this are also similar to those of the one-stage models: the indicator model provides a greater likelihood of a

Figure 5.7: Three-Stage Efficiency Score KDE Distributions, 2002/03 - 2019/20



high efficiency score, when compared to the other models, but all models have a modal efficiency of about 1, which again suggests that there may still be issues with dimensionality, or that there are yet more factors unaccounted for. Interestingly, the three-stage model seems to no longer have the slight peak in density in the middle efficiency scores, which suggests that the removal of environmental differences has positively impacted some companies by improving their comparative efficiencies once their operating conditions were accounted for - a result that adheres nicely to the expected consequences of using the three-stage adjustments.

In summary, many of the behaviours in the results exhibited by the one-stage models are reflected in the three-stage models, giving more robustness to those outcomes. Though it seems that the procedure has effectively accounted for operating differences by removing some lower comparative efficiency scores and lowering the average efficiency overall, there is still a present issue with many observations displaying full efficiency. While this could well suggest, for example, that TMS and WSX are robustly industry-leading in terms of technical efficiency with the inclusion of a quality output, it more likely suggests that the models are still suffering from small-sample bias issues that the bootstrap procedures have not fully removed.

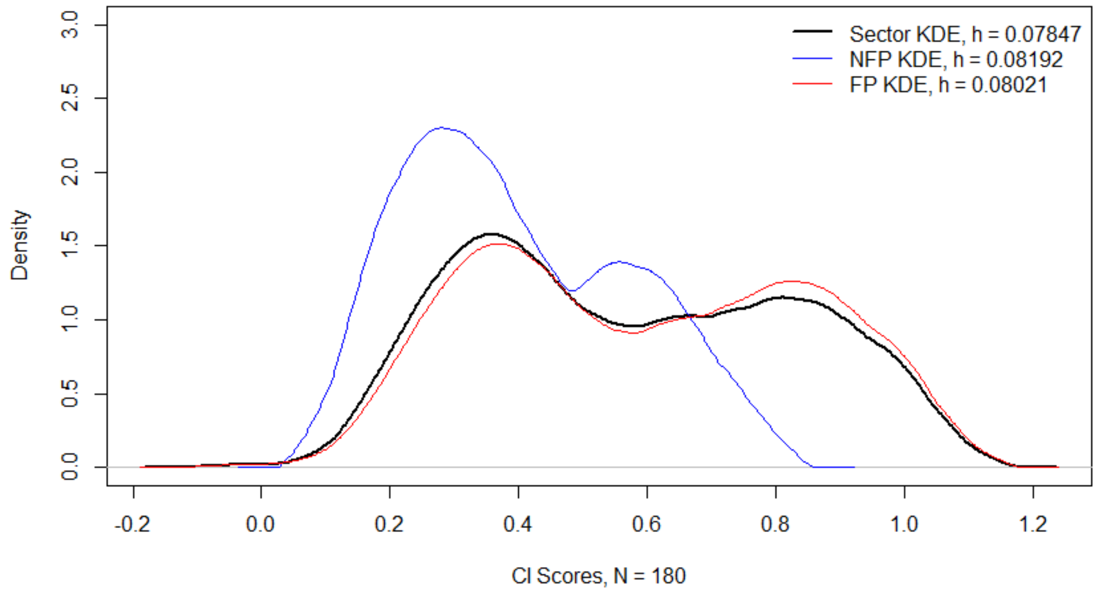
Though a weaker argument than the former, there is also a case where other factors are not properly considered. In one sense, this could include factors that have influences similar to the operating environmental factors, but there may also be factors omitted that consistently act upon the model regardless of operating conditions. An immediate suggestion, then, is that the composite indicator itself might contain quality factors that certain companies excel in, to the detriment of other companies in the industry. Perhaps, similarly, if other quality factors were consolidated into the indicator somehow, the models would better account for industry quality, and the scores would therefore decrease on average, lowering the spuriously high amount of fully efficient observations.

For-Profit vs. Non-Profit Results

Lastly, the entirety of the preceding analysis is re-considered according to the fifth research question of this thesis, and assesses the differences between For-Profit and Not-for-Profit results.

Figure 5.8 illustrates the KDE distribution of the Lower Bound composite indicator, which is the 'default' specification in that it is further used in the DEA models and so bares the most importance. In decomposing the Sector efficiency scores into two sub-samples - the For-Profit companies' scores and the Non-Profit scores - it can be seen that the NFP quality scores appear to be somewhat uniformly lower than the rest of the industry. On one hand, this could suggest again that there are issues pertaining to the exclusion of quality factors that the Non-Profit company, WSH, excels in, but on the other hand it perhaps more likely suggests that the company simply isn't performing as well

Figure 5.8: Lower Bound CI KDE Distributions, 2002/03 - 2019/20



as its For-Profit counterparts. These results do also lend credence to the idea of a need for indicator-level adjustments of operating heterogeneities, however, as despite not doing so in this chapter, it may well be that the quality performance is impacted by operating conditions that leaves WSH at a disadvantage.

Continuing on, Figures 5.8 and 5.9 show the average Efficiency Scores of the sample time period, and the respective KDE distributions of those efficiencies, in their decomposed form for the CI DEA model. These results once again seems to suggest that, when compared to the For-Profit behaviours of the other companies, WSH appears to be performing worse overall, with a modal efficiency of about 0.97 after bootstrap correction. Uniquely, WSH seems to have a somewhat symmetrical distribution of its efficiencies about its peak, suggesting that its behaviour, though less efficient on average, is more consistent than the rest of the industry.

Figure 5.9: One-Stage Efficiency Scores, 2002/03 - 2019/20

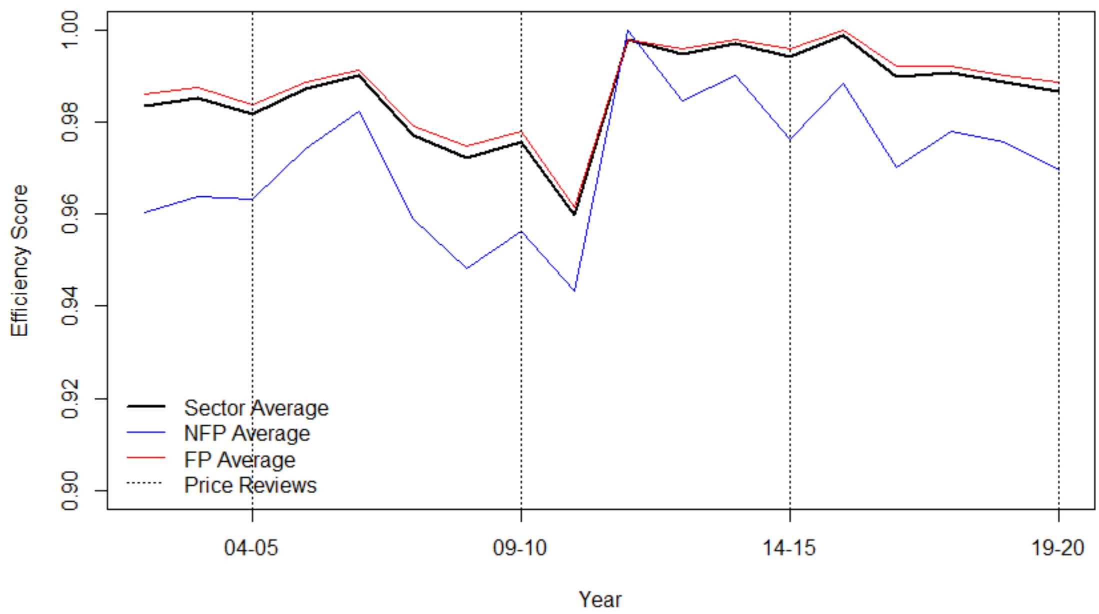


Figure 5.10: One-Stage Efficiency Score KDE Distributions, 2002/03 - 2019/20

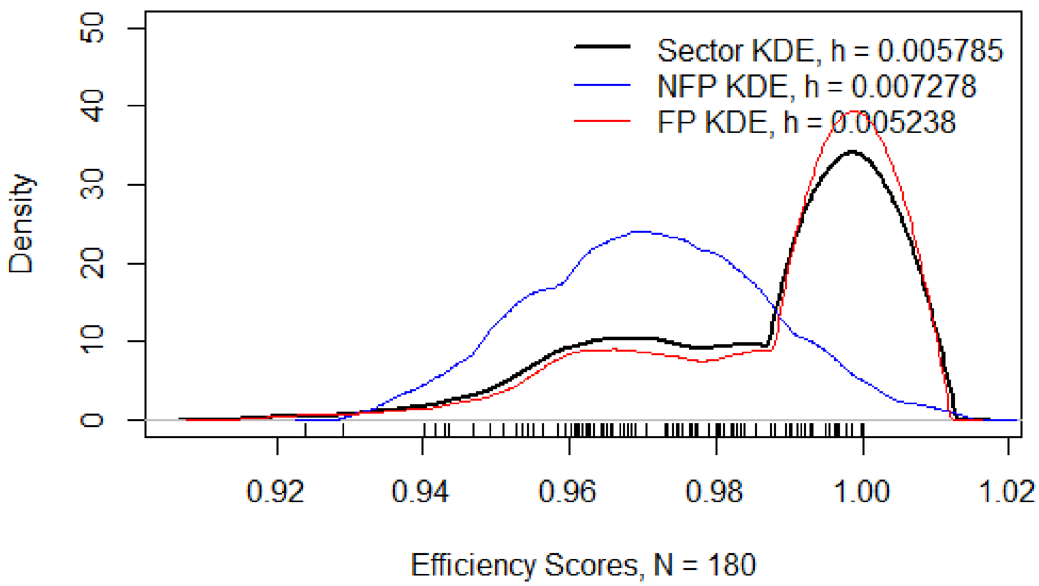


Figure 5.11: Three-Stage Efficiency Scores, 2002/03 - 2019/20

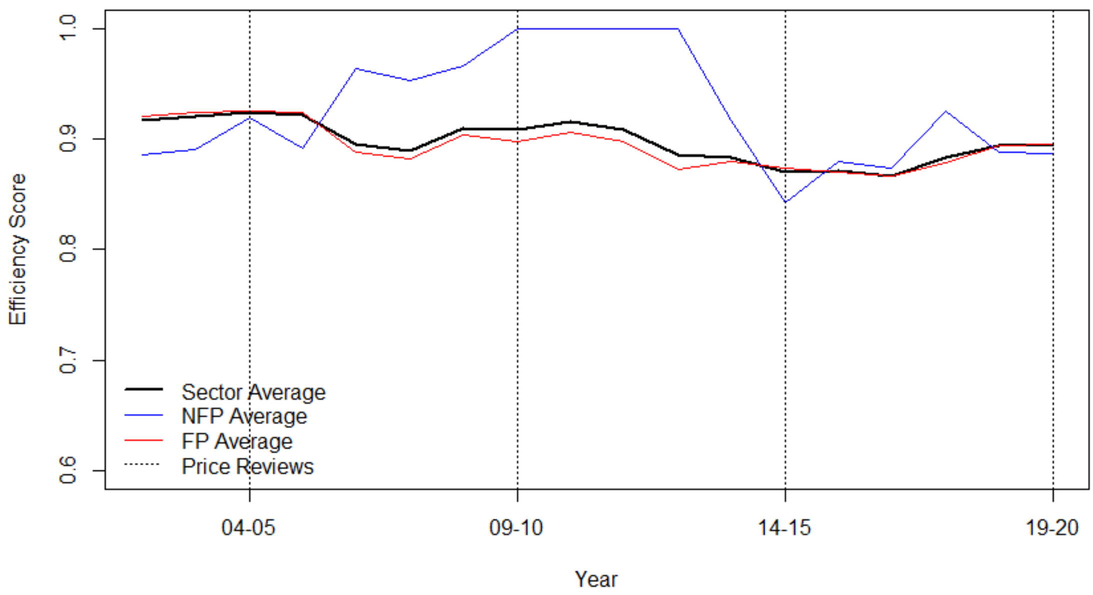
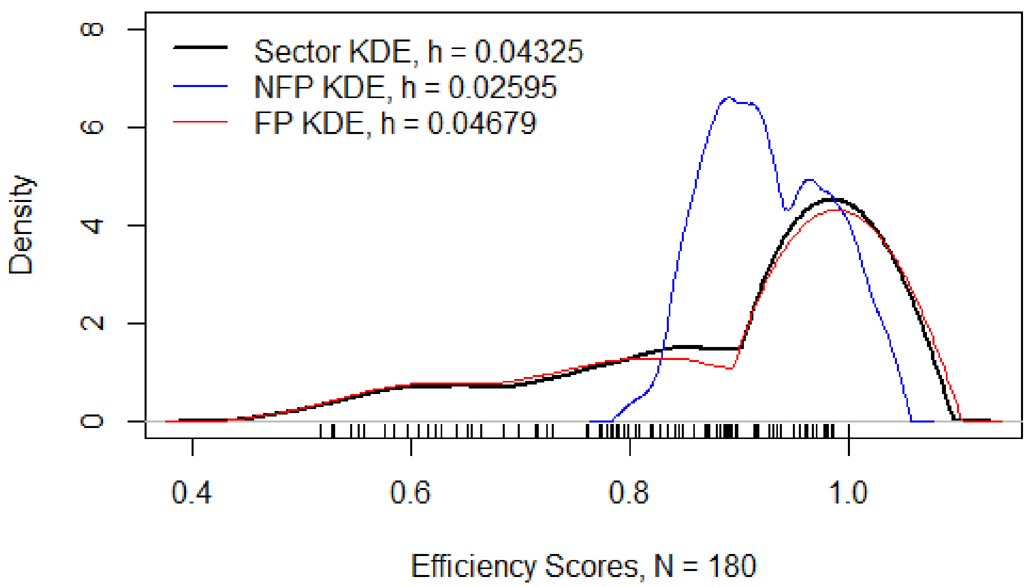


Figure 5.12: Three-Stage Efficiency Score KDE Distributions, 2002/03 - 2019/20



Finally, Figures 5.10 and 5.11 show the same decomposed average efficiency scores and respective KDE distributions for the three-stage indicator DEA model. These results are the most striking of this part of the analysis, by virtue of their opposition to the previously drawn conclusions of this section. In accounting for the environmental differences within the industry, it appears as though NFP behaviour has instead yielded a significant improvement in efficiency over the 2006/07 - 2013/14 period, as well as the 2015/16 - 2017/18 period. This observation, most interestingly, suggests that a large reason for the previous relative inefficiencies of WSH were perhaps because of the operating conditions, as it appears as though there is a generally greater efficiency in the sample once all companies are considered under the same operating environments.

The KDEs do reflect this conclusion as well, if a little obtusely, and support the positivity around the idea that WSH, and by extension in this analysis NFP behaviour, is more consistent than its for-profit counterparts. The NFP distribution exhibits a bi-modal shape similar to the composite indicator, with it most modal value at about 0.85, and its next-most modal value at about 0.95. Where this distribution supports the idea of consistent efficiency scores being a good thing is in its lack of skew: as in the one-stage distribution, the NFP efficiency scores are relatively without skew, especially when compared to the high negative skew of the rest of the industry. As the average score trends show in Figure 5.11, the consistent performance of WSH has meant that it appears to be a more efficient company on average compared to for-profit companies, which despite having far more fully efficient observations, are pulled downwards by the incidence of far less efficient results as well.

5.5 Conclusion

This chapter aimed to develop a novel measure of quality across the various facets of the English and Welsh Water and Sewerage industry, in such a way that multiple areas of interest related to quality improvement were accounted for, while minimising the consequences associated with the small-sample issues known to cause losses in explanatory power from the addition of too many variables into conventional DEA models. In doing so, the first two Research Questions of thesis have been addressed.

What this chapter has hopefully shown is that, using a Composite Indicator approach as had been used in various macro-economic comparative measures, a single value can be built out of many quality factors, according to a design specification that is, in this case, somewhat autonomous in its determination of which factors matter most to companies in a given year. Further, though the resulting indicator has moderate uncertainty related to its design choice and generally moderate variance, it has addressed the concerns of reports such as Saal et. al. (2017) when applied to conventional DEA models by providing a statistically significant difference in the resulting efficiency scores of the model that applies it as a new independent production output, especially when compared to models with no quality or the older measures of quality, which are not significantly different to each other in this data window.

The DEA applications of this indicator have shown two key results related to the total suite of quality measures: first, when quality is treated a production output, and therefore the choice between service production and quality investment is presented, technical efficiency scores significantly change relative to models without a quality output, implying that the addition of quality as an output might well present a more accurate reflection of the production decisions firms are facing, which previously would have be obfuscated within the service outputs. Second, that the old measures of quality, applied in a DEA setting, do not produce significantly different technical efficiency scores to the same model with no quality adjustment at all implies that the measures are statistically superfluous - models hereafter that refer to older quality measurements are no different to models with no quality at all. As it pertains to future modelling, this seems to imply that, unless quality is treated with more novel ideas and more significance in the production process, it might not be statistically prudent to include quality at all.

Ultimately, this indicator is, hopefully, only the first iteration of its kind. The results of the indicator show that there is certainly room for new regulatory targets on improving multiple factors of quality across the industry, following this approach, but the indicator itself could well require improvement to potentially better reflect how quality affects efficiency when treated as a result of production.

From the indicator design point-of-view, various changes could be made to better describe how quality is collated into the final quality output. For example, though the weighting systems used in this chapter are specifically such that no personal opinions are used - with the exception of a lower bound - it might well be more agreeable to somehow generate these weights by also including consensus from a collection of industry experts, such as regulators and company economists. Similarly, following the example of the Innovation Index, one other change in the future, given that data on other quality aspects can be found and measured consistently over time, could be to used Sub-Indices for each area of quality, which are themselves weighted and organised before their final aggregation. On the point of referring to at least the sampled indicators, more initial outside information or modelling could be used to find the most suitable variables for each indicator or sub-indicator: Regression Analysis or Principal Component Analysis might, for instance, find unlikely factors to heavily affect quality that were unexplored here; similarly, other non-discretionary factors might have their own position in determining quality outside of adjusting the indicator's weights.

All in all, though the indicator could be improved upon, it serves as an interesting point of potential development in the industry with respect to how quality is measured for the purpose of its improvement, as well as its place in other modelling procedures. Further analysis, through the proper specification of a test for Uncertainty Analysis, or other evaluative procedures such as Sensitivity Analysis through methods such as those in Saltelli et. al. (2010) or Rahman (2016), could also be insightful in future iterations by further providing explanation of the contributions of each inputted factor to the outputted index, be it independently or otherwise. Different applications of composite indicators in the same context could also be considered, such as El Gibari et. al. (2022)'s use of composite indicators to create DEA outputs.

One final thing to consider about the results presented in this chapter, specifically those technical efficiency scores which are modelled with the composite indicator, is the potential differences between these scores and the scores one could achieve with each quality factor modelled separately. As the thesis has already discussed, one pragmatic advantage of the composite indicator is that it reduces the dimensionality problem of DEA models in this industry, which suffers from a relatively small sample size in each year. However, the CI measure could, on the other hand, obscure some companies' relatively strong performances in some of the quality factors. For example, a company could be relatively leading in leakage reduction, and so they have a high measure, and relatively lower weightings for that factor. The indicator would then provide a much more pessimistic take on that company's leakage performance, as their lead in the industry is offset by low weighting in the end. This is more pertinent in the common weights model, where companies by definition share common weights which might not correctly reflect their individual priorities.

This chapter has also sought to partly answer the fifth research question of the thesis, by addressing at all empirical stages the differences in results due to what is assumed to be behavioural differences based on profit-seeking and profit-deviating choices. In so doing, it appears that while Non-Profit behaviours appear to yield worse quality scores, such a conclusion could well be an insufficiency of the indicator used. Further, though Non-Profit behaviours appear to yield less technical efficiency in static models relative to for-profit behaviour on average, adjustments for differences in operating environments result in a quite interesting change in position on average, with non-profit behaviours exhibiting instead a more consistent, greater efficiency on average over the time period assessed. This suggests that, though not necessarily the most efficient at a company-level, the extreme distribution of for-profit efficiencies is detrimental from a technical efficiency point-of-view on average, especially after any environmental advantages are removed.

The next stage of the thesis is to extend this indicator concept to dynamic DEA models, rather than static models repeated independently over time. As will be discussed, in treating quality as a quasi-fixed variable, the following chapter aims to see how Capex Bias is affected by the explicit inclusion of quality by assessing changes in Allocative Efficiency over time, as to see if the well-evidenced industry habit of over-investment into Capital is accounted for in some way by investing into longer-term improvements in quality.

Chapter 6

Topic 2: Dynamic DEA and Allocative Efficiency

The previous chapter's results inform us of a very important insight: inclusion of the Composite Indicator of quality yields significantly different improvements in Technical Efficiency on average.

In utilities industries, however, there is another type of efficiency not explored in Chapter 5, that provides even greater evaluation of how companies operate, and if they do so efficiently. The notion of Capex Bias, defined in Averch & Johnson (1962), is the idea that companies, in this case DMUs in the water and sewerage industry, over-use Capital expenditures relative to other Operational expenditures, as they prefer investment into shorter-term project that rely heavily on capital over, say, improving Labour via long-term training programs. As a result, the industry is likely to face low Allocative Efficiency, which is defined by how optimal the 'choice' of production factor quantities is - if Capex bias does exist in the industry, then the companies are choosing far too much capital which, while potentially still giving high technical efficiency, will give low allocative efficiency as a result.

As mentioned previously in the thesis, Averch & Johnson (1962)'s model, in which they define capex bias, relates to RoR regulation. The solution proposed for this problem in the water industry was the $RPI - X$ price cap regulation, and following this principle, the industry now employs an $RPI + K$ price cap structure, which covers an efficiency improvement term as in the original price cap definition, and a quality improvement factor. This chapter follows the basis aforementioned, where the additional point of interest in $RPI + K$ calculations, while good from a regulatory standpoint, may actually be causing capex bias by permitting another avenue down which companies can over-invest into short-term capital intensive projects.

Furthermore, one assumption of the previous models is that they are Static - each year of observations, and therefore each companies' decisions, are independent from year to year, which also implies that all factors of production are Variable - they can be adjusted year-on-year without consequence. In practice, even if capital-intensive projects are considered shorter-term, they are not instantaneous and are likely to occur over multiple time periods in the sample of the models. With that in mind, capital should more accurately be defined as Quasi-Fixed: in the Short-Run, capital in production is fixed, but becomes variable in the Long-Run.

$$\underbrace{f : (\bar{K}, L, OC) \mapsto (W, WW, CI)}_{Short-Run} \longrightarrow \underbrace{f : (K, L, OC) \mapsto (W, WW, CI)}_{Long-Run} \tag{6.1}$$

Where the distinction between short- and long-run capital is the use of a bar, $\bar{K}$, for fixed short-run capital. For this to be operationalised into a DEA model similar to Chapter 5, Nemoto & Goto (2003) propose a Dynamic DEA model that also accounts for Quasi-Fixed Inputs.

This chapter intends to use this model to create efficiency scores with dynamic behaviour due to quasi-fixed capital. In doing so, two major points of interest are assessed, relating to Research Questions 3 and 4: to what extent has Allocative Efficiency changed over time in the industry, with particular focus on regulatory changes in PR14; and to what extent does Capex Bias, measured by over-use of Quasi-Fixed Capital and Allocative Efficiency, change over time in dynamic models that include the composite indicator of quality.

Research Question 3 looks at a standard dynamic model, with no quality measures at all given that Chapter 5 also finds that there is no significant difference between old quality measures and no inclusion of quality.

6.1 Capex Bias Research

All in all, there isn't too much research specifically on Capex Bias, let alone how it impacts and is reflected in the water and sewerage industry. The effect is attributed to Averch & Johnson (1962), and is therefore otherwise known as the Averch-Johnson Effect, whereby a regulated monopoly responds to a revenue constraint by increasing their Capital input in production - they bias their operations towards the use of capital, over other inputs such as Labour.

This section on the research around Capex bias will go cover the model of Averch & Johnson (1962), as well as some more recent research that address the bias, including a recent development of Fixed Opex-Capex Shares (FOCSs) as a solution to the biasing problem. What research exists specific to the water industry is also examined, to explore what solutions were drawn by the industry as well.

6.1.1 Models of Capex Bias

Starting with Averch & Johnson (1962), a monopoly firm is assumed to have Profits defined by:

$$\pi = py - r_1x_1 - r_2x_2 \tag{6.2}$$

With y and x's referring to output and production inputs respectively, and p and r denoting the respective output and input prices. x_1 is assumed to be the capital input of interest. Assuming no depreciation and a unit Acquisition Cost of Capital, the model's regulatory constraint is defined by:

$$\frac{py - r_2x_2}{x_1} \leqslant s_1,\ py - s_1x_1 - r_2x_2 \leqslant 0 \tag{6.3}$$

Where s_1 is the regulated Rate of Return. The paper then defines a set of Kuhn-Tucker conditions for the model, ultimately yielding two important results: first, the Factor Prices, which can be interpreted as rates of return on the inputs, can be defined as:

$$r_i = \left(p + y\frac{\partial p}{\partial y}\right)\frac{\partial y}{\partial x_i},\ \forall i \tag{6.4}$$

Second, when the regulatory constraint is binding on the monopolist, the Marginal Rate of Substitution between the inputs is:

$$-\frac{\partial x_2}{\partial x_1} = \frac{r_1}{r_2} - \frac{\lambda}{1-\lambda}\frac{s_1 - r_1}{r_2} \tag{6.5}$$

Where λ is the model's Lagrange Multiplier. Given that the second part of the above equation is found to be strictly positive, it can then be concluded that:

$$\frac{\partial x_2}{\partial x_1} < -\frac{r_1}{r_2} \tag{6.6}$$

That is, under the rate of return regulatory constraint, the marginal rate of substitution is less than the socially optimal ratio of input prices, $-\frac{r_1}{r_2}$. To adjust to this constraint, the firm is found to increase their capital input, relative to the other variable input. This increase in capital expenditure, as to account for the constraint and best profit-maximise, is the behaviour that defines Capex Bias.

More recent literature on the bias proposes an alternative solution to the over-investment into capital, that of Fixed Opex-Capex Shares, or FOCS. Brunekreeft & Rammerstorfer (2020) approach this idea by modelling Opex Risk as a source of Capex Bias: Defining a Risk Factor, $\beta(K, L)$ as a function of Labour and Capital, L and K, which is a measure of systematic risk. The costs to a firm are now defined as:

$$C(K,L) = wL + (r + \beta(K,L))K \tag{6.7}$$

Where the additional risk factor appears as a risk premium on top of the risk-free rental rate for capital, r. In the case where the model is not constrained by a revenue cap, the equilibrium Marginal Rate of Technical Substitution is found to be:

$$\frac{f_K^*}{f_L^*} = \frac{(r + \beta^*) + \beta_K^* K^*}{w + \beta_L^* K^*} \tag{6.8}$$

On the other hand, if the constraint to revenue does bind, and there exists a Lagrange Multiplier $\lambda \neq 0$, then the Distorted Equilibrium is defined by:

$$\frac{f_K^D}{f_L^D} = \frac{(r + \beta^D) + \left(\frac{1}{1-\lambda}\right) \beta_K^D K^D}{w + \left(\frac{1}{1-\lambda}\right) \beta_L^D K^D} \tag{6.9}$$

Comparing both equations, the paper proposes that, in the long-run equilibrium, the opex-risk leads to a systematic capex-bias, illustrated by the fact that $\frac{f_K^D}{f_L^D} > \frac{f_K^*}{f_L^*}$, and therefore $\frac{K^D}{L^D} > \frac{K^*}{L^*}$.

Brunekreeft & Rammerstorfer (2020) propose two possible solutions to this distorted equilibrium born of the systematic risk premium of capital. The first approach, and the one we will focus on, is the FOCS approach, which is a variant of Totex regulation that will be shortly discussed. In this approach, costs are collected totally, and then split according to a Capitalisation Rate, α:

$$C(K,L) = (wL)^F + ((r + \beta)K)^F = (1 - \alpha)C(K,L) + \alpha C(K,L) \tag{6.10}$$

Where, though mathematically equivalent to the left-hand side cost function, the costs are considered to be split by the capitalisation rate α, which defined the fixed opex and capex shares. The fixed portion of capital, known as Quasi-

Capex, then faces a mark-up μ, which re-specifies the revenue constraint as:

$$R(f(K,L)) \leqslant (1-\alpha)C(K,L) + \mu\alpha C(K,L) = (1-\alpha+\mu\alpha)C(K,L) \quad (6.11)$$

Where a mark-up of $\mu = 0$ would simplify the equation to (6.10). Solving the model via the same Lagrangian process as the previous model, using the mid-section of (6.10) as the fixed labour and and capital, the equilibrium condition is defined as:

$$\frac{f_K^F}{f_L^F} = \frac{(r+\beta^F) + \beta_K^F K^F}{w + \beta_L^F K^F} \quad (6.12)$$

That is, (6.12) is identical to the undistorted equilibrium (6.8) - the use of a fixed capitalisation has removed the systematic risk distortion causing capex bias.

Interestingly, this idea has seen some practical use in the English and Welsh water and sewerage industry. Smith et. al. (2019) report on the capex-biases and corresponding regulations of various UK industries, and note Ofwat's use of Totex regulation, rather than Opex or Capex regulation specifically. By incorporating both costs first, and then regulating the resultant Totex, the industry regulates the aggregate in a fashion similar to the model described above. Even more recently, Ofwat adapted their regulatory objective further by changing to Botex regulation, wherein the Base Opex and Base Capex are instead totalled and then targeted. This avoids any biases caused by different enhancement costs in opex and capex projects, and instead only focuses on the base costs required for upkeep, and whether they suffer any capital over-investment. Smith et. al. (2019) later look to identifying sources of Capex in the industry and, following the quotes from Ofwat that the bias can be explained by a 'risk of failure and penalty strength', and 'wider requirements and incentives', interestingly note that the additional environmental regulators and additional mechanisms for regulatory targets - the ODIs in PR14 - might also be potential reasons for industry capex bias.

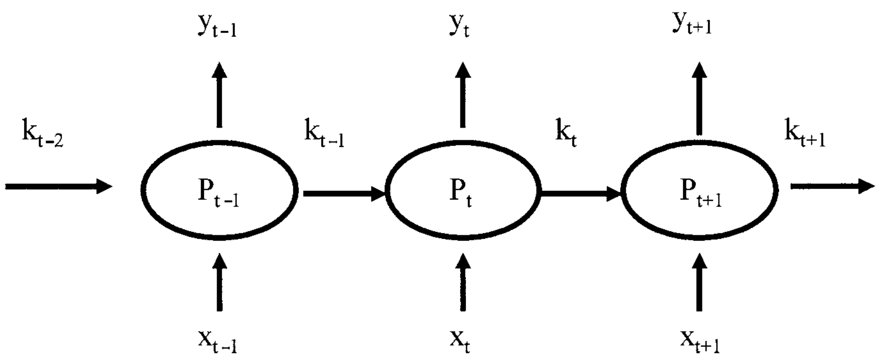
Ultimately, the research literature seems to find that the capex bias issues could well have been solved by Totex regulation-type modelling through the FOCS approach. As this was used in the industry, despite the apparent sources of bias from incentivised regulatory targets and additional regulatory pressures, there might well exist some degree of evidence that capex bias has reduced following Ofwat's cost regulation changes. The remainder of the chapter, via models that also account for dynamic behaviours of Capital in the industry, will look to see how the bias is changed over time via Allocative Efficiency.

6.2 Dynamic DEA Models

This section will first cover the model of Nemoto & Goto (2003), which is the dynamic model with quasi-fixed capital that will be used in this chapter to investigate allocative efficiency. The three-stage procedure for accounting for operating environment heterogeneities will be briefly recalled, and then other dynamic models will also be briefly described.

To understand the model of the paper, Figure 1 of Nemoto & Goto (2003) can be used to illustrate the channels of capital throughout the model over time.

Figure 6.1: Nemoto & Goto (2003) Model Diagram of Dynamic DEA



Nemoto & Goto (2003)'s model supposes that a firm has production P_t at time t, and requires not only current-time variable inputs x_t, but the rolling capital investment from the previous time period, k_{t-1}. The process yields outputs y_t, as well as remaining capital k_t which is then used as an input in the following time period, $t+1$.

The principal difference from Chapter 5's static DEA models is the separation of dynamic variables, Labour and Other Costs in context, from Capital, which is itself split into an input from the previous time period, and an output from the current period which feeds into the next period's DEA as an input. In effect, the capital invested from the previous year into the current years production process helps not only to produce the industry's water and sewerage services, by also to create capital to then be invested into the companies in the following year.

The resulting efficiency scores from this model are not quite Technical Efficiencies, which Chapter 5 drew from its models. Instead, the model's output Overall Efficiency, which can then be decomposed into Technical Efficiency and, more importantly for this chapter's research questions, Allocative Efficiency, according to the following:

$$OE_{i,t} = TE_{i,t} \cdot AE_{i,t} \tag{6.13}$$

Where, for each DMU i in each time t, overall efficiency can be decomposed into the product of technical and allocative efficiency. In Nemoto & Goto (2003), overall efficiency can be further decomposed to incorporate Dynamic Efficiency, the efficiency owed specifically to the inter-temporal aspects of the model:

$$OE_{i,t} = DE_{i,t} \cdot SE_{i,t}, = DE_{i,t} \cdot TE_{i,t} \cdot AE_{i,t}$$

In this chapter, however, the former decomposition will be preferred to the latter: Though dynamic efficiency, DE, and Static Efficiency, SE, could provide interesting insights about the performances of companies specifically because of inter-temporal factors, or the lack thereof, respectively, as the main focus is the on allocative efficiency, this chapter will prefer a simpler decomposition.

Formally, the overall cost function of a firm is defined as:

$$
\begin{aligned}
C(\bar{k}_0) = & \min_{\{x_t,k_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \left(w_t' x_t + v_t' k_{t-1} \right), s.t. \\
& (x_t, k_{t-1}, k_t, y_t)_{t=1}^{T} \in \times_{t=1}^{T} \Phi_t, \\
& k_0 = \bar{k}_0
\end{aligned}
\tag{6.14}
$$

The function minimises the total value of Variables Inputs x_t at time t, and Quasi-Fixed Capital given from the previous period, k_{t-1}, subject to the constraints that the combination of the inputs and corresponding outputs k_t and y_t are in the feasible Production Set Φ_t, for all t, and that the initial capital stock k_0 is fixed to $\bar{k}_0$. This function can be estimated via the following DEA model:

$$
\begin{aligned}
\hat{C}(\bar{k}_0) = & \min_{\{x_t,k_t,\lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \left(w_t' x_t + v_t' k_{t-1} \right), s.t. \\
& X_t \lambda_t \leqslant x_t,\ K_{t-1} \lambda_t \leqslant k_{t-1}, \\
& K_t \lambda_t \geqslant k_t,\ Y_t \lambda_t \geqslant y_t, \\
& \iota' \lambda_t = 1,\ k_0 = \bar{k}_0,\ x_t, k_t, \lambda_t \geqslant 0,\ \forall t
\end{aligned}
\tag{6.15}
$$

Overall Efficiency is then estimated as:

$$
OE = \frac{\hat{C}(\bar{k}_0)}{C} \tag{6.16}
$$

Technical Efficiency is found by finding the radial efficiency measure ϕ_t for the variables inputs, as is defined a the following problem:

$$
\begin{aligned}
\hat{C}_{TE} = & \min_{\{\phi_t,\lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \left(\phi_t w_t' \bar{x}_t + v_t' \bar{k}_{t-1} \right), s.t. \\
& X_t \lambda_t \leqslant \phi_t \bar{x}_t,\ K_{t-1} \lambda_t \leqslant \bar{k}_{t-1}, \\
& K_t \lambda_t \geqslant \bar{k}_t,\ Y_t \lambda_t \geqslant y_t, \\
& \iota' \lambda_t = 1,\ \phi_t \geqslant 0,\ \lambda_t \geqslant 0,\ \forall t
\end{aligned}
\tag{6.17}
$$

Technical Efficiency is then defined by:

$$TE = \frac{\hat{C}_{TE}}{C} \tag{6.18}$$

Finally, Allocative efficiency is defined as $AE = OE/TE$, which is identical to $\hat{C}(\bar{k}_0)/\hat{C}_{TE}$.

One noted insight from the results of the empirical application in Nemoto & Goto (2003) was the idea that typical DEA models might have biased results in comparison to the dynamic model, in terms of their resultant efficiency scores, due to the mistreatment of the quasi-fixed variables as variable. To test this, both the paper and this chapter compare the dynamic model results to static DEA models, where all inputs are considered variable:

$$\begin{aligned}
\bar{C}_{OE} = & \min_{\{x_t,k_t,\lambda_t\}_{t=1}^T} \sum_{t=1}^{T} \gamma^t \left(w_t' x_t + v_t' k_{t-1}\right), s.t. \\
& X_t \lambda_t \leqslant x_t,\ K_{t-1}\lambda_t \leqslant k_{t-1}, \\
& Y_t \lambda_t \geqslant y_t, \\
& \iota' \lambda_t = 1,\ x_t, k_t, \lambda_t \geqslant 0,\ \forall t
\end{aligned} \tag{6.19}$$

$$\begin{aligned}
\bar{C}_{TE} = & \min_{\{\phi_t,\lambda_t\}_{t=1}^T} \sum_{t=1}^{T} \gamma^t \phi_t \left(w_t' \bar{x}_t + v_t' \bar{k}_{t-1}\right), s.t. \\
& X_t \lambda_t \leqslant \phi_t \bar{x}_t,\ K_{t-1}\lambda_t \leqslant \bar{k}_{t-1}, \\
& Y_t \lambda_t \geqslant y_t, \\
& \iota' \lambda_t = 1,\ \phi_t \geqslant 0,\ \lambda_t \geqslant 0,\ \forall t
\end{aligned} \tag{6.20}$$

$$OE^S = \frac{\bar{C}_{OE}}{\bar{C}},\ TE^S = \frac{\bar{C}_{TE}}{\bar{C}},\ AE^S = \frac{OE^S}{TE^S} = \frac{\bar{C}_{OE}}{\bar{C}_{TE}} \tag{6.21}$$

As with the models of Chapter 5, there may still be a bias in the results due to small-sample dimensionality issues. To remedy this, the Three-Stage DEA approach will again be used to control for environmental heterogeneities between companies, bootstrapping the adjusted efficiency scores in the process. This process will be undertaken at the Overall Efficiency stage, and the adjustments made by the three-stage procedure will then carry on through the other results.

The above first set of models seeks to test the model with a relatively usual specification - Inputs, Quasi-Fixed Variables, and Outputs. Following this, this chapter also wishes to test the models with the inclusion of the novel composite indicator of quality as an output. Taking the first OE model as an example, the models including the CI quality measure will be defined, analogously to the previous models, as:

$$\begin{aligned} \tilde{C}(\bar{k}_0) = & \min_{\{x_t,k_t,\lambda_t\}_{t=1}^T} \sum_{t=1}^{T} \gamma^t \left(w_t' x_t + v_t' k_{t-1} \right), s.t. \\ & X_t\lambda_t \leqslant x_t,\ K_{t-1}\lambda_t \leqslant k_{t-1}, \\ & K_t\lambda_t \geqslant k_t,\ \tilde{Y}_t\lambda_t \geqslant \tilde{y}_t, \\ & \iota'\lambda_t = 1,\ k_0 = \bar{k}_0,\ x_t, k_t, \lambda_t \geqslant 0,\ \forall t \\ & \tilde{Y}_t = [Y_t, CI_t] \end{aligned} \tag{6.22}$$

The decomposition of OE into technical and allocative efficiencies, comparison to Static equivalent models, and the use of 3SDEA models are all carried out analogously as well, each with the additional quality output.

6.3 Data

One point of note about operationalising the models defined in section 6.2 is the difference between the theoretical and actual definitions of costs, inputs and input prices used in the industry. As will be discussed later in this section, various factors such as input prices and capital must be proxied in this chapter, as was discussed in Chapter 4, owing to the availability of data for the project, and to avoid difficulties with data gathering throughout the sample time period.

Furthermore, in practice, Ofwat's definitions of so-called total costs, Botex, are not technically the same as the theoretical definitions - base costs, by definition, exclude enhancement costs that would otherwise contribute to the total costs of a company. Hereafter, then, it is worth noting these data discrepancies, and how they might influence the forthcoming results.

Summary Statistics for this chapter's data are shown in Table 6.1. This chapter uses the same data for its outputs, composite indicator factors and three-stage variables as in Chapter 5 : Water Delivered and Equivalent Population are the water and wastewater outputs respectively, and in some models the Composite Indicator is also included; factors entering the indicator are Leakage, Pollution Incidents, and Total Complaints; the environmental factors accounted for in the 3SDEA procedure are Water and Wastewater Density, the Proportion of Distribution Input abstracted from Rivers, and the Proportion of Trade Effluent. Section 5.3, provides those definitions respectively, and data is again taken from annual WaSC reports (JARs and APRs) for the years 2002/03 - 2019/20.

Table 6.1: Data Descriptive Statistics - 2011/12 - 2019/20

Variable (Units)	Mean	Std. Dev	Min.	Max.
WaterDelivered (Ml/Day)	1004.82	544.49	270.96	2170.98
EqvPopulation (Pop.(000s))	5783.71	3847.29	1620.85	15771.01
CI (Nr.)	0.5742	0.2463	0.0522	1.0000
TotalEmployees (Nr.)	1765.17	751.16	555.00	3969.60
LPrice (£)	0.2046	0.0523	0.1348	0.3534
KStock (km)	61124.71	27872.63	24257.07	104016.3
KPrice (£/km)	0.0069	0.0021	0.0031	0.0128
TotalLength (km)	61252.23	27911.95	24406.18	104178.4
OCPrice (£/km)	0.0001	0.00009	0.00002	0.0003
Leakage (Ml/Day)	263.83	184.00	61.35	649.65
PollutionIncidents (Nr./1000km)	67.66	55.43	12	289
TotalComplaints (Nr.)	9662.72	7452.93	1467	34466
PropDIRivers (%)	0.3298	0.2202	0.007	0.732
WDensity (Pop.(000s)/km)	0.1614	0.0564	0.1070	0.3198
WWDensity (Pop.(000s)/km)	0.1732	0.0239	0.1324	0.2283
PropTradeEffluent (%)	0.0238	0.0102	0.0092	0.0438

What differs in this chapter are the inputs, which are required to represent physical stocks of the production factors, as to estimate prices for the Cost Efficiency models. The Stock of Labour is defined by the Total Number of Full-Time Equivalent Employees, and the corresponding Labour Price is calculated as the Total Base Opex divided by the total number of employees.

Some difficulties arise in the data collection for Capital and Other Costs variables. Typically, the MEA value of Capital would be used for capital stock, and the price can be proxied by a sum of depreciation rates and determinations by a WACC model. Other Costs are referred to as energy costs and other miscellaneous factors, and are priced by a corresponding energy price index. This chapter, in failing to find the appropriate data, instead uses known substitutes in the literature. To achieve this, the time frame of the models is reduced in comparison to the other contributing chapters, spanning from 2011/12 to 2019/20.

Other Costs have been alternatively defined by Bottasso et. al. (2011) as the sum of Water Mains and Sewer Mains, otherwise known as the Length of Mains and the Length of Sewers respectively. Following this approach, the Other Costs Price is then defined as the monetary value of Other Costs, defined in Chapter 5 as Total Costs less Base Opex and Base Capex costs, divided by the stock of Other Costs.

Water capital stock is defined in this chapter as the Length of Mains less the Length Relined and Length Refurbished - the idea being that this measure proxies capital stock for water services, removed of the captial required to enhance or improve its quality, giving a figure that should be analogous to Base Water Opex. Wastewater capital stock is similarly defined as the Length of Sewers less Gravity Sewers Rehabilitated and Rising Mains Replaced or Refurbished - Total captial stock is, then, the sum of water and wastewater capital, and the Capital Price is defined as Base Capex divided by the capital stock.

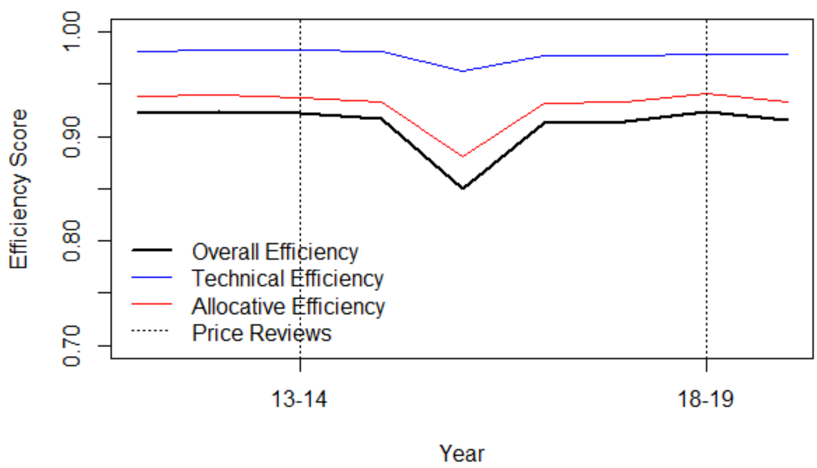
As discussed at the beginning of the section, some data - namely captial stock and the inputs prices, have been defined on an ad hoc basis owing to limited data availability for this project - is is again therefore worth attaching this caveat to the results of the following section, though the definitions chosen for those variables should be sufficient representations of their variables nonetheless.

6.4 Results and Discussion

6.4.1 Base Model Results

First looking at the base model, without the inclusion of the Composite Indicator, Figure and Table 6.2 report the average efficiency scores over time and by company, respectively. Figure 6.3, as with the right half of Table 6.2, illustrate these same efficiencies for the Static models.

Figure 6.2: 3SDEA Average Efficiency Scores, 2011/12 - 2019/20



All dynamic efficiencies are relatively high, with technical efficiency TE being close to full efficiency. While this could be a consequence of dimensionality issues, it could also be believed that the industry is, in general, performing well from a technical and overall efficiency perspective, with a lower but still relatively high level of allocative efficiency.

Figure 6.3: 3SDEA Static Efficiency Scores, 2011/12 - 2019/20

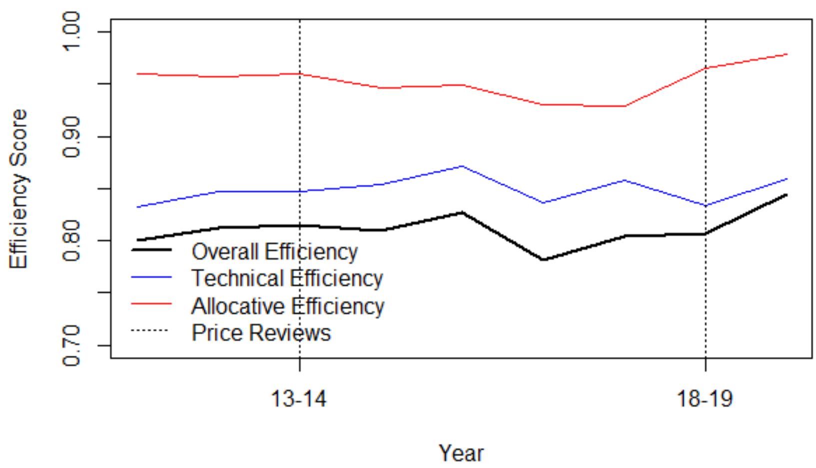


Table 6.2: 3SDEA Efficiency Scores, 2011/12 - 2019/20

DMU	OE	AE	TE	OE^S	AE^S	TE^S
ANG:	0.7502	0.8426	0.8903	0.5415	0.9677	0.5598
NWL:	0.7634	0.7778	0.9811	0.6600	0.8551	0.7729
SRN:	0.8560	0.8782	0.9719	0.6527	0.9244	0.7062
SVT:	1	1	1	0.7021	0.9769	0.7188
SWT:	0.9716	0.9729	0.9983	0.8594	0.9272	0.9274
TMS:	1	1	1	1	1	1
UUW:	0.9523	0.9660	0.9857	0.9277	0.9752	0.9512
WSH:	0.8314	0.8709	0.9536	0.7711	0.9019	0.8546
WSX:	1	1	1	1	1	1
YKY:	0.9890	0.9890	1	1	1	1
Average:	0.9114	0.9297	0.9781	0.8114	0.9528	0.8491

* Significant at 10%, ** Significant at 5%, *** Significant at 1%.

The results over time are more-or-less consistent over the entire time period, with only one period of significant change in average efficiency scores: in 2015/16, all efficiencies fall on average, with a 6% decrease in overall efficiency, 2% in allocative efficiency, and 5% in technical efficiency, all of which recover to previous average scores in the following year. This sudden decrease in all efficiencies could be a result of the enaction of PR14 regulatory policies, such as the change in focus from Totex costs to Botex costs, but could also be a reflection of the data proxies used in modelling that were affected by this change.

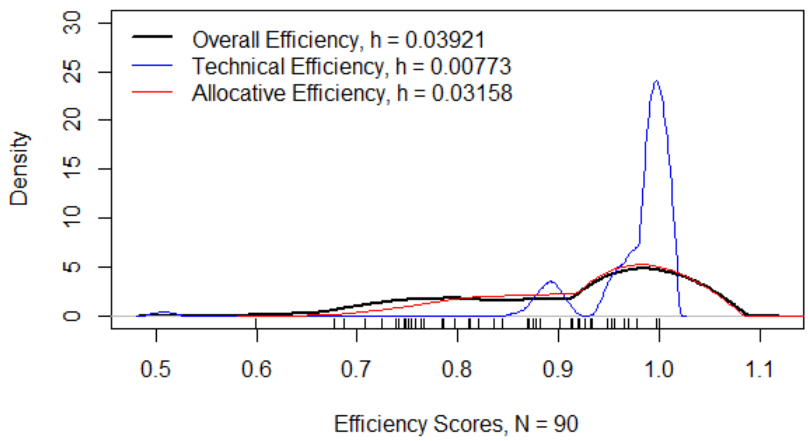
In comparison, the overall and technical efficiency scores are on average significantly lower in the Static models, with a 10% decrease in overall efficiency driven primarily by a 12.9% fall in technical efficiency. Interestingly, allocative efficiency instead increased slightly by 2.31% on average. Table 6.3 provides Wilcoxon-Signed Rank test results between the dynamic and static scores, which shows that though the allocative efficiency change is not significant, both of the remaining efficiencies are significantly different as a result of not treating Capital as a quasi-fixed input.

Table 6.3: Wilcoxon Signed-Rank Tests for Pairwise Efficiency Scores

Model Pair:	Significance Level:
OE, OE^S	0.0001***
AE, AE^S	0.3343
TE, TE^S	0.0000***

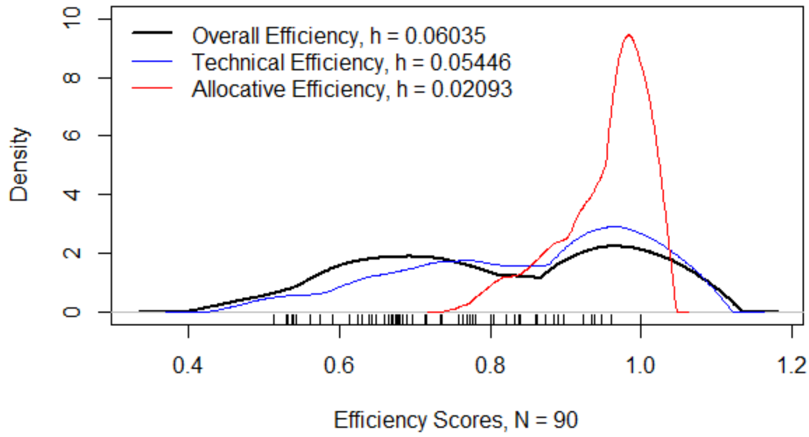
* Significant at 10%, ** Significant at 5%, *** Significant at 1%.

Figure 6.4: 3SDEA Dynamic Efficiency Distributions, 2011/12 - 2019/20



Figures 6.4 and 6.5 provide the same results from the efficiency distributions: the dynamic model PDFs concur with the idea that technical efficiency is almost uniformly close to full efficiency, with an interesting slight bimodality in

Figure 6.5: 3SDEA Static Efficiency Distributions, 2011/12 - 2019/20



the distribution around full efficiency and values close to 0.9. The behaviour is flattened in the static distributions, but traded-off by a uniform increase in the frequency of lower efficiency scores. Instead, the PDF of allocative efficiency reflects the relatively consistently high average allocative efficiency scores in the static models, compared to the large falls in overall and technical scores.

The most unusual part of these results is the relatively minimal changes in allocative efficiency between the two sets of models. On the initial belief that allocative efficiency reflects the presence of Capex Bias, the expectation of the results is that there is a significant drop in allocative efficiency in the static models, which detail the case where Capital is mis-treated and not accurately represented as a quasi-fixed input.

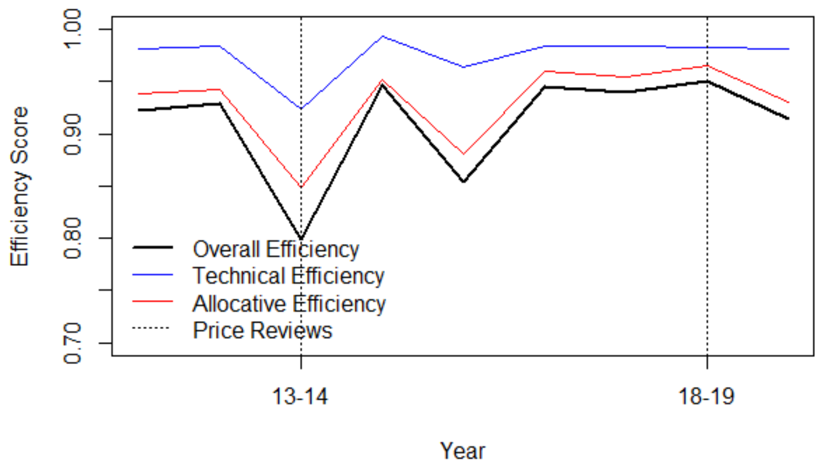
However, the fact is that this expectation was not only not met, but reversed according to these results. This therefore suggests that, on average, Capital is not mis-specified or mis-treated by the industry at large, and is appropriately thought of as something close to a quasi-fixed input.

What of the technical efficiencies? Believing that the high technical efficiency is a mark of good performance from the companies, rather than a poor performance of the models, the aforementioned 12.9% decline in technical efficiency in the static models suggests that, if Capital is not being mis-represented in company production processes, then it might instead be being poorly used in production, while also being considered correctly.

A final point to consider, in line with previous discussion in this chapter, is that the significances and efficiency score behaviours between the models could be due to the ad hoc specification of some of the data used in estimation. This is further compounded by the potential differences between botex-defined total costs and total costs in theory - this could reflect the departure from the more expected results in Pointon & Matthews (2016)'s similar investigation into the industry, who uses totex instead.

6.4.2 CI Model Results

Figure 6.6: CI 3SDEA Average Efficiency Scores, 2011/12 - 2019/20



The same figures and tables have been analogously drawn for the 3SDEA models with the addition of the composite quality indicator as an output. As in the previous section, Figures 6.6 and 6.7 respectively show the average efficiencies over time in the industry, and demonstrate an unusual change in the driving efficiency between the dynamic and static models: technical efficiency appears to

Figure 6.7: CI 3SDEA Static Efficiency Scores, 2011/12 - 2019/20

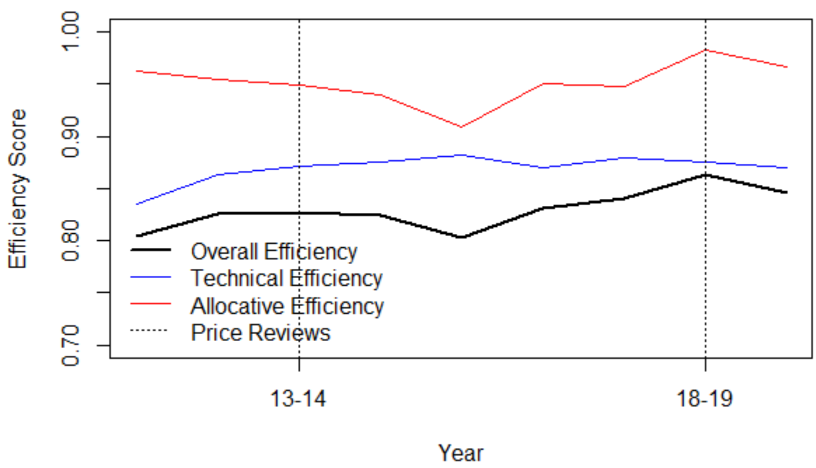


Table 6.4: CI 3SDEA Efficiency Scores, 2011/12 - 2019/20

DMU	OE	AE	TE	OE^S	AE^S	TE^S
ANG:	0.7381	0.8287	0.8797	0.5476	0.9546	0.5740
NWL:	0.8159	0.8322	0.9732	0.8103	0.8813	0.9167
SRN:	0.8278	0.8543	0.9655	0.6433	0.9181	0.7012
SVT:	1	1	1	0.7181	0.9702	0.7402
SWT:	0.9606	0.9631	0.9968	0.8498	0.9207	0.9219
TMS:	1	1	1	1	1	1
UUW:	0.9850	0.9886	0.9963	0.9669	0.9829	0.9837
WSH:	0.8209	0.8209	0.9508	0.7697	0.8982	0.8566
WSX:	1	1	1	1	1	1
YKY:	0.9597	0.9671	0.9910	0.9874	0.9874	1
Average:	0.9108	0.9296	0.9753	0.8293	0.9513	0.8694

* Significant at 10%, ** Significant at 5%, *** Significant at 1%.

drive the overall dynamic efficiency over time, but allocative efficiency becomes that main driver in the static models.

The same behaviour between both sets of models in companies can also be observed, via Table 6.4: Overall Efficiency sees an 8.15% decrease on average when changing from a dynamic model to a static one; allocative efficiency sees little change, increasing on average by 2.17%; technical efficiency again observes the largest change of the three, decreasing on average by 10.59%.

Table 6.5 displays the Wilcoxon Signed-Rank Test results between the dynamic-static pairs of efficiency scores, and again finds that, while both overall and technical efficiencies significantly change between the models, allocative efficiency does not.

Table 6.5: Wilcoxon Signed-Rank Tests for CI Efficiency Scores

Model Pair:	Significance Level:
OE, OE^S	0.0038***
AE, AE^S	0.4631
TE, TE^S	0.0000***

* Significant at 10%, ** Significant at 5%, *** Significant at 1%.

Table 6.6: Wilcoxon Signed-Rank Tests for Inter-Model Efficiency Scores

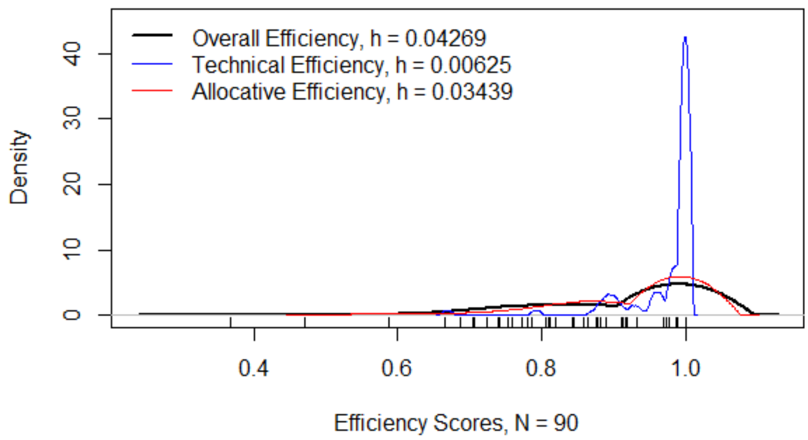
Model Pair:	Significance Level:
OE, $CI\ OE$	0.4075
AE, $CI\ AE$	0.2385
TE, $CI\ TE$	0.4165
OE^S, $CI\ OE^S$	0.2368
AE^S, $CI\ AE^S$	0.5682
TE^S, $CI\ TE^S$	0.3010

* Significant at 10%, ** Significant at 5%, *** Significant at 1%.

Of additional interest is the difference in all models, between those models with the composite indicator and those without. Table 6.6 reports Wilcoxon Signed-Rank test results for all pairs of dynamic and static models. The results are markedly unimpressive: the addition of the composite indicator into the model - any model - does not provide a significant change in efficiency.

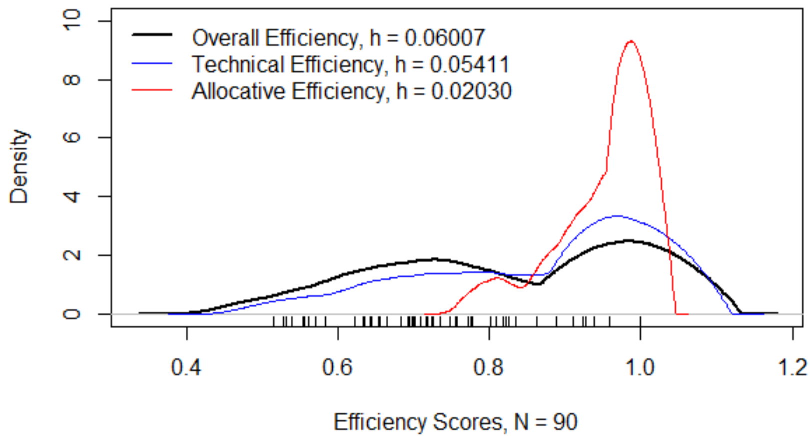
What does this suggest? We can again find both a positive and negative spin on the implications of these tests. On one hand, the fact that there are no significant changes in the scores might imply that the capex bias (pervasive in the industry) is not excused by investing into quality improvements - there are still some omitted reasons that could explain the short-term capital bias, which are currently unaccounted for. On the other hand, that the efficiencies aren't significantly different, but allocative efficiency is consistently high, suggests that, to some extent, the allocations of production factors - particularly capital - are already more-or-less achieved - the lack of significance isn't because of poor capex bias that excludes quality, but companies instead have already allocated correctly for quality improvements, giving high and consistent allocative efficiency while still exhibiting some form of bias, reflected by the significantly less consistent technical efficiency.

Figure 6.8: CI 3SDEA Dynamic Efficiency Distributions, 2011/12 - 2019/20



Figures 6.8 and 6.9 report the estimated density functions of the efficiency scores, for the dynamic and static models respectively. The same distributional patterns emerge here as in the models without the composite indicator: in the dynamic models, technical efficiency ha by far the strongest density of near- or fully-efficient scores, though with less evidence of a second mode in lower scores, relative to its counterpart Figure 6.4. In the static distributions, allocative efficiency again demonstrates the highest density of fully-efficient scores, and overall efficiency shows some evidence of bi-modality by virtue of a significant amount of technical efficiency scores that takes lower values, from around 0.6 to 0.9.

Figure 6.9: CI 3SDEA Static Efficiency Distributions, 2011/12 - 2019/20



The results when adding in the composite indicator as a production output, as well as the insignificance between both groups of model, could also be explained by misspecification due to the ad hoc data used in this chapter. Allocative efficiency in particular requires precision in the definition of the data, to accurately define what is being misallocated. Similarly, given the expectations from Chapter 5 that the composite indicator's addition should cause significant differences in the models, while it could be the case that the specifications of the production models estimated here simply response less significantly to the addition of the quality measure, it is likely that, as with the other results in the chapter so far, that the data issues present in the model, compounded with potential dimensionality issues, have led to the insignificance.

6.4.3 FP vs. NFP Results

The final important point of observation is how the efficiency of primary interest - allocative efficiency - differs in all models between For-Profit and Non-For-Profit companies in the industry. Figures 6.10 and 6.11 show the average allocative efficiency over time for the industry overall, for for-profit companies

Figure 6.10: 3SDEA Dynamic Allocative Scores Comparison, 2011/12 - 2019/20

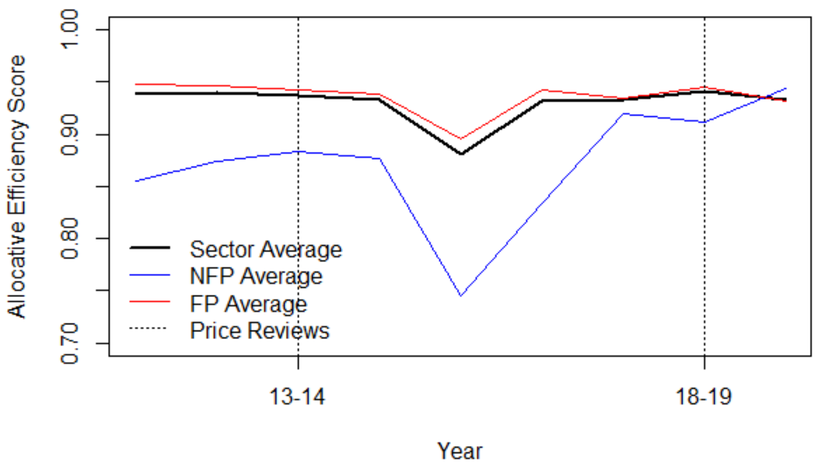
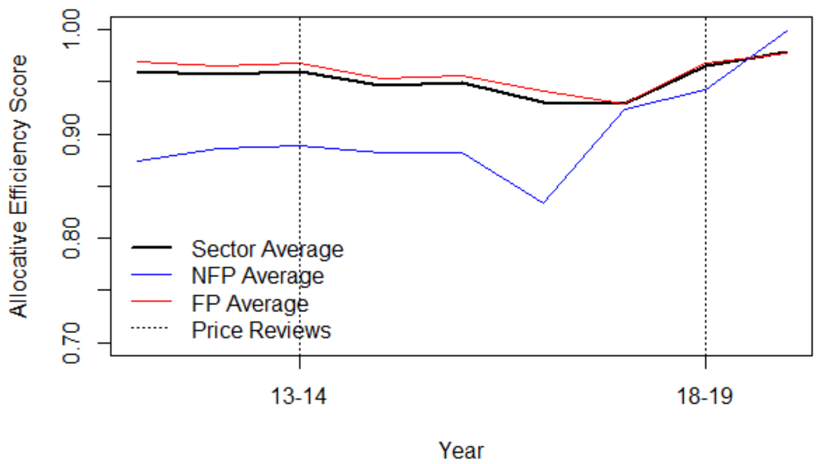


Figure 6.11: 3SDEA Static Allocative Scores Comparison, 2011/12 - 2019/20

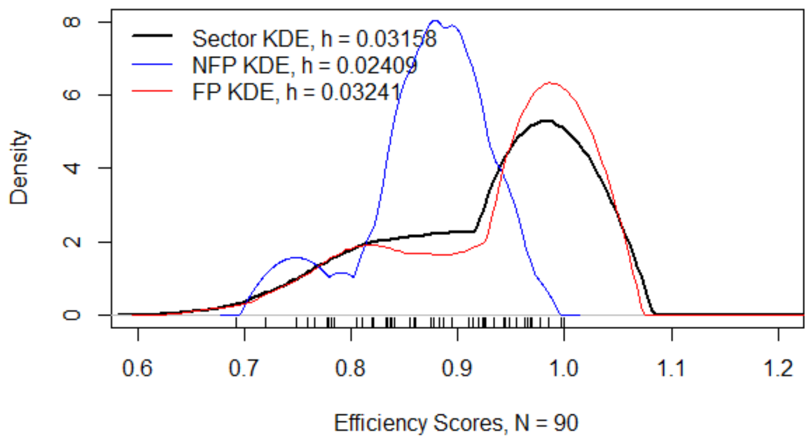


on average, and for non-profit companies - WSH - on average, for the dynamic and static models respectively.

For of the time period, WSH shows a reasonably lower allocative efficiency, compared to the sector and for-profit averages. Interestingly, from 2015 - 2016 to the end of the period, they then see a significant improvement in their allocative efficiency, ultimately ending the period with a marginally higher efficiency that the for-profit companies. This behaviour is consistent in both dynamic and static models, which suggests that, regardless of the treatment of capital, the company has improved how it apportions its production factors in its production processes.

Looking at the estimated PDFs of the scores in the same fashion, via Figures 6.12 and 6.13, can show how the overall distributions of allocative efficiency differs between company type.

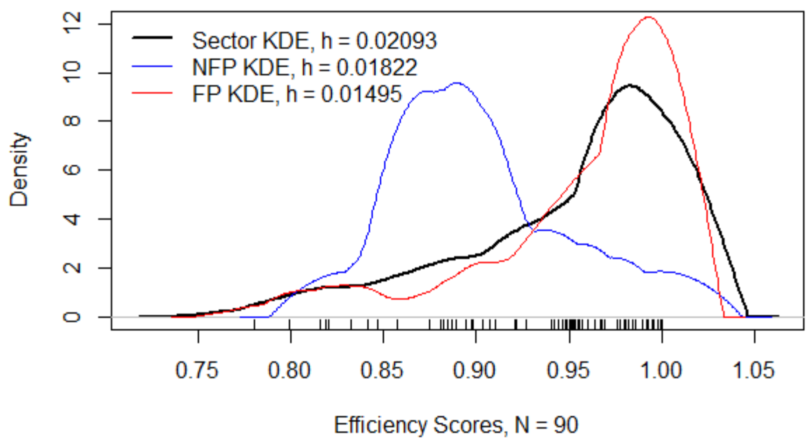
Figure 6.12: 3SDEA Dynamic Allocative Score Distributions, 2011/12 - 2019/20



In both distributions, the modal allocative efficiency of WSH is lower than their for-profit counterparts on average. The distributions in the dynamic models are a similar shape, with WSH's distribution being close to a translation of the other distributions down the efficiency score scale - both have a collection of scores in their left tails, but the for-profit and industry distributions observe a collection of scores from 0.8 to 0.9, before beginning their peak around full efficiency. On the other hand, WSH sees a slight peak around 0.75, with the majority of the scores building the efficiency score mode of around 0.88.

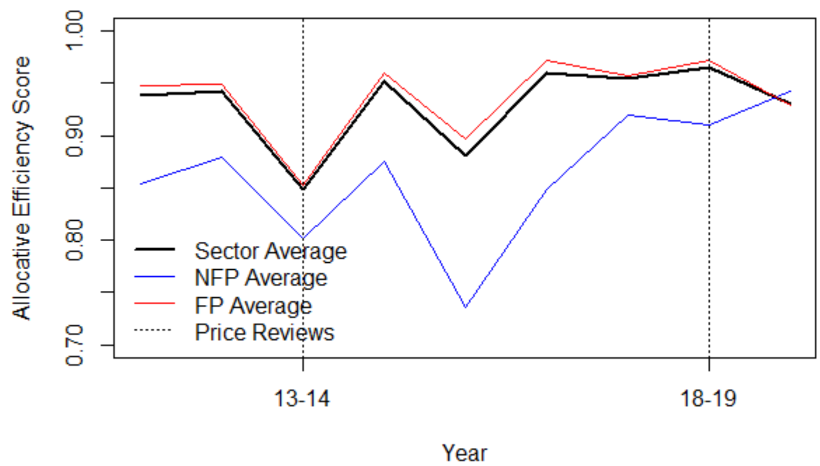
The static models, interestingly, see a significant change in the non-profit distribution of allocative scores: whereas the for-profit and sector distributions remain a similar shape to their dynamic model equivalents, WSH's distribution

Figure 6.13: 3SDEA Static Allocative Score Distributions, 2011/12 - 2019/20



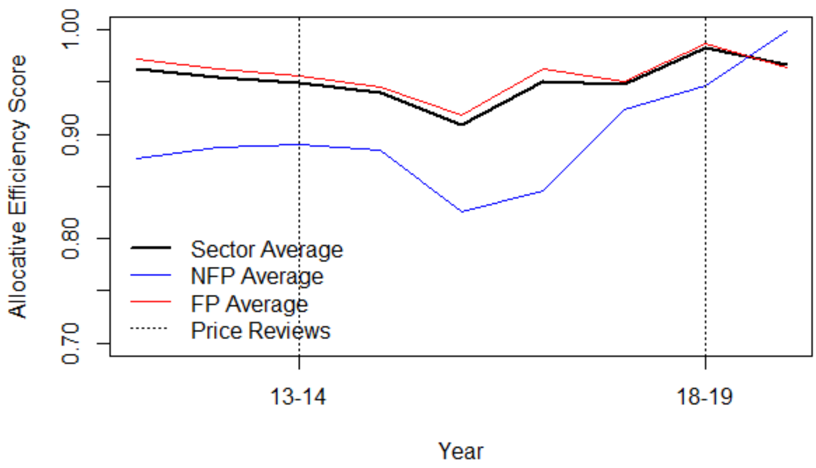
sees something like a reversal of its skew, becoming distribution with roughly the same modal score, but far more values higher than the mode, ranging from 0.9 to full efficiency. This also explains the wider tail of the sector average distribution as well, as there now exists a greater density of nearly-efficient scores.

Figure 6.14: CI 3SDEA Dynamic Allocative Scores Comparison, 2011/12 - 2019/20



As with the previous analysis between the models with and without a com-

Figure 6.15: CI 3SDEA Static Allocative Scores Comparison, 2011/12 - 2019/20



posite quality output, the behaviour of allocative efficiency between non-profit and for-profit companies remains largely the same: WSH sees in general a lower allocative efficiency than other companies until 2015/16, after which it improves and ultimately overtakes the other companies on average at the end of the time period. This behaviour also persists in the static models, with WSH having slightly higher scores on average over the whole time period, as before.

Figure 6.16: CI 3SDEA Dynamic Allocative Score Distributions, 2011/12 - 2019/20

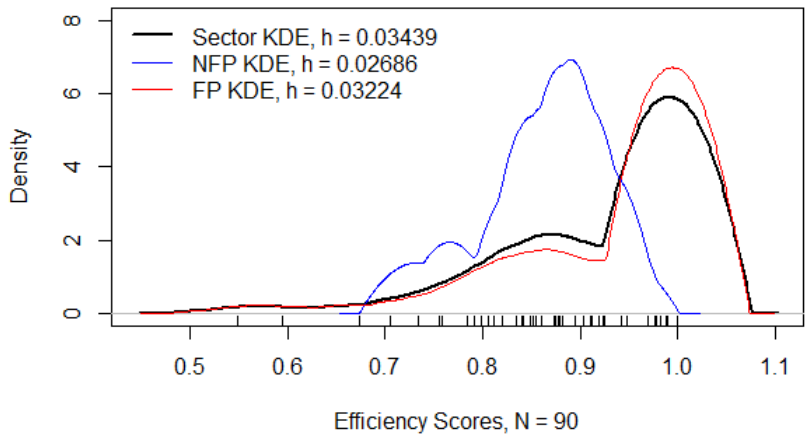
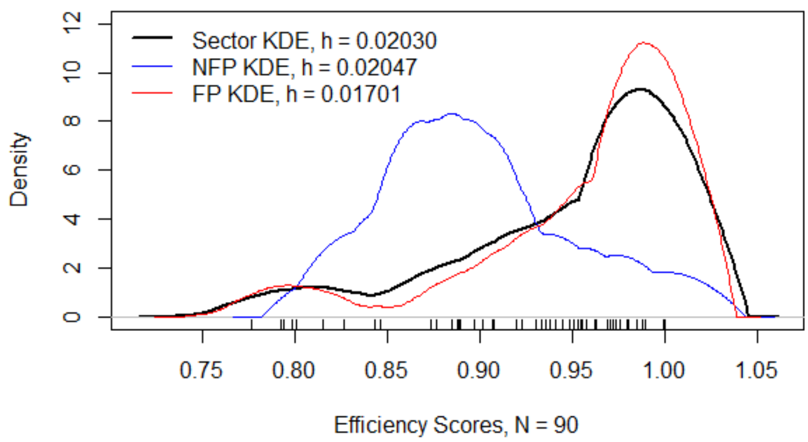


Figure 6.17: CI 3SDEA Static Allocative Score Distributions, 2011/12 - 2019/20



Unsurprisingly then, the distributions of the scores also concur with this evaluation of allocative efficiency. As before, the dynamic models produce allocative efficiency that, for WSH, is largely uniformly lower distributionally than the for-profit and sector distributions, with a slight peak in frequency of scores around 0.75, and a modal peak aroung 0.88.

The 'skew reversal' of the static models is again observed in the static CI models, where the non-profit allocative distribution retains a modal peak around 0.88, but then has a collection of scores to the right of the mode, ranging from the mode to full efficiency. The explanation of the sector's distribution of allocative efficiency's increase in left tail density is also the same, in that the addition of higher non-profit allocative efficiencies drives a higher density of scores in the left tail, close to the modal peak of the distribution.

6.5 Conclusion

In conclusion, the results of these dynamic models with quasi-fixed capital have yielded mixed results. On one hand, the models show a consistent and significant change in technical, and therefore overall, efficiency when comparing dynamic and static DEA models; on the other hand, allocative efficiency remains more-or-less the same on average.

This leads us to an interesting conclusion: It might be, according to this chapter's results, that there are few allocative efficiency issues that cause the ever-present Capex Bias issue. Instead, given that technical efficiency is the most subject to change between dynamic and static models, it may be that, though companies are on average allocating inputs in the correct proportions, they are still using sub-optimal amounts: the capex bias may not be from the behaviour and proportions of captial being mis-allocated, but that the scale of inputs used is instead too much - there is an over-investment into short-term capital that arises from technical inefficiency that drives overall inefficiency, while allocative efficiency remains relatively stable.

The comparison of models with and without Chapter 5's composite indicator of quality yields no significant results. Though perhaps a little less exciting than the alternative, what this suggests in fact supports the previous concluding statement: even with quality investment added in as an additional model output, the allocative efficiency remains consistently high in both static and dynamic models, suggesting that capex bias cannot be excused significantly by considering a focus on quality. Instead, some other source, either unobserved, unaccounted for, or technically inefficient, still gives the persistent industry bias.

In comparing the for-profit and not-for-profit allocative efficiency, it is found that, in general, WSH performs worse than for-profit companies and the sector on average. However, there is a distinct improvement in the latter half of the modeled time period that suggests that the company vastly improves over time, surpassing the allocative efficiency of the sector and for-profit companies by 2019/20. This behaviour is consistent in dynamic and static models, with or without the composite indicator output, with the distributions of allocative efficiency illustrating a change in WSH's allocative efficiency distribution from having a denser left tail to a denser right tail, indicating an increase in high allocative efficiency in the static models, relative to the dynamic models.

The contributions of this chapter are subject to improvement. For example, Nemoto & Goto (2003)'s suite of models also include a model that finds Static and Dynamic Efficiency as components of technical efficiency. Though we can infer from the results how the behaviour of technical efficiency changes between the dynamic and static models, inclusion of these models could provide additional insights otherwise missed in this chapter.

Similarly, Nemoto & Goto (2003) also find the Dual to their dynamic models, which grants them measures of Investment Paths for Variable, Quasi-Fixed, and Net Investment into quasi-fixed inputs. These measures could provide confirmation of the results drawn from this chapter, in particular evaluating whether there is an over-investment into quasi-fixed captial that drives the models' technical inefficiencies.

Another issue that could better the chapter's findings refers to data. As discussed when declaring the data, issues in data collections have led to alternative definitions of data used in this chapter which, while sufficient, are considered sub-optimal. Were the collection of the most appropriate data be doable, the results of the models consequently could be considered better than the results here, and could all potentially be subject to change.

On the topic of the models used, though the dynamic DEA models were a good choice, a future research endeavour could be to instead use the aforementioned Network DEAs, which allow for more nuanced dynamic behaviour. If the same investment paths and statistical comparison can be drawn from network models, greater insight into what drives the industry's inefficiencies - and by extension its capex bias - could be investigated. In a similar vein, the notion of a dynamic composite indicator could also be considered, perhaps modeled as some form of 'quasi-fixed output', on the intuition that improvements into quality are not a short-term project, and are likely to take multiple time periods to complete.

Chapter 7

Topic 3: Additional Research Directions

The last two chapters have sought to address the research question defined at the start of the thesis. This last contributing chapter seeks instead to ask further research questions, as to prompt future research directions in the industry.

The chapter first sets up the current state of the industry, drawing from some the matters highlighted in Chapter 2, as well as some recent examples of industry issues, and how they might pertain to these future research ideas. Then, the chapter examines the scope for the following research questions:

1. How might a Composite Indicator of Quality be further developed for use in the industry?

2. How might a Composite Indicator of Quality be further developed methodologically?

In looking at these questions, a large part of the chapter is devoted to discussing the first new research question, and taking a first-pass investigation into the previously derived composite indicator's relationship with 'Extreme Weather'. The second question, relating to the dynamic of quality indicators, follows some methods for decomposing a Composite Indicator into constituent dynamic parts to reflect various causes for change over time. The chapter then concludes by discussing various other ways in which this thesis' indicator could be improved and iterated upon.

7.1 Current State of the Industry

To begin this chapter on a dour note, three quotes, which all relate to same facet of quality, can be compared:

> "By 1870 . . . Sewage was treated before disposal into watercourses . . . in only 46 out of the 178 towns." – *Hassan J. (1998, pg. 26)*

> "The UK has twice been prosecuted by the European Court of Justice over failure to fully implement EEC directives relating to water quality." – *Howes R., Skea J. & Whelan B. (1997, pg. 76)*

> ". . . 373,000 cases of sewage discharge were reported in 2021, even before this year's heatwave." – *Jenkins S., The Guardian, 22/08/2022*

These quotes paint an albeit selective picture of the state of the industry - despite some of the problems that had led to privatisation in the first place being resolved, such as the water and wastewater quality EC directive mentioned in the second quote, the same issues persist now as they have always done, or at least in a state recognisable in the 1870s.

To be more specific, the picture this paints is one that suggests that each main objective of quality in the industry - that is, the betterment of water quality and wastewater quality - is more complicated than the resolution of EC directives of the 1980s, a matter whose completion is reflected by the older measures of quality being almost uniformly at one hundred percent compliance for all companies. To this end, the introduction of the relatively new CPCs, and their use as parts of quality measurement going forward, is a prudent step to take.

Though perhaps a little late in to the thesis to defend again, it is nonetheless this need for wider coverage of over-arching quality issues that presents

Composite Indicators as a sensible direction forward in the topic of empirically measuring quality, as evidenced by Henriques et. al. (2020), D'Inverno et. al. (2021) and Yakymova et. al. (2022), whose papers appeared concurrent to this project.

As it pertains to the composite indicator from this thesis, many of the industry's current issues can relate to the CPCs, and hence parts of the indicator. As will be further explored in this chapter, some of the contemporary issues in the industry are exacerbated by recent droughts and floods. Take, for example, Hosepipe Bans: water companies, foreseeing drought weather and an increase in the demand for water, may issue bans on the excessive use of hosepipes, as to preserve water usage at a household level. One might relate this to good Drought Resilience perhaps, but the need to enforce responsible water use might instead be a failing of another CPC - the target to reduce the Average Consumption per Capita of water.

Another example is aforementioned sewerage discharge issues, which relate to Pollution Incidents in the data. Flood weather, and a lack of Flood Resilience, will inflate the damage and number of these incidents on account of increased water flow in places such as Combined Sewerage Overflows. The same extreme weather, where companies lack the preparation to deal with the weather, can cause increased Leakage and Mains Bursts - issues with the water infrastructure.

Other issues in the industry, not unrelated to the measurement of CPC factors such as leakage or pollution incidents, is that of consistent monitoring in the industry. Looking specifically to rivers as an example, though there already exists a great deal of water monitoring taking place, through the guidance and initiative of the EA in England and the NRW in Wales, various issues still remain unresolved. For example, though the industry does well at monitoring the issues set out by the Water Frameworks Directive (WFD), emerging threats in the water such as Microplastics or pharmaceutical waste, remain unaccounted for.

Another more specific example is the River Wye, wherein a report by UK Parliament (2022) note a tremendous level of phosphate pollution in the Wye catchment, primarily onset by agriculture, but secondarily by water companies. In this case study there is an active example of failure to monitor Pollution Incidents, then, resulting in potentially catastrophic damage to the river.

All of these issues, while not exhaustive, help to frame what the remainder of the chapter seeks to investigate. The notes of the prevalence of extreme weather's impacts on the industry's regulatory targets raises in interesting question: referring to this thesis' composite indicator of quality, does extreme weather have any correlation with industry quality? In theory, there may well be an effect on quality overall, on account of the strain droughts or floods cause on other targets such as leakage or pollution incidents.

The remainder of the chapter will look at the interactions of quality and extreme weather first, and then cap the chapter's broader theme of considering future research with composite quality measurements, with a discussion of next steps to improve the design and analysis of the quality indicator.

7.2 Composite Quality and Extreme Weather

The focus of the extension of quality for this chapter is how the composite indicator interacts with the weather.

The quotes in the earlier contextual section of this chapter highlight that some of the industry issues, despite various attempts to solve them, remain a constant over time. In particular, these quotes refer to the persistent issues surrounding Pollution Incidents in the industry. So, a good question to ask is what factors might still cause these problems to arise.

One such factor is the weather. With the increase in global warming and climate change, incidents of extreme weather are becoming more frequent. Quoting the National Academies of Sciences, Engineering, and Medicine (2016):

> "The observed frequency, intensity, and duration of some extreme weather events have been changing as the climate system has warmed... (W)arming is expected to increase the likelihood of extremely hot days and nights ... Warming also is expected to lead to more evaporation that may exacerbate droughts and increased atmospheric moisture that can increase the frequency of heavy rainfall and snowfall events." - *Committee on Extreme Weather Events and National Academies of Sciences, Engineering, and Medicine (2016), pg. 1*

Given these changes in extreme weather patterns, it stands to reason that water industry quality ought to have measures of resilience against extreme weather. Some measures already exist in the form of CPCs such as the risk of sever restrictions in a drought, and the risk of sewer flooding in a storm. However, the resilience of this thesis' composite quality indicator remains untested. This section, as mentioned previously, will, then, assess how the composite indicator, and its component quality factors, interact with measures of extreme weather; the aim of this exploration is, therefore, one of determining whether this new quality measure is, or has the potential to be, resilient to climate change and its consequential changes in weather.

7.2.1 Weather Data

For England and Wales, the Met Office provides frequent data on a selection of weather statistics. The two weathers used in this chapter is Mean Temperature and Total Rainfall, collected monthly for the whole of England and Wales. This examination of the quality and weather interactions will be limited to the overall England and Wales values, which is equivalent to looking at the sector-wide average behaviours. Data for these measures of weather have a long time window, with the data for Rainfall and Mean Temperature beginning in January of 1836 and 1884, respectively. Table 7.1 provides a table of summary statistics for both variables.

Table 7.1: Rainfall and Mean Temperature Length, T, and Descriptive Statistics

	T	$Mean$	$Std.\ Dev.$	$Min.$	$Max.$
$Rainfall,\ mm$	187	74.93	34.47	3.9	209.8
$Mean\ Temp.,\ C$	139	9.11	4.57	-2.3	19.1

7.2.2 Defining Extreme Weather

So, what defines extreme weather? An intuitive, if overly literal, approach to take is to define outlier data points as 'extreme', by some means of outlier detection. There are various ways to achieve this, and a few are considered here:

Inter-Quartile Range, which defines an observation if it falls outside of the $(P_{25} - 1.5IQR, P_{75} + 1.5IQR)$ range.

Winsorisation, which defines an outlier if an observations falls outside the $(P_{\xi}, P_{1-\xi})$ range, for some quantile $\xi \in (0, 1)$.

Cook's Distance, which detects Influential Points of a dataset, which are similar to outliers.

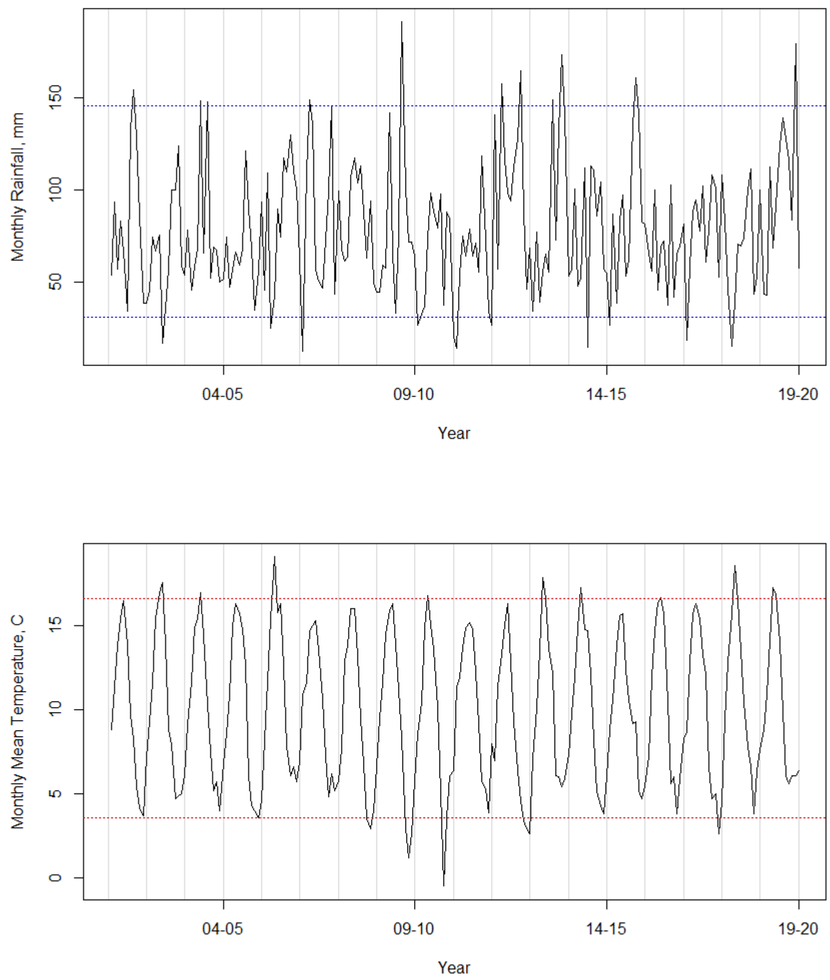
Where P_q is the qth Percentile of the data. In practice, only Winsorisation is chosen for this analysis, as no influential points were detected by Cook's Distance, and negligible outliers were detected using the Inter-Quartile Range.

More formally, then, an Outlier in this model is defined as follows, for an observation $x_i \in X$:

$$x_i \in X \text{ is an Outlier if } x_i \notin [P_5, P_{95}] \tag{7.1}$$

Full application of Winsorisation would then censor any outlier in the data to the appropriate floor or ceiling value of P_5 or P_{95}, but here the method is used only to detect which values are outliers instead. Initially, the threshold past which outliers are determined is found using the years appropriate to the samples of the previous chapters, that is from 2002/03 to 2019/20. However, in accordance with the years defined by the industry APRs, the years in the weather are re-defined to be from April to March, and so an industry year is, for example, defined as April 2002 to March 2003. Figure 7.1 illustrates the data for both variables, with the outlier thresholds shown for both in their respective plots.

Figure 7.1: Short-Form Outlier Detection for Monthly Weather Data, 2002/03 - 2019/20



For the analysis of quality and extreme weather, yearly data is used. To achieve this in a naïve sense, dummy variables can be defined as follows:

$$
\begin{aligned}
ER_t &= \begin{cases} 1, \text{ if any outlier } Rainfall \text{ in year }, t \\ 0, \text{ otherwise} \end{cases} \\
ET_t &= \begin{cases} 1, \text{ if any outlier } MeanTemp \text{ in year}, t \\ 0, \text{ otherwise} \end{cases}
\end{aligned}
\tag{7.2}
$$

That is, the variable indicating extreme rainfall, ER_t, takes a value of one if, for any month in the year t, the data is found to be an outlier according to the winsorisation thresholds, and is zero otherwise. The same explanation is analogous for ET_t, which indicates extreme mean temperature.

7.2.3 Results

As this is primarily an exploratory angle, the analysis is limited to looking at the correlations between these extreme weather measurements, the composite indicator of quality, and the input factors of the indicator.

Table 7.2: Short-Form Correlation Coefficients

	CI	$TotalComplaints$	$PollutionIncidents$	$Leakage$
ER	0.1564	-0.1662	0.3532	-0.1212
ET	0.0956	-0.2561	-0.0427	-0.1541

* 10% significance; ** 5% significance; *** 1% significance

Table 7.2 displays the correlations as described above, and from these a fairly spurious conclusion can be drawn. All correlations are statistically insignificant, and the principal correlations of interest - that of the extreme weather dummies and CI - suggest that years of extreme weather are positively correlated with increases in composite quality, which is counter-intuitive to the notion that quality often falls due to extreme weather. Furthermore, looking at the individual input factors of the indicator, their correlations also seem unusual: in years of extreme, Total Complaints and Leakage appear to fall, whereas Pollution Incidents increase in years of extreme rainfall, but functionally do not change in response to years of extreme temperature. The increase in pollution incidents could be attributed to the increase in combined sewer overflows incidents, where wastewater flows from the sewers into water sources, due to a rise in sewer water levels.

Table 7.3: Short-Form Correlations for different CI Weights

	ER	ET
$Default$	0.1564	0.0956
$Unbound$	0.1478	-0.1354*
$Equal$	0.1417	0.5359**
$Common$	-0.1940	-0.1288

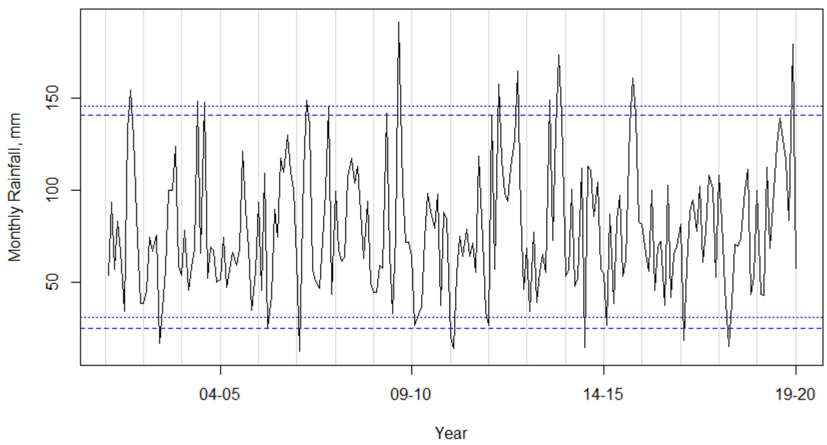
* 10% significance; ** 5% significance;
*** 1% significance

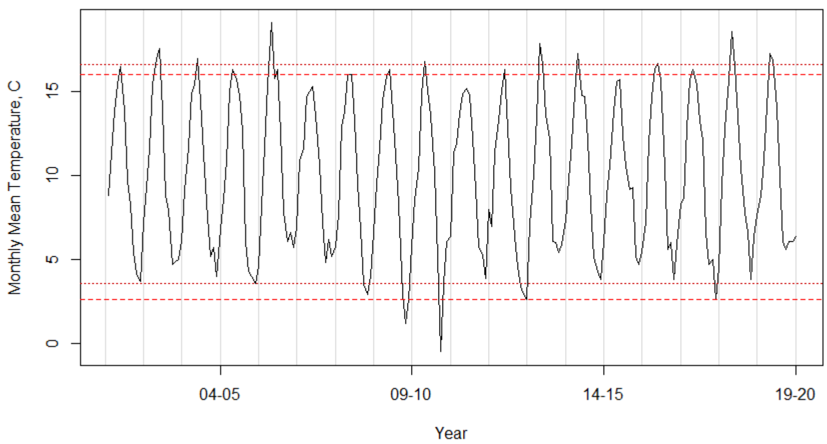
To see if these oddities are robust, Table 7.3 also reports the principal correlations for all of the different Weights specifications that Chapter 5 defined for the composite indicator, to determine if the make of the indicator is perhaps at fault. These results offer little clarity compared to the results beforehand: with Unbound BoD and Equal weights, CI correlates positively with extreme rain, and then negatively when Common Weights are used instead; under Unbound BoD and Common weights CI correlates with extreme temperature negatively and with 10 percent significance when with Unbound weights, and correlates positively with 5 percent significance when equal weights are used.

To try and re-consider these behaviours, a longer time period is considered when detecting outliers. Analogous to the definition of (7.2), two new dummies are constructed, ER_t^l and ET_t^l, which are defined for the same sample time frame, but whose outliers are defined by winsorisation of the data from 1884 to 2020. Figure 7.2 illustrates the data again, with the addition of the new long-form outlier thresholds.

The new long-from thresholds are uniformly lower than the in-sample, short-

Figure 7.2: Long-Form Outlier Detection for Monthly Weather Data, 2002/03 - 2019/20





form thresholds used previously, which suggests that, for both variables, there has been a general upwards trend over time, which means that the data of the sample are far more likely to be caught as outliers, compared to the short-form detection.

Tables 7.4 and 7.5 show the correlations between the long-form dummy variables and the various quality representations. This time, the correlations

Table 7.4: Long-Form Correlation Coefficients

	CI	$TotalComplaints$	$PollutionIncidents$	$Leakage$
ER^l	-0.2565	-0.0893	0.0743	-0.4405*
ET^l	-0.1820	-0.3376	0.1193	-0.1687

* 10% significance; ** 5% significance; *** 1% significance

Table 7.5: Long-Form Correlations for different CI Weights

	ER^l	ET^l
$Default$	-0.2565	-0.1820
$Unbound$	-0.3465	-0.3161*
$Equal$	0.0161	-0.2198**
$Common$	-0.4488**	-0.2313

* 10% significance; ** 5% significance; *** 1% significance

between CI and both extreme weathers, while not statistically significant, are both moderately negative. This behaviour is mostly robust to changes in weights specifications, with the one non-negative result being sufficiently close to zero as to be considered ambiguous. The relationships between the input factors are more consistent as well, but still offer unusual directions, with leakage and total complaints having the expected negative correlation with both extreme weathers, and pollution incidents seeing a positive, if small in magnitude, correlation instead.

To check one other facet of the analysis, an alternative correlation statistic is used instead. The Point-Biserial Correlation Coefficient is possibly more appropriate than the former standard correlation measures, as it is specifically for correlations between one continuous and one binary variable. The statistic, and its corresponding test statistic, are defined as:

$$\begin{aligned} r_{pb} &= \frac{M_1 - M_0}{s_n}\sqrt{\frac{n_1 n_0}{n^2}} \\ Z_{pb} &= r_{pb}\sqrt{\frac{n_1 + n_0 - 2}{1 - r_{pb}^2}}, \ \sim t(n_1 + n_0 - 2) \end{aligned} \tag{7.3}$$

Where M_1 and M_0 are the Means of those observations with a valuable of the binary variable of 1 or 0, respectively; n_1 and n_0 are the analogous Sample Sizes, whose sum totals to n, and s_n is the full-sample Standard Deviation. The resultant test statistic lies on a t-distribution with $n_1 + n_0 - 2 = n - 2$ degrees of freedom.

Table 7.6: Point-Biserial Correlation Coefficients

	CI	$TotalComplaints$	$PollutionIncidents$	$Leakage$
ER	0.0245	-0.0734	0.1710**	-0.0117
ET	0.0149	-0.1130	-0.0206	-0.0148
ER^l	-0.0589	-0.0173	0.0024	-0.0025
ET^l	-0.0285	-0.1490**	0.0578	-0.0162

* 10% significance; ** 5% significance; *** 1% significance

Table 7.6 report the point-biserial correlations for both short- and long-form extreme weather dummies. The results are again somewhat ambiguous, with correlations between CI and the extreme weather dummies being insignificantly positive, and then insignificantly negative, for the short-form and long-form extremes respectively; total complaints and leakage are still consistently negatively correlated, with middling significance between total complaints and long-form extreme mean temperatures. Pollution incidents again provide a relatively intuitive direction, remaining positive or about zero, with a significant positive relation with short-form extreme rainfall likely owed an increase in combined sewer overflows incidents.

7.2.4 Future Directions

A great deal can be done to improve the modelling of weather and quality index interactions[1]. First and foremost to improve is the currently naïve definition of extreme weather. From a more meteorological perspective, extreme weather is better defined as persistent high or low weather observations, rather than the currently defined case that calls for extreme weather if an outlier occurs at least once in a year. To that same end, another improvement would be the use of

[1]With particular thanks to the people who gave feedback to, and ideas for, my presentation of this work on the 12th of October, 2022.

finer data - daily observations, rather than monthly averages. That way, the persistence of weather can be more accurately assessed, and a better measure of extreme weather can be drawn. Were such an improvement made, the resulting extreme weather variables would likely be intensities of extreme weather in a given year, defined in some fashion akin to:

$$ER_t := \frac{\#Obs.ExtremeWeatherPatterns}{Total\#WeatherObs.}, \in [0,1]$$

Where, for example, Extreme Rain intensity for a year t is defined as the total number of observations in a year that form patterns of extreme rain, as a proportion of the total number of rainfall observations in that year.

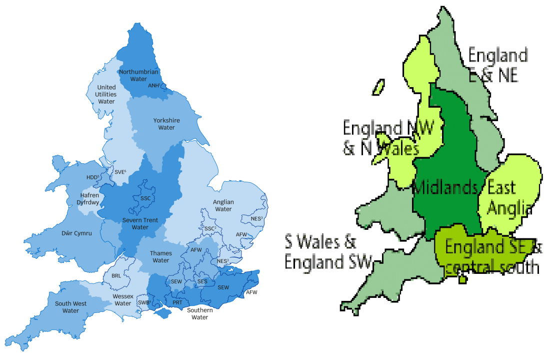
Another interesting angle to consider, to better measure when weather can be defined as extreme, refers instead to a historical perspective. In looking at historical data, one could infer patterns of extreme weather from instruments such as Harvest Yield and Infant Mortality - the former reflects poor crop yield potentially because of extreme weather; the latter reflecting the dangers that arise from extreme weathers to individuals' health, particularly children. Though these measures aren't used in modern meteorology, the patterns of the historical instruments could be extrapolated forward, to reflect a sort of 'technology unadjusted' instrument reflecting extreme weather patterns. Similar to the previous suggestion, an intensity variable, whose measure is reflected from the instruments, could be defined as something like:

$$\begin{aligned} &ER_t := \frac{\#\hat{O}bs.ExtremeWeatherPatterns(X)}{Total\#WeatherObs.}, \in [0,1]\,, \\ &\#Obs.Ext. = f(HY_{ex}, IM_{ex}, C; \beta) + \varepsilon \end{aligned}$$

Where HY_{ex} and IM_{ex} would be the Extrapolated Harvest Yield and Infant Mortality, respectively in this example. These figures, found in essence as counterfactuals to the actual values, could then be used with other Characteristics C to estimate the number of extreme rainfall observations in year t, which then is measured relative to the actual total number of rainfall observations.

Similar to the notion of finer weather observations, one key naïvety assumed in this chapter is that these weather observations are homogenous across England and Wales - the data used is the average rainfall or temperature over the entire geographical sample. In reality, there will be heterogenous weather patterns across the sample region and, therefore, an improvement that could be made would be to incorporate regional-level observations, which could with sufficient matching be used to effectively create 'DMU-specific weather'. The difficulty of achieving this, however, can be best expressed by Figure 7.3, which displays the rather different regions captured by DMUs and by the Met. Office, respectively.

Figure 7.3: Ofwat (left) and Met. Office (right) Geographical Regions



As can be observed by Figure 7.3, the regions each WaSC oversee and the meteorological regions of England and Wales do not match - this difference in regional definitions, and the consequent difficulties in reconciling these regions for regional analysis, are a significant contributor to the use of national-level, industry-wide analysis of the interactions of weather on quality in this chapter. Future research should consider the regional impacts of weather on quality, by WaSC operating region, and to achieve this one strategy for data collection could be to aggregate local or sub-regional data into the appropriate regions, requiring high-detail information on weather events in England and Wales, as well as complete information of the constituent areas of each WaSC.

Some other data sources, such as the National River Flow Archive's catchment rainfall data (https://nrfa.ceh.ac.uk/rainfall-data) or the UK Government's Historic Flood Warnings data (https://www.data.gov.uk/dataset/d4fb2591-f4dd-4e7f-9aaf-49af94437b36/historic-flood-warnings) could be useful here[2].

Speaking more holistically to this exploration's idea, a further development could be to form some sort of theoretical structure around the behaviours of quality and extreme weather. By doing so, there could then be an operationalised empirical model that more easily reflects the relations between extreme weather measures and the composite indicator or its factors. Extending from this idea, some form of time-series analytics could also be employed to better understand how the weather data, its extreme patterns, and its interactions with quality, behave, though this does not require the aforementioned theory to be used.

On a similar topic, as it tenuously pertains to VAR modelling, the relationships between extreme weather patterns and quality could be better investigated, with additional insight gleaned via qualitative data, such as surveys on the perception of quality and appropriate definitions of extreme weather, or more anecdotal evidence on the relationships of weather and quality, such as the belief that total complaints may increase seasonally in the summer, when the use of water increases in response to hotter weather. This angle of evaluation may, in tandem with the previous quantitative improvements, add more detail to the behaviours of the extreme weather and quality, providing more accurate and more realistic conclusions on companies' resilience to extreme weather.

A final point of discussion in this exploration of quality and extreme weather interactions is in reference to the composite indicator itself, and how its inputted data may influence the resulting correlations with extreme weather. As the thesis has discussed, the composite quality indicator contains quality factors from various parts of the industry, as to reflect a form of overall quality. While

[2]I would like to thank my viva examiners for providing directions to these alternative data sources.

advantageous in earlier DEA analysis due to the composite indicator's quantity of information and small dimensions, in this chapter's analysis the fact that only one small facet of, say, wastewater quality measurement has been included may have lead to the relatively inconclusive results. Future research using composite indicators for quality measurement in the industry may benefit from additional data for both water and wastewater quality, such as other CPCs not used in the indicator's construction (see Table 4.2). As will be discussed shortly, one method that may allow for this additional information without compromising the BoD method's dimensionality issues could be to create sub-indicators within the final composite indicator.

7.3 Further Composite Indicator Development

Chapter 5 introduced the notion of a Composite Indicator of Quality and, in doing so, sought to create a quality index that had some factors of quality from each of the general facets of industry quality: water environmental quality, wastewater environmental quality, and service quality. To do so, some of the array of CPCs were used as measurements. However, a trade-off in the indicator's design was made to account for the relatively long timespan covered by the DEA models in the chapter, which spanned eighteen years of data. As Chapter 4 describes, most of the CPCs that could have been useful metrics for quality were discarded on the grounds of a lack of data - most CPCs have only existed since their formal introduction in PR14, whereas the data ultimately used in the indicator had been previously measured throughout the time period used in the models.

The design of the indicator was such that the indicator could be relatively simple and well-justified from a mechanical point-of-view, with specific attention drawn to the weights of the inputted quality factors. The BoD method of weight selection circumvented any designer-side biases in the weights by mathematically solving a linear program for the weights, with the only designer restriction assuming that each weight for each factor had a lower bound of one over the total number of CPCs plus the service quality measurement. Though this led to defensible results, it is valid to consider also whether these weights should have left to an optimisation problem alone, or if other choices about them should have been made by the designer.

This section will look into greater detail how the composite might be iterated upon, as to be as in-depth and as accurate as possible, while also retain a degree of simplicity in its interpretation for the sake of use in policy or further modelling, such as in future DEA models like in Chapters 5 and 6.

7.3.1 Data Selection and Nested Indicator Design

Assuming away all of the practical restrictions that existed in the previous chapters, an important question to ask is on the topic of data selection: if there are no restrictions on the choice of variable due to the time period of the indicator, which variables should be chosen?

A blunt answer might be to select all of the variables that were otherwise valid choices for quality measurement. In this case, drawing on the discussion in Chapter 4, the number of quality factors may increase from three to five, or to seven if the Water and Treatment Quality Compliance commitments were also considered viable. Rather quickly, it seems, the amount of factors for the index increases - more so if further service quality measures, as in Molinos-Senante et. al. (2015b, 2016, 2017b) are also applied. This leads quickly to a few issues, the first of which is familiar to the already-discussed issue of dimensionality frequent in small-sample DEA models. Since the BoD weighting method is a DEA-type model, the quantity of input factors for the weighting program far exceeds the Cooper's Rule recommended amount, thereby leading to inaccurate results that are no longer tenable.

Following Cooper's Rule, then, the maximal number of input factors for the BoD weights should still remain at three; how, then, can all of these other factors be implemented in an effective manner, if it believed that their addition better relates actual industry quality behaviour than their exclusion?

A good answer to this problem is the idea of Sub-Indicators, which was seen in the Porter & Stern (2001) case study in Chapter 5. This idea quite readily allows for this supposed larger set of quality factors to be grouped and use in separate sub-indicators, which are themselves components for the final composite indicator. Choosing this approach to larger sets of factors of interest raises further questions related to what would be a significantly more complex design process:

- How should quality factors be sub-divided to produce sub-indicators?
- How should the sub-indicators be designed, and should they all take the same design?
- What is the design process for the final composite indicator, given the use of sub-indicators?

Some of these issues are addressed more in the next section, but each question above warrants investigation. The first question of grouping variables seems trivial, given the emphasis of the thesis of having representative environmental water, environmental wastewater and service quality factors, but with the addition of further quality factors, the grouping could differ from the default idea. With inclusion of Water Compliance, Treatment Compliance, and Households above Reference Pressure - as used in Molinos-Senante et. al. (2015b, 2016, 2017b) - one could instead create a 'Compliance' quality group, taking each factor out the previously defined water, wastewater and service groups. Similarly, Leakage, Pollution Incidents, and Total Unplanned Interruptions, could form something like an 'Infrastructure' quality group.

This complication feeds directly into the next question, on the matter of sub-indicator design. A default option might be to repeat this thesis' design for sub-indicators, at it finds tenable results and has, hopefully, been sufficiently justified as a choice of design for an indicator. However, as the question continues to mention, should each sub-indicator have the same design? Depending on the options for grouping as mentioned above, perhaps certain combinations of quality factors will not produce sensible results with the current design specification. Furthermore, the assumption that each quality factor is independent is unlikely to hold in practice, and so if groups of two groups of quality factors contain members that highly correlate or are dependent on the same underlying process, does this confound the sub-indicator design process with the need to address interactions between factors, and so between sub-indicators?

This kind of problem also extends to the last question, that of the design process for the final composite indicator. One solution is to take an unweighted arithmetic mean of the sub-indicators, but that assumes that all sub-indicators are of equal importance. If each sub-indicator is represents the water, wastewater and service quality groups, for example, is it sensible to weight them equally in the final result, or are they weighted by the contribution of the sector they represent to total costs? If groups are differently defined, such as with the compliance and infrastructure quality example groups, how are the indicators weighted then?

All of these parts to the argument of sub-indicators circle around a principle concern in indicator design overall: the trade-off between depth and accuracy, and ease of explanation and evaluation. I believe that the indicator ought to be iterated on in the manners discussed here and in the following sections, but it quickly becomes a crucial part of the process to balance any greater technical complexities with the ability to have the final result still be understandable.

7.3.2 Weighting and Aggregation

In a continuation of the balancing act of depth versus transparency, the matters of Weights and Aggregation as a whole arise as important factors of the process to discuss. For as much as the whole process of the indicator's design is important, the weights and summation of the factors into the outputted indicator values are, perhaps, the crux of the design process, short of selecting the data to use in the first place.

The indicator, as designed in Chapter 5, utilised the BoD method of weighing its inputted quality factors, as to try to weigh the factors mathematically and therefore without any bias from designer choices. Therein, as was already mentioned, the only designer-made restriction on the weight model was the inclusion of a lower bound, such that all factors had non-zero weights that were at least equal to one over the total number of CPCs plus service quality measures used, $w_l = 1/15$. This assumption was based on the idea that, in reality, all companies should be paying some attention to all of these measures in some capacity, even if the weights end up choosing one factor to prioritise, as was frequently the case in the results.

But the ever-present question in these discussive parts of this chapter still remains - is this good enough? Referring first to the BoD methodology, it can be noted that, by definition, the BoD method used in this project's composite indicator is considered 'Optimistic', in that it maximises the indicator such that the indicator cannot exceed a value of one. As Zhou et. al. (2010) define it, this version of BoD optimisation effectively finds the 'best weights for the indicator, and so by definition there can also exist a 'worst' set of weights, leading a 'Pessimistic' indicator instead. They conclude their indicator development by defining a mixed final indicator:

$$CI_{i,t} = \lambda \frac{CI_{i,t}^{opt} - \min(CI_{i,t}^{opt})}{\max(CI_{i,t}^{opt}) - \min(CI_{i,t}^{opt})} + (1-\lambda) \frac{CI_{i,t}^{pess} - \min(CI_{i,t}^{pess})}{\max(CI_{i,t}^{pess}) - \min(CI_{i,t}^{pess})} \quad (7.4)$$

Which is a weighted sum of the min-max-normalised optimistic and pessimistic CIs, respectively, for each company i in each year t. An immediate extension of this model addresses the definition of $\lambda \in (0, 1)$, the weighting parameter, which is assumed by default to be uniform for all DMUs in all sample years. By some measure or instrumentation, a natural extension is to have this final weighting depend on companies and time, giving $\lambda_{i,t} \in (0,1), \forall i, t$, instead.

Zhou et. al. (2010) cover other areas of potential improvement, as it pertains to this project's indicator's weighting and aggregation choices. One idea is that of Relative Importance Weights, as in Cherchye et. al. (2007b). Paraphrasing their definition of this, the relative importance weights, $\omega_{q,i,t}$, are defined as:

$$\omega_{q,i,t} := \frac{w_{q,i,t} I_{q,i,t}}{\sum_{q=1}^{Q} w_{q,i,t} I_{q,i,t}}, \omega_{q,i,t} \in [w_l, w_u] \tag{7.5}$$

That is, the relative importance weight of factor q for DMU i in year t is determined by the relative contribution of the weighted factor to the total sum of the the weighted factors, the CI, such that the importance is within an upper and lower bound, w_u and w_l. This definition would be a good alternative the the BoD with a lower bound used in this project, as is still allows for boundaries to be set on the importance in the final CI, now defined with the importance weights, thereby meeting the assumptions that all factors are at least somewhat important to the measure of quality. By similar extension, the use of an upper bound could mitigate the consequences of BoD vastly prioritising one factor over all others, leading to a potentially more parsimonious mathematical choice for factor weights.

Both Cherchye et. al. (2007b) and Zhou et. al. (2010) also argue via their initial designs for a Geometric indicator, rather than a Linear one. In a similar argument to that of the arithmetic and geometric means, both papers believe that a geometric indicator presents a slightly more neutral definition than the arithmetic alternative, though the latter can also be reached by log-linearising the indicator. An argument arises, though, from the resulting slight complexity over the arithmetic mean, which is far more recognised as an average measure. Further, without log-linearising the model, the determination of these weights becomes a non-linear optimisation problem, thus creating more technical difficulty as it pertains to programming and therefore finding the optimal weights.

A last idea to iterate on the BoD weighting method follows Wang (2015) who incorporates Slacks into the BoD design. In effect, this inclusion of slack components allows for the weights to be optimised, without necessarily being totally maximised - there may be some non-zero optimal slack for a quality factor, if mathematically over-investing into the factor is sub-optimal. This may also relate a more pragmatic assumption of Satisficing in the determination of the weights. As per the behavioural economics definition (Dixon (2020)), there might instead be a region wherein the weights for each quality factor are 'satisfactory', if not optimal. This could also account for any determinations that fall slightly away from the optimal weights in practice, but may confound the resulting composite indicator by creating a sort of 'region' of satisfactory CI values, rather than a well-defined exact value for the indicator.

Finally, and in a more general sense of weighting, one argument against BoD is that is ignores, by definition, any expert opinions on the importance of the quality factors to the indicator. For as advantageous as BoD might appear to be, owing to its mathematical approach, there is significant merit in actually incorporating practised opinion into the indicator's design. The mathematics of BoD might find an optimal selection of weights, but by no means does that imply that it finds the correct weights, agreeable with the actual choices of the industry. For future indicator design, expert influence from each company would be equally wise, for the purpose of best iterating upon this thesis' indicator design. As with Zhou et. al. (2010)'s contributions to BoD design, one could even create some sort of 'mix' of the weights:

$$w_{q,i,t}^{mix} = \xi w_{q,i,t}^{BoD} + (1 - \xi) w_{q,i,t}^{Exp}$$

Which could include the aforementioned process of determining relative importance of the weights within certain, perhaps also industry-determined, DMU- and year-specific, upper and lower bounds.

7.3.3 Further Indicator Analysis

Regardless of how the indicator design is iterated upon, there is also scope to further the analysis of the computed results for the purpose of understanding how the indicator behaves, how it reacts to changes and robustness checks, and overall whether the indicator's results stand up to statistical scrutiny. Chapter 5 achieved some of this type of further analysis, by looking at the indicator's reaction to random changes in various stages of the design process through Uncertainty Analysis. The results of that analysis indicated a moderate amount of volatility due to uncertainty in the data selection, normalisation, weighting and aggregation steps of the design process, which adds some credence to the importance of the choices of these parts of the design, as was discussed both earlier in this section and in Chapter 5.

As with most results of an empirical nature, there is a fair amount more that could be looked at from an analytical perspective in theory. As per OECD (2008)'s handbook for constructing composite indicators, one omitted activity in this thesis was Sensitivity Analysis, owing primarily to the common methods of analysis (Saltelli (2002), Saltelli et. al. (2010)) requiring more data than is reasonably available. Alternative methods of such analysis, using distributional characteristics to produce sensitivities, can be found for example in Rahman (2016), although this has a separate difficulty of being complex from a computational perspective.

Another idea considered but not employed with the thesis that focuses more on the 'robustness' of the indicator, is the notion of a Three-Stage CI, analogous to the 3SDEA model used in Chapters 5 and 6. In principle, many of the CPC measures, or other measures of quality that could be used in a composite industry quality indicator, are tied to the social and geographical make-up of the company's catchment area. To that end, as in the 3SDEA model, accounting for these differences in operating environment could yield an indicator that better demonstrates the ‘real’ quality performance. Intuitively, one idea was to adjust the BoD weighting structure in the thesis’ indicator to that of a three-stage procedure, but this particular weighting method does not possess the inputs used in the three-stage adjustment model elsewhere presented. So, a future direction of deriving such an adjustment would need to either determine how the BoD weighting method could adjust for operating conditions, or another, better suited method of weighting will need to be found.

7.4 Conclusion

This chapter sought to be very exploratory in nature, and achieved so by investigating two possible extensions to the application of a composite indicator of quality in the English and Welsh water and sewerage industry, motivated by contemporary issues in the space. This chapter first derived naïve measures for the incidence of droughts and floods, and found limited correlations between those measures of extreme weather and the quality indicator. Then, a large part of this chapter was also dedicated to the discussion of improvements in the make of the composite indicator, and how the indicator could be further analysed or made more robust.

The hope of this chapter, ultimately, is to demonstrate not only the scope of the current indicator derived in Chapter 5, but also to give some idea of the scope of possible future research directions for this kind of indicator, as to better provide measures to improve quality in the industry. One example of this scope in application is found in Appendix C, where the dynamics of the composite indicator are defined and analysed.

Chapter 8

Conclusion

To conclude this thesis, this short chapter will summarise the findings of the three previous chapters which aimed to contribute to the research literature of the industry. These findings will be discussed, measuring how the findings have answered the research questions set at the start of the thesis, and what the implications of the findings are. Then, the chapter will discuss future research directions, drawing partly from some discussion in the contributing chapters, and partly from other ideas. The chapter, and the thesis, will then conclude with some final thoughts.

8.1 Summary and Discussion of Findings

This thesis proposed five research questions:

1. To what extent can a new measurement of Quality be derived, which accounts for the newer, broader definitions of factors of quality, as illustrated by the Common Performance Commitments?

2. Does this new measure of quality, when included as an Output in production, yield significantly different Technical Efficiency Scores for companies, compared to older models?

3. Using Dynamic Models, to what extent have the recent regulatory changes affected measures of Efficiency and Capex Bias over time?

4. To what extent are there Dynamics of quality, and, in dynamic models, to what extent does the novel inclusion of quality affect Capex Bias, by way of affecting Allocative Efficiency?

5. Throughout the previous research questions, how does Welsh Water, the only Non-Profit company in the industry, differ in terms of results from its other industry counterparts?

These research questions were covered by Chapters 5 and 6. With an equivalent list, the answers to each question, according to the findings of this thesis, can be summarised as follows:

1. A Composite Indicator can be constructed from CPCs and other relevant quality factors, and has produced an index 'overall' quality which has a far lower average quality and far more volatility than the predecessor indices - there is, therefore, great scope for quality improvement.

2. Yes: DEA Technical Efficiency Scores are significantly higher on average, according to the Wilcoxon Signed-Rank test, in models with the CI measure as an output, compared to models with the old quality measures ans models with no quality adjustments. These results are robust to differences in environmental heterogeneities, and also show that the models of old quality and no quality are not significantly different.

3. Using dynamic DEA models with quasi-fixed Capital, there is little evidence of changes in dynamic efficiency scores due to moving through regulatory periods - though some movement is observed throughout PR14 on average, the overall change in negligible. Including a CI quality output in the models, there exists some evidence of efficiency improvements throughout PR14; all results are robust to environmental heterogeneities.

4. Interestingly, little evidence of Capex Bias appears to exist via changes in Allocative Efficiency; instead, inefficiencies in the models appear to be driven by Technical Efficiency, suggesting an appropriate treatment and allocation of Capital, but inefficient employment of the input. Results are robust to the inclusion of the quality output and environmental differences.

5. In those models of efficiency which also account for operational heterogeneities between the companies, the industry's Not-for-Profit company - Welsh Water - demonstrates periods of greater technical efficiency compared to the average For-Profit performance. This behaviour is not carried over to Allocative Efficiency, suggesting that the company still over-invests in Capital relative to other companies, though there are improvements in the company's allocative efficiency over time.

So what do these findings mean in sum for the industry? The first, and perhaps the most important point, is that this new measure of quality derived in Chapter 5 is significantly different, significantly more volatile, and significantly worse than the older, stagnant' measures in terms of industry compliance. Though outwardly negative, from the perspective of regulation, this new indicator tells of a great scope for improvement throughout the industry, in various aspects of quality. By extension, this could also provide support for the need of newer regulatory targets, such as the Common Performance Commitments, as they represent areas in the industry that require significant quality improvement still.

The employment of the indicator as an output in DEA models, and the resultant significant differences in technical efficiency scores compared to models with older quality approaches, suggests that assigning a more important role to quality - one where investment into it is considered a trade-off with the production of water and wastewater services - is fruitful for industry benchmarking. The differences in the models' resultant efficiency scores being consistent after accounting for differences in the operating environments of companies lends greater credence still, suggesting that the technical efficiencies caused by this quality output are due to company decision-making, rather than their operating conditions.

The assessment of whether this quality indicator could explain some of the industry's prevalent Capex Bias yielded particularly interesting results in Chapter 6. It was found that, in terms of allocative efficiency, no significant differences exist between the quality output models and older quality models, when quasi-fixed Capital was included to create dynamic DEA specifications. This result, which itself implies that the capex bias is not caused by the investment into quality improvements, is further confounded by the fact that there was significant changes in Technical Efficiency in these dynamic models, when compared to static model equivalents. Overall, these findings suggest that specifying capital-intensive projects as quasi-fixed does not affect capex bias through allocative efficiency, but instead through technical efficiency - companies are allocating their inputs similarly, but are employing them poorly in a technical sense. From a regulatory standpoint, then, it would appear that the reduction of capex bias may not come from better allocation, but from greater cost minimisation using the correctly-allocated inputs. This bears some support to Ofwat's use of Totex and later Botex regulation, where, as Chapter 6 describes, fixed Opex-Capex shares are used, and the total (base) costs are regulated in total, with fixed proportions of resources given to both Labour and Capital.

When further examining the scope of the composite indicator, Chapter 7 finds some promising, albeit limited, results. Using very rough measures to define the incidence of extreme weather, defined as Droughts or Floods, the correlations of these measures with the composite quality indicator and its input factors show little significance overall, though some correlations directionally show that companies may, on average, have some resilience to the weathers when it comes to their effects on overall quality. Intuitively, that there is some evidence in any direction is positive, and future research should, as the chapter discusses, better incorporate extreme weather variables.

Looking at the dynamics of the composite indicator gives an interesting, if slightly dismal, addition to the results of Chapter 5. The indicator's changes over time are, ultimately, minimal, showing even a slight decrease in the indicator's value on average. This supports the notion that the indicator gives great scope for companies to improve quality throughout their operations, but also suggests that the attention to quality is, if only slightly, in decline. From a policy perspective, quality needs to be increased throughout the industry, and given the results of the composite indicator over time, this work gives a new metric on which Ofwat can base the importance of future quality-related regulatory targets, be it the increased need for common commitments to quality such as the CPCs and ODIs by extension, or through other targeted benchmarks.

Lastly, the differences in results between Welsh Water and the other WASCs has some interesting behavioural implications for the industry. Chapter 5 finds that, once environmental differences in company operations are accounted for, non-profit behaviour appears to yield more technical efficiency over time than the average for-profit levels. On the other hand, Chapter 6 finds that this behaviour does not hold for allocative efficiency, with the for-profit average allocative efficiency remaining consistently above that of non-profit efficiency scores.

Specific to non-profit companies, these findings suggest that, in terms of underlying performance, the issues of company performance arise when allocating resources correctly, and in context reducing capex expenditures to remove capex bias. On the other hand, from a wider industry perspective, it is interesting that non-profit behaviours seem to yield better technical efficiency in terms of underlying performance. Though the issue is no doubt more complicated given that environmental differences must always be a factor in practice, the evidence that there is some advantage to non-profit behaviours could suggest, from a behavioural economics point-of-view, that there is scope to see at least technical efficiency improvements by changing or incentivising for-profit behaviours to be closer to non-profit behaviour. There could be interesting future research to be carried out in this area, and similarly there is scope to look at, based on Chapter 6's results, how non-profit behaviour could be adjusted to better allocative efficiency.

Much of the conclusions of each contributing chapter discuss not just the results of the chapter, but the flaws and next steps in the development of the ideas presented, and this conclusion has acted much the same. To end the thesis as a whole, some final future research directions will now be briefly discussed, as to demonstrate where this thesis could be used as a stepping stone to better the understanding of quality, its measurement, and its impacts in the industry.

8.2 Future Research Directions

This section aims to look more towards the future research directions that could be meritorious in the field of improving water and wastewater industry quality. There are three topics that will be postulated on here, as to close the thesis with the hope of such topics being undertaken in the near future.

8.2.1 Cost of Quality Models

The use of quality as a production output in standard CCR DEA models presents the economic idea that firms must choose between producing more water or wastewater outputs, or choose to invest in the improvement of their overall quality, as was represented by the composite indicator. These models yielded technical efficiency as the primary means of benchmarking companies against each other, and in Chapter 6 the dynamic DEA models looked at Overall Efficiency, composed of technical and allocative efficiency.

These results are all well and good, but for the sake of a practical understanding of how quality affects the industry's current regulation, one idea that could be considered is that of the 'Cost of Quality': how, in a monetary sense, can quality and investments into its betterment be accounted for in, say, Totex or Botex cost functions, which are then regulated by Ofwat? Suppose that the thesis' composite was the 'Quantity' of quality; then, how might a Price be derived?

The problem of finding a cost to quality extends past just the need to determine its monetary value. If quality is considered an investment into all aspects of industry quality, as it is in this thesis, then how might be incorporated into a cost function? That quality improvements would then affect water and wastewater services could well result in complications when it comes to the estimation of the whole function. An interesting future direction, regardless of how quality is defined, would be to consider how its measurement, when treated as an important factor output, could be accounted for in total costs, and by extension this could provide greater elucidation on how quality might impact Capex Bias via the regulatory cost models.

8.2.2 Mindful Behaviour of Firms

This next section borders well into a value judgment, but nonetheless has a basis in recent investigations into how companies of any kind 'ought' to behave to best benefit society. This topic will also tie into the previous discussion on the implications of non-profit versus for-profit behaviours within the industry, tying the more behavioural economic ideas of Altruism from a non-profit perspective into the idea of Mindful business as a future direction for the industry.

The idea of a Mindful, or Purposeful business, as is presented here, comes from Mayer (2018)'s book, *Prosperity: Better Business Makes the Greater Good.* Therein, the idea of a business' purpose, and how it should be achieved is questioned, with the principle philosophy being that companies, having been created to fulfill a specific purpose, should strive only and completely to achieve that purpose, such that it is also best for its staff, customers, the environment, society, and so on. After reaching such a state of company performance, then and only then is supernormal profit permissible - any profit before that goal should be used for bettering the company's delivery of their purpose, and should account for costs not normally considered in total costs.

As it pertains to this industry, this idea of purposeful business appears most aligned with non-profit behaviours. Welsh Water quote that "*[They are] owned by Glas Cymru, a single purpose company with no shareholders and is run solely for the benefit of customers.*" Assuming that this holds true, and that their continued existence is a positive sign of that principle working effectively, a future research question ought to be the extent to which the industry can achieve mindful business practice throughout, on account of water and sewerage companies, providing necessary goods such as utilities, ought to best provide for their customers before thinking about profits. To that end, behavioural economics ideas, such as Altruism, could be useful here, to model non-profit behaviours and, if not regulate them upon for-profit companies, then incentivise such modeled decisions as to have for-profit companies behave sufficiently like an altruistic, mindful business.

8.2.3 Efficiency Shock Recovery

This last future direction strays largely away from the work covered in the thesis, and focuses more on the more general ways by which efficiency is measured in the first place[1].

Buncic et. al. (2023) look at the idea of macroeconomic 'Stars' - key variables that are recovered from a macroeconomic model. The paper notes that, in a lot of contemporary models that address finding these stars, there is actually a significant issue in their recovery, on account of the models being short - there are more shocks in the model than there are variables to solve the model with. Much of the discussion ends similarly, in that the amount of shocks, in prac-

[1]Special thanks go to Prof. Adrian Pagan for his visit to Cardiff University in October 2023, and the subsequent communications with me on this topic.

tice, actually make the star variables unrecoverable, or at least not accurately recoverable.

Why does this matter in topics concerning microeconomic efficiency? Consider the standard SFA model, which decomposes the error term into two distinct parts, a random white noise shock, and a non-negative inefficiency shock. If this is held true, and only one dependent variable is used in the model, then the same issue arises - there are too many shocks to accurately recover desirable results. In the case of efficiency, if this problem exists, then the efficiency scores themselves may not be reliably estimated at all. As it pertains to this thesis, this could mean that the 3SDEA adjustments, which are estimated with an SFA model in the second stage, might not accurately reflect operational differences between companies. More generally, as using SFA is a common model for cost functions in the literature, perhaps they too are inaccurate. A useful future direction would be to explore this possibility, and if there is a problem with efficiency score recovery, to also determine how such an issue could be overcome.

In conclusion, this thesis has found a novel way to measure quality in the English and Welsh Water and Sewerage industry, hopefully addressing the reported issue of stagnation in older quality indices. As the latter chapters conclude, there is a great deal of iteration and improvement that can be done with composite indicators, to best create tools for policy, regulation, and as the thesis also demonstrated, measuring efficiency in the industry. On the premise that the thesis began with, that water is the '*sine qua non*' of the city, one can only hope that this new principle of measuring quality and incorporating it more prominently into empirical models, allows for future directions, and future industry decisions, to be made with a better understanding of quality.

References

Afriat S. N. (1967). The construction of a utility function from expenditure data. *International Economic Review, Vol. 8*, pp. 67-77.

Afriat S. N. (1972). Efficiency estimation of production functions. *International Economic Review, Vol. 13*, pp. 568-598.

Aigner D., Lovell C. A. K. & Schmidt P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics, Vol. 6*, pp. 21-37.

Arocena P., Saal D. & Coelli T. (2009). Measuring Economies of Horizontal and Vertical Integration in the US Electric Power Industry: How Costly is Unbundling? *Aston Business School, Aston University.*

Ashton J. K. (2000a). Cost Efficiency in the UK Water and Sewerage Industry. *Applied Economics Letters, Vol. 7(7)*, pp. 455-458.

Ashton J. K. (2003). Capital Utilisation and Scale in the English and Welsh Water Industry. *The Service Industries Journal, Vol. 23(5)*, pp. 137-149.

Asmild M., Kronborg D. & Matthews K. (2016). Introducing and Modelling Inefficiency Contributions. *European Journal of Operational Research, Vol. 248*, pp. 725-730.

Atkins & Peirce (2024). Protecting the natural capital and biodiversity in agricultural supply chains: introduction. In Atkins (Ed.), *Protecting natural capital and biodiversity in the agri-food sector.* Cambridge: Burleigh-Dodds Science Publishing.

Atkinson A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory, Vol. 2*, pp. 244-263.

Averch H. & Johnson L. L. (1962). Behaviour of the Firm under Regulatory Constraint. *American Economic Analysis, Vol. 52(5)*, pp. 1052-1069.

Bandura R. (2008). A Survey of Composite Indices Measuring Country Performance: 2008 Update. *United Nations Development Programme, Working Paper.*

Berkelaar M. & others (2023). _lpSolve: Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs_. R package version 5.6.18, https://CRAN.R-project.org/package=lpSolve.

Bernini C., Guizzardi A. & Angelini G. (2013). DEA-Like Model and Common Weights Approach for the Construction of a Subjective Community Well-Being Indicator. *Social Indicators Research, Vol. 114*, pp. 405-424.

Blank J. L. T. & Valdmanis V. (2005). A Modified Three-Stage Data Envelopment Analysis. *European Journal of Health Economics, Vol. 50*, pp. 65-72.

Bogetoft P. & Otto L. (2022), Benchmarking with DEA and SFA, R package version 0.31.

Bottasso A. & Conti M. (2003). Cost Inefficiency in the English and Welsh Water Industry: A Heteroskedastic Stochastic Cost Frontier Approach. *University of Essex.* Retrieved from: http://repository.essex.ac.uk/8872/1/dp573.pdf

Bottasso A. & Conti M. (2009a). Scale Economies, Technology and Technical Change in the Water Industry: Evidence from the English Water Only Sector. *Regional Science and Urban Economics, Vol. 39(2)*, pp. 138-147.

Bottasso A. & Conti M. (2009b). Price cap regulation and the ratchet effect: a generalised index approach. *Journal of Productivity Analysis, Vol. 32(3)*, pp. 191-201.

Bottasso A., Conti M., Piacenz M. & Vannoni D. (2011). The appropriateness of the poolability assumption for multiproduct technologies: Evidence from the English water and sewerage utilities. *International Journal of Production Economics, Vol. 130*, pp. 112-117.

Brea-Solis H., Perelman S. & Saal D. S. (2017). Regulatory incentives to water losses reduction: the case of England and Wales. *Journal of Productivity Analysis, Vol. 47*, pp. 259-276.

Brunekreeft G. & Rammerstorfer M. (2020). OPEX-risk as a source of CAPEX-bias in monopoly regulation. *Jacobs University Bremen, Bremen Energy Working Papers.*

Buncic D., Pagan A. & Robinson T. (2023). Recovering stars in macroeconomics. *Melbourne Institute Applied Economic and Social Research, Working Paper No.12/23.*

Chambers R. G., Chung Y. & Färe R. (1996). Benefit and Distance Functions. *Journal of Economic Theory, Vol. 70*, pp. 407-419.

Chambers R. G., Chung Y. & Färe R. (1998). Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimisation Theory and Applications, Vol. 98(2)*, pp. 351-364.

Champely S (2018). _PairedData: Paired Data Analysis_. R package version 1.1.1, https://CRAN.R-project.org/package=PairedData.

Charles V., Aparicio J. & Zhu J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research, Vol. 279*, pp. 929-940.

Charnes A., Cooper W. W. & Rhodes E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research, Vol.*

2, pp. 429-444.

Cherchye L., Lovell C. A. K., Moesen W. & van Puyenbroeck T., (2007a). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review, Vol. 51*, pp. 749-779.

Cherchye L., Moesen W., Rogge N. & van Puyenbroeck T. (2007b). An Introduction to 'Benefit of the Doubt' Composite Indicators. *Social Indicators Research, Vol. 82*, pp. 111-145.

Cooper W. W., Li S., Seiford L. M., Tone K., Thrall R. M. & Zhu J. (2001). Sensitivity and Stability Analysis in DEA: Some Recent Developments. *Journal of Productivity Analysis, Vol. 15*, pp. 217-246.

Cordero-Ferrera J. M., Pedraja-Chaparro F. & Santín-González D. (2010). Enhancing the inclusion of non-discretionary inputs in DEA. *Journal of the Operational Research Society, Vol. 61(4)*, pp. 574-584.

Cubbin J. & Tzanidakis G. (1998). Regression versus Data Envelopment Analysis for Efficiency Measurement: An Application to the England and Wales Regulated Water Industry. *Utilities Policy, Vol. 7*, pp. 75-85.

Debreu G. (1951). The Coefficient of Resource Utilisation. *Econometrica, Vol. 19(3)*, pp. 273-292.

Diewert W. E. & Fox K. J. (2017). Decomposing Productivity Indexes into Explanatory Factors. *European Journal of Operational Research, Vol. 256*, pp. 275-291.

D'Inverno G., Carosi L. & Romano G. (2021). Environmental sustainability and service quality beyond economics and financial indicators: A performance evaluation of Italian water utilities. *Socio-Economic Planning Sciences, Vol. 75*.

Dixon H. (2020). Almost-Maximisation as a Behavioural Theory of the Firm:

Static, Dynamic and Evolutionary Perspectives. *Review of Industrial Organisation, Vol. 56*, pp. 237-258.

El Gibari S., Gómez T. & Ruiz F. (2022). Combining reference point based composite indicators with data envelopment analysis: application to the assessment of universities. *Scientometrics, Vol. 127*, pp. 4363-4395.

Erbetta F. & Cave M. (2007). Regulation and Efficiency Incentives: Evidence from the England and Wales Water and Sewerage Industry. *Review of Network Economics, Vol. 6(4)*, pp. 425 - 452.

Espiner T. (2024, July 11). Row over rise in water bills as firms say it's not enough. *BBC News.*

Farrell M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Vol. 120(3)*, pp. 253-290.

Fried H. O., Lovell C. A. K., Schmidt S. S. & Yaisawarng S. (2002). Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. *Journal of Productivity Analysis, Vol. 17*, pp. 157-174.

Fried H. O., Schmidt S. S. & Yaisawarng S. (1999). Incorporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency. *Journal of Productivity Analysis, Vol. 12*, pp. 249-267.

Garcia S. & Thomas A. (2001). The Structure of Municipal Water Supply Costs: Application to a Panel of French Local Communities. *Journal of Productivity Analysis, Vol. 16*, pp. 5 - 29.

Hassan J. (1988). *A History of Water in modern England and Wales.* Manchester University Press.

Henriques A. A., Camanho A. S., Amorim P. & Silva J. G. (2020). Performance benchmarking using composite indicators to support regulation of the Portuguese wastewater sector. *Utilities Policy, Vol. 66.*

House of Commons. (2022). *Water Quality in Rivers* (4th Report of Session 2021-22). House of Commons.

Howes R., Skea J. & Whelan B. (1997). *Clean and Competitive: Motivating Environmental Performance in Industry.* Earthscan.

Hunt L. C. & Lynk E. L. (1995). Privatisation and Efficiency in the UK Water Industry: An Empirical Analysis. *Oxford Bulletin of Economics and Statistics, Vol. 57(3)*, pp. 371-388.

Jenkins S. (2022, August 22). England's water industry now represents the unacceptable face of capitalism. *The Guardian.*

Kao C. & Hung H-T. (2005). Data envelopment analysis with common weights: the compromise solution approach. *Journal of the Operational Research Society, Vol. 56(10)*, 1196-1203.

Kneip A., Simar L. & Wilson P. W. (2016). Testing Hypotheses in Nonparametric Models of Production. *Journal of Business & Economic Statistics, Vol. 34(3)*, pp. 435-456. Kohl S. & Brunner J. O. (2020). Benchmarking the benchmarks - Comparing the accuracy of Data Envelopment Analysis models in constant returns to scale settings. *European Journal of Operational Research, Vol. 285(3)*, pp. 1042-1057.

Koopmans T. C. (1951). Analysis of Production as an Efficient Combination of Activities. In T. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation: Proceedings of a Conference* (pp. 33-97). New York: John Wiley & Sons.

Kuosmanen T. (2008). Representation theorem for convex nonparametric least squares. *The Econometrics Journal, Vol. 11(2)*, pp. 308-325.

Kuosmanen T. & Kortelainen (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis, Vol. 38*, pp. 11-28.

Lai H. & Kumbhakar S. C. (2018). Endogeneity in Panel Data Stochastic Frontier Model with Determinants of Persistent and Transient Inefficiency. *Economics Letters, Vol. 162*, pp. 5-9.

Lai H. & Kumbhakar S. C. (2019). Technical and allocative efficiency in a panel stochastic frontier system model. *European Journal of Operational Research, Vol. 278*, pp. 255-265.

Lambsdorff J. G. (2005). The Methodology of the 2005 Corruption Perceptions Index. *Transparency International (2005).* Retrieved from: https://www.transparency.org/files/content/tool/2005_CPI_LongMethodology_EN.pdf

Lampard, E. E. (1973). The Urbanising World. In Dyos H. J. & Wolff M. (eds.), *The Victorian City, 1, Images and Reality*, (pp. 21). Routledge.

Littlechild S. (1988). Economic Regulation of Privatised Water Authorities and some Further Reflections. *Oxford Review of Economic Policy, Vol. 4(2)*, pp. 40-68.

Lovell C. A. K. (2003). The Decomposition of Malmquist Productivity Indexes. *Journal of Productivity Analysis, Vol. 20*, pp. 437-458.

Lynk E. L. (1993). Privatisation, Joint Production and the Comparative Efficiencies of Private and Public Ownership: The UK Water Industry Case. *Fiscal Studies, Vol. 14(2)*, pp. 98-116.

Mas-Colell A., Whinston M. D. & Green J. R. (1995). *Microeconomic Theory.* Oxford University Press.

Mayer, C. (2018). *Prosperity: Better Business Makes the Greater Good.* OUP Oxford.

Maziotis A., Saal D. S., Thannasoulis E. & Molinos-Senante M. (2014). Profit change and its drivers in the English and Welsh water industry: is output quality important? *Water Policy.*

Maziotis A., Saal D. S., Thanassoulis E. & Molinos-Senante M. (2015). Profit, productivity and price performance changes in the water and sewerage industry: an empirical application for England and Wales. *Clean Technology and Environmental Policy, Vol. 17*, pp. 1005-1018.

Meeusen W. & Van den Broeck J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review, Vol. 18(2)*, pp. 435-444.

Milfeldt M (2016). _epandist: Statistical Functions for the Censored and Uncensored Epanechnikov Distribution_. R package version 1.1.1, https://CRAN.R-project.org/package=epandist.

Mocholi-Arce M., Sala-Garrido R. & Molinos-Senante M. (2020). Performance assessment of water companies: A metafrontier approach accounting for quality of service and group heterogeneities. *Socio-Economics Planning Sciences.*

Mocholi-Arce M., Sala-Garrido R., Molinos-Senante M. & Maziotis A. (2021). Water company productivity change: A disaggregated approach accounting for changes in inputs and outputs. *Utilities Policy, Vol. 70.*

Molinos-Senante M. & Maziotis A. (2018). Flexible versus common technology to estimate economies of scale and scope in the water and sewerage industry: an application to England and Wales. *Environmental Science and Pollution Research, Vol. 25*, pp. 14158-14170.

Molinos-Senante M. & Maziotis A. (2019). Cost Efficiency of English and Welsh Water Companies: a Meta-Stochastic Frontier Analysis. *Water Resources Management, Vol. 33*, pp. 3041-3055.

Molinos-Senante M. & Maziotis A. (2020). Drivers of productivity change in water companies: an empirical approach for England and Wales. *International Journal of Water Resources Development, Vol. 36(6)*, pp. 972-991.

Molinos-Senante M., Maziotis A., Mocholí-Arce M. & Sala-Garrido R. (2015a). Assessing the relative efficiency of water companies in the English and Welsh water industry: a metafrontier approach. *Environmental Science and Pollution Research, Vol. 22*, pp. 16987-16996.

Molinos-Senante M., Maziotis A., Mocholí-Arce M. & Sala-Garrido R. (2015b). Accounting for service quality to customers in the efficiency of water companies: evidence from England and Wales. *Water Policy, Vol. 18*, pp. 513-532.

Molinos-Senante M., Maziotis A. & Sala-Garrido R. (2016). Estimating the cost of improving service quality in water supply: A shadow price approach for England and Wales. *Science of the Total Environment, Vol. 539*, pp. 470-477.

Molinos-Senante M., Maziotis A. & Sala-Garrido R. (2017b). Assessing the Impact of Quality of Service on the Productivity of Water Industry: a Malmquist-Luenberger Approach for England and Wales. *Water Resources Management, Vol. 31(2)*, pp. 2407-2427.

National Academies of Sciences, Engineering, and Medicine (2016). *Attribution of Extreme Weather Events in the Context of Climate Change*. Washington DC: The National Academies Press. Retrieved from: https://doi.org/10.17226/21852.

Nemoto J. & Goto M. (2003). Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities. *Journal of Productivity Analysis, Vol. 19*, pp. 191 - 210.

Nithammer C. M., Mahabir J. & Dikgang J. (2022). Efficiency of South African water utilities: a double bootstrap DEA analysis. *Applied Economics, Vol. 54(26)*, pp. 3055-3073.

OECD (2008). *Handbook on Constructing Composite Indicators.* ISBN 978-92-64-04345-9.

Ofwat (1994). *Future Charges for Water and Sewerage Services.* Retrieved from: https://webarchive.nationalarchives.gov.uk/20150603192123/ http://www.ofwat.gov.uk/pricereview/det_pr_fd94.pdf

Ofwat (1999). *Future Water and Sewerage Charges 2000 - 05.* Retrieved from: https://webarchive.nationalarchives.gov.uk/20150603222823/ http://www.ofwat.gov.uk/pricereview/pr99/det_pr_fd99.pdf

Ofwat (2004). *Future Water and Sewerage Charges 2005 - 10.* Retrieved from: https://webarchive.nationalarchives.gov.uk/20150603192547/ http://www.ofwat.gov.uk/pricereview/pr04/det_pr_fd04.pdf

Ofwat (2008) *The Development of the Water Industry in England and Wales.* Retrieved from: https://www.ofwat.gov.uk/publication/the-development-of-the-water-industry-in-england-and-wales/.

Ofwat (2009). *Future Water and Sewerage Charges 2010 - 15: Final Determinations.* Retrieved from: https://webarchive.nationalarchives.gov.uk/ 20150603201359/https://www.ofwat.gov.uk/pricereview/pr09phase3/ det_pr09_finalfull.pdf

Pointon C. & Matthews K. (2016). Dynamic Efficiency in the English and Welsh Water and Sewerage Industry. *Omega, Vol. 58*, pp. 86-96.

Portela M. C. A. S., Thanassoulis E., Horncastle A. & Maugg T. (2011). Productivity change in the water industry in England and Wales: application of the meta-Malmquist index. *Journal of the Operational Research Society, Vol. 62(12)*, pp. 2173-2188.

Portela M. C. A. S., Thanassoulis E. & Simpson G. (2004). Negative Data in DEA: A Directional Distance Approach to Bank Branches. *Journal of the Operational Research Society, Vol. 55(10)*, pp. 1111-1121.

Porter M. E. & Stern S. (2001). National Innovative Capacity. *Harvard University.*

Rahman S. (2016). The f-Sensitivity Index. *Society for Industrial and Applied Mathematics, Uncertainty Quantification.*

Renzetti S. & Dupont D. P. (2003). Ownership and Performance of Water Utilities. *Greener Management International.*

Rogge N. (2018). Composite indicators as generalised benefit-of-the-doubt weighted averages. *European Journal of Operational Research, Vol. 267*, pp. 381-392.

Saal D. S. & Parker D. (2000). The Impact of Privatisation and Regulation on the Water and Sewerage Industry in England and Wales: A Translog Cost Function Model. *Managerial and Decision Economics, Vol. 21*, pp. 253-268.

Saal D. S. & Parker D. (2001). Productivity and Price Performance in the Privatised Water and Sewerage Companies of England and Wales. *Journal of Regulatory Economics, Vol. 20(1)*, pp. 61-90.

Saal D. S., Parker D. & Weyman-Jones T. (2007). Determining the Contribution of Technical Change, Efficiency Change and Scale Change to Productivity Growth in the Privatised English and Welsh Water and Sewerage Industry: 1985-2000. *Journal of Productivity Analysis, Vol. 28*, pp. 127-139.

Saal D. S. & Reid S. (2005). Estimating Opex Productivity Growth in English and Welsh Water and Sewerage Companies: 1993 - 2003. *Faculdade de Economia da Universidade do Porto.* Retrieved from: https://www.fep.up.pt/conferences/earie2005/cd_rom/Session%20VI/VI.N/Saal_Reid.pdf

Saal, D., Wright J. & Huggins, M. (2017). *Productivity Improvement in the Water and Sewerage Industry in England Since Privatisation: Final Report for Water UK.* Loughborough University.

Sala-Garrido R., Mocholi-Arce M., Molinos-Senante M. & Maziotis A. (2021). Assessing the marginal cost of reducing greenhouse gas emissions in the English and Welsh water and sewerage industry: A parametric approach. *Utilities Policy, Vol. 70.*

Saltelli A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications, Vol. 145*, pp. 180-297.

Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M. & Tarantola S. (2010). Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index. *Computer Physics Communications, Vol. 181.* pp. 259-270.

Sarkis J. (2010). Preparing Your Data for DEA. In Zhu J. & Cook W. D. (Eds.). *Modelling Data Irregularities and Structural Complexities in Data Envelopment Analysis.* Springer.

Simar L. & Wilson P. W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science, Vol. 44(1)*, pp. 49-61.

Simar L. & Wilson P. W. (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics, Vol. 136*, pp. 31-64.

Simar L. & Wilson P. W. (2011). Performance of the Bootstrap for DEA Estimators and Iterating the Principle. In Cooper W., Seiford L. M. & Zhu J. (Eds.). *Handbook on Data Envelopment Analysis*, pp. 241-271. Springer US, New York.

Smith A., Wheat P., Thiebaud T-C. & Stead A. (2019). CAPEX Bias and Adverse Incentives in Incentive Regulation: Issues and Solutions. *International Transport Forum.*

Stone & Webster Consultants. (2004). *Investigation into Evidence for Economics of Scale in the Water and Sewerage Industry in England and Wales.* Report for Ofwat.

Thanassoulis E. (2000). The use of data envelopment analysis in the regulation of UK water utilities: Water distribution. *European Journal of Operational Research, Vol. 126*, pp. 436-453.

The Economist. (2021). Our Big Mac index shows how burger prices are changing. *The Economist.* Retrieved from: https://www.economist.com/big-mac-index

UK Parliament. (2022). *Water quality in rivers* (Fourth Report of Session 2021-22). UK Government Publishing Office.

Van Puyenbroeck T. & Rogge N. (2017). Geometric mean quantity index numbers with Benefit-of-the-Doubt weights. *European Journal of Operational Research, Vol. 256*, pp. 1004-1014.

Varian, H. R. (2009). *Microeconomic Analysis.* W. W. Norton & Company.

Wang H. (2015). A generalised MCDA-DEA (multi-criterion decision analysis-data envelopment analysis) approach to construct slacks-based composite indicator. *Energy, Vol. 80*, pp. 114-122.

Wickham H, Bryan J (2023). _readxl: Read Excel Files_. R package version 1.4.2, https://CRAN.R-project.org/package=readxl.

Wong W. P. (2021). A Global Search Method for Inputs and Outputs in Data Envelopment Analysis: Procedures and Managerial Perspectives. *Symmetry, Vol. 13*, pp. 1155.

Wye Salmon Association. (2019). *River Wye Salmon Action Plan, Bold and Urgent Action: A way forward.* Wye Salmon Conservation Society.

Yakymova L., Novotná A., Kuz V. & Tamándl L. (2022). Measuring industry digital transformation with a composite indicator: A case study of the utility industry. *Journal of International Studies, Vol. 15(1)*, pp. 168-180.

Zhou P., Ang B. W. & Zhou D. Q. (2010). Weighting and Aggregation in Composite Indicator Construction: a Multiplicative Optimisation Approach. *Social Indicators Research, Vol. 96*, pp. 169-181.

Zohrehbandian M., Makui A. & Alinezhad A. (2010). A compromise solution approach for finding common weights in DEA: an improvement to Kao and Hung's approach. *Jounral of the Operational Research Society, Vol. 61(4)*, pp. 604-610.

Appendices

A: Bootstrapping Methods

The following appendix covers the two different bootstrapping methods discussed in this thesis, for one-stage bias reduction and three-stage adjustments respectively.

Simar & Wilson (1998)

Let $\hat{\theta}_1, \ldots, \hat{\theta}_N$ be the N Technical Efficiency Scores from the original, biased one-stage DEA model. To generate bootstrapped scores, a simple bootstrap sample of these scores is taken with drawing samples with replacement, giving a vector of sampled scores β^*. An intermediate set of scores, $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$, are then generated via the following random generator:

$$\tilde{\theta}_i = \begin{cases} \beta_i^* + h\varepsilon_i^*, \text{ if } \beta_i^* + h\varepsilon_i^* \leqslant 1, \\ 2 - \beta_i^* - h\varepsilon_i^*, \text{ otherwise} \end{cases} \tag{.1}$$

For some bandwidth $h = 0.014$[2], and some deviation $\varepsilon_i^* \sim N(0,1)$. The bias-corrected efficiency scores are then calculated as:

$$\theta_{i,b}^* = \bar{\beta^*} + \left(1 + \frac{h^2}{\hat{\sigma}_\theta^2}\right)^{-\frac{1}{2}} (\tilde{\theta}_i - \bar{\beta^*}) \tag{.2}$$

Where $\bar{\beta^*} = \frac{1}{N}\sum_i \beta_i^*$ and $\hat{\sigma}_\theta^2 = \frac{1}{N}\sum_i \left(\hat{\theta}_i^2 - \bar{\hat{\theta}}\right)^2$, $\bar{\hat{\theta}} = \frac{1}{N}\sum_i \hat{\theta}_i$. For each bootstrap $b = 1, \ldots, B = 2000$, a new set of scores are collected.

From these scores, and for each bootstrap b, the inputs of the initial DEA models can be adjusted by the relative efficiency between the original technical efficiency and bootstrapped scores:

$$x_{i,b}^* = \frac{\hat{\theta}_i}{\theta_{i,b}^*} x_i \tag{.3}$$

[2]Which is, as per Simar & Wilson (1998), an optimal bandwidth for this bootstrapping process.

Using these adjusted inputs, the one-stage model is then re-run with updated inputs, giving bootstrapped technical efficiency scores $\hat{\theta}^*_{i,b}$. Finally, after all bootstraps are completed, the scores are averaged over the bootstraps to give an average, bias-adjusted measure of technical efficiency for each company:

$$\bar{\hat{\theta}}^*_i = \frac{1}{B}\sum_b \hat{\theta}^*_i \tag{.4}$$

Simar & Wilson (2007)

The three-stage procedure takes the idea of re-estimating the input slacks of the DEA model in Fried et. al. (2002), and bootstraps it via Algorithm 1 of Simar & Wilson (2007). The overall methods, as it is used in this paper, similarly begins with the initial one-stage model of technical efficiency, but finds the Input Slacks of the model via:

$$S_{m,i} = x_{m,i} - x^{bench}_{m,i}, = x_{m,i} - X'_i\lambda_i \tag{.5}$$

That is, the slacks for each of m inputs are found as the difference between the actual data and the 'benchmark' inputs, which are the inputs scaled by the optimal constraints λ also taken from the initial DEA model. The slacks are minimally equal to zero, if the actual inputs are optimal, and positive if there is some scope for further input minimisation. Next, the slacks are regressed on the Environmental Variables, Z, in a Stochastic Frontier Analysis (SFA) Model (Aigner et. al. (1977), Meeusen & van der Broeck (1977)):

$$S_{m,i} = Z'\alpha + v_{m,i} - u_{m,i} \tag{.6}$$

Where $v \sim N(0, \sigma^2_v)$ and $u \sim N^+(0, \sigma^2_u)$ are the White Noise Errors and Inefficiency Errors respectively. From the m SFA regressions, the parameter and residual estimates, $\hat{\alpha}$ and $\hat{\varepsilon} = \hat{v} - \hat{u}$ are collected. From the parameter estimates, slack estimates can be calculated, which are then used to create the

Bootstrapped Slacks, S^b, by randomly drawing from the collected residuals:

$$\begin{aligned} \hat{S}_{m,i} &= Z'\hat{\alpha}, \\ S^b_{m,i} &= \hat{S}_{m,i} + \hat{\varepsilon}^b_{m,i} \end{aligned} \tag{.7}$$

Using these new bootstrapped slacks, the SFA regressions are re-ran, and the final slack estimates of the bootstrap iteration are calculated with the new SFA parameters:

$$\begin{aligned} S^b_{m,i} &= Z'\alpha^b + v^b_{m,i} - u^b_{m,i}, \\ \hat{S}^b_{m,i} &= Z'\hat{\alpha}^b \end{aligned} \tag{.8}$$

As with the previous method for bootstrapping, the final step of the procedure is to average the bootstrapped slacks over the iterations b for each input and each DMU:

$$\bar{\hat{S}}_{m,i} = \frac{1}{N}\sum_b \hat{S}^b_{m,i} \tag{.9}$$

This mean-bootstrapped slack is then the estimate used to adjust the model's inputs in the third stage of the 3SDEA model.

Some notes on Bootstrapping Bias

Particularly to Simar & Wilson (1998) in this thesis, one important thing to note about the use of bootstrapping is potential biases from the methods. Despite the objective of bootstrapping to be the removal of small-sample biases, the actual method may still apply a sort of 'bootstrapping bias'.

In Simar & Wilson (1998)'s case, the method employs re-sampling in each bootstrap which then re-generates efficiency scores further in the process. However, suppose the small sample bias issues take shape in the form of many fully efficient scores. Then, though the bootstrap procedure does remove some of this upward bias from small samples, there inherently remains an amount of that bias precisely because the method of bootstrapping requires the use of the already-biased efficiency score sample. The same issue could arise when there are dimensionality issues, which the method does not resolve.

Other research around bootstrapping is not used in this thesis, but is worth noting nonetheless. For example, Simar & Wilson (2011) further develop their bootstrapping processes by incorporating bias correction and the creation of confidence intervals. Nithammer et. al. (2022) provide a different approach to improving bootstrapping accuracy, by using a 'double bootstrap analysis' in their DEA modelling approach. Lastly, in reference more to the dimensionality problem discussed in the thesis, Wong (2021) propose a global search method to find the best inputs and outputs for a DEA model which, in conjunction with the aforementioned bootstrapping procedures, could provide accurate efficiency estimates that are unbiased, and use the 'optimal' inputs and outputs.

B: RStudio Packages

This appendix lists the packages used throughout the thesis to produce the results therein. All packages are also listed in the References chapter. Packages in this appendix will be accompanied by a brief explanation of their purpose, and where they were utilised in the thesis.

The following packages in RStudio are used throughout the thesis:

Benchmarking: Bogetoft & Otto (2022). Provides tools for DEA and SFA modelling. Used for those models in the thesis, particularly for the SFA process in 3SDEA models.

epandist: Milfeldt (2016). Allows the use of the Epanechnikov distribution in code. Used to provide an appropriate distribution for kernel density estimation of efficiency scores.

lpSolve: Berkelaar et. al. (2023). Provides code for solving linear programs. Used to solve the DEA models employed throughout the thesis.

PairedData: Champely (2018). Used to analyse paired data. Used in thesis to analyse results via Wilcoxon signed-rank tests.

readxl: Wickham & Bryan (2023). Allows for the reading of Excel files by RStudio. Used in the thesis to import the datasets into the coding software.

C: Dynamic Composite Indicators

This appendix explores the notion of dynamic composite indicators, including the development and analysis of a dynamic indicator based on the work in Chapter 5. References to further developments of the indicator are found in Chapter 7's discussion of further composite indicator development, but discussion of future directions for dynamic indicators specifically are found in this appendix.

Chapter 6 concludes with some mention of the potential of a dynamic composite indicator, for the purpose of better modelling the dynamic properties of quality and investment into it, highlighting a potential future interest in treating quality as quasi-fixed, as Capital was in the chapter.

This section addresses the existing literature around the dynamics of composite indicators, using the notion of Performance Change between two time periods. This project's indicator is then developed into various extensions of performance change, and is then examined such that some idea of the dynamics of industry quality can be observed.

Defining a Dynamic Measure

First, it is worth quickly introducing the baseline specification of the indicator for this section:

Normalisation: $I_{q,i,t} = \frac{X_{q,i,t} - \min(X_{q,i,t})}{\max(X_{q,i,t}) - \min(X_{q,i,t})} + \varepsilon,\ \varepsilon > 0$

Weighting: $\max\limits_{w} \left(\sum_q w_{q,i,t} I_{q,i,t} \right),\ s.t.$

$$\sum_q w_{q,i,t} I_{q,i,t} \leqslant 1,$$

$$w_{q,i,t} \geqslant w_l > 0$$

Aggregation: $CI_{i,t} = \sum_{q=1}^{Q} w^*_{q,i,t} (I_{q,i,t} - I^b_{q,t})$

Where the specification of the indicator is near-identical to equations (5.5), (5.8) and (5.13), with the principal differences being a time subscript t, and a

Base Measure that scales the final indicator's factors, $I^b_{q,t}$. This base adjustment is introduced to reflect a general industry level of the input factors, but in practice can represent any other reference points from which company-specific quality factors are compared to - in this case, as with the Linear Threshold Aggregation (5.15), this base is the Industry Mean of factor q at time t.

So, how can this indicator's dynamics be addressed? By design, there are no inter-temporal factors built into the indicator, meaning that, though the indicator does exist for every year of the sample, each year's indicators are functionally independent to each other. One way to turn this measurement into a dynamic one is to consider its 'growth' between years, which can be achieved by the Malmquist-type indices, which were discussed in Chapter 3.

Cherchye et. al. (2007a) defines the notion of Performance Change in their dynamic assessment of EU internal markets. This change is defined as:

$$\begin{aligned} PC_{i,t} &= CI_{i,t} - CI_{i,t-1} \\ &= \sum_{q=1}^{Q} w^*_{q,i,t}(I_{q,i,t} - I^b_{q,t}) - \sum_{q=1}^{Q} w^*_{q,i,t-1}(I_{q,i,t-1} - I^b_{q,t-1}) \end{aligned} \tag{.10}$$

That is, the performance change of a DMU i at time t is the difference between the time t and time $t-1$ composite indicator score, which differs from Cherchye et. al. (2007b)'s ratio measure of change:

$$PC'_{i,t} = \frac{CI_{i,t}}{CI_{i,t-1}}, = \frac{\sum_{q=1}^{Q} w^*_{q,i,t}\left(\frac{I_{q,i,t}}{I^b_{q,t}}\right)}{\sum_{q=1}^{Q} w^*_{q,i,t-1}\left(\frac{I_{q,i,t-1}}{I^b_{q,t-1}}\right)}$$

From this, assuming the linear form of performance change hereafter, the measure can be decomposed onto multiple parts, according to van Puyenbroeck & Rogge (2017):

$$PC_{i,t} = \Delta OWN_{i,t} + \Delta BP_{i,t} + \Delta W_{i,t} \tag{.11}$$

Which consists of the change in DMU-Specific performance, $\Delta OWN_{i,t}$, the change in Industry Base performance, $\Delta BP_{i,t}$, and the change in Weights, $\Delta W_{i,t}$. These decomposed terms can then formally define performance change as:

$$\begin{aligned} PC_{i,t} =& \underbrace{\frac{1}{2}\sum_q [(w_{q,i,t} + w_{q,i,t-1})(I_{q,i,t} - I_{q,i,t-1})] +}_{\Delta OWN_{i,t}} \\ & \underbrace{\frac{1}{2}\sum_q [(w_{q,i,t} + w_{q,i,t-1})(I^b_{q,t} - I^b_{q,t-1})] +}_{\Delta BP_{i,t}} \\ & \underbrace{\frac{1}{2}\sum_q [(w_{q,i,t} - w_{q,i,t-1}) \left[[(I_{q,i,t} + I_{q,i,t-1}) - (I^b_{q,t} + I^b_{q,t-1})]\right]}_{\Delta W_{i,t}} \end{aligned} \tag{.12}$$

Performance Change Extensions

Other forms of the performance change can be evaluated as in Cherchye et. al. (2007b) and Rogge (2018), who considers a Fisher Ideal (or Bennet Ideal for the equivalent linear form) and a 'general model' that contains a sensitivity parameter, ρ, respectively. Following Rogge (2018), both of these points can be combined and then transformed to match the linear form of the change:

$$\begin{aligned} PC^{\rho}_{i,t} &= \frac{1}{2}\left[(PC^L_{i,t})^{\rho} + (PC^P_{i,t})^{\rho}\right], \\ &= \frac{1}{2}\underbrace{\left[\left[\sum_{q=1}^{Q} w^*_{q,i,t-1}(I_{q,i,t} - I^b_{q,t})^{\rho}\right]^{\frac{1}{\rho}} - \left[\sum_{q=1}^{Q} w_{q,i,t-1}(I_{q,i,t-1} - I^b_{q,t-1})^{\rho}\right]^{\frac{1}{\rho}}\right]}_{(PC^L_{i,t})^{\rho}} \\ &+ \frac{1}{2}\underbrace{\left[\left[\sum_{q=1}^{Q} w^*_{q,i,t}(I_{q,i,t} - I^b_{q,t})^{\rho}\right]^{\frac{1}{\rho}} - \left[\sum_{q=1}^{Q} w_{q,i,t}(I_{q,i,t-1} - I^b_{q,t-1})^{\rho}\right]^{\frac{1}{\rho}}\right]}_{(PC^P_{i,t})^{\rho}} \end{aligned} \tag{.13}$$

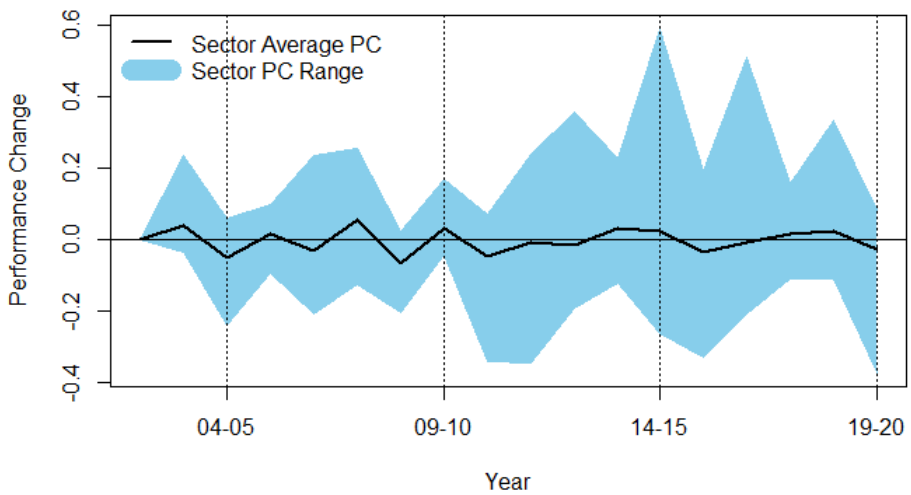
Where $(PC^L_{i,t})^{\rho}$ is the Laspeyres Performance Change, which uses time $t-1$ weights in all terms, and $(PC^P_{i,t})^{\rho}$ is the Paasche Performance Change, which instead uses time t weights in all terms. The equally weighted combination of both terms gives a Bennet Ideal index over the two time periods, with $\rho \in \mathbb{R}$ acting as the general sensitivity parameter.

The desire for a Bennet Ideal indicator is based around the idea of having symmetric contributions to the index between time periods. Rather than the simpler case of (7.4), the resulting components evenly incorporate information from both time periods through the weights, similar to how the decomposition (7.5) captures that effect as an isolated part of the total change. The general parameter ρ is similar to Atkinson (1970)'s measure of income inequality, and, analogous to the Atkinson index's inequality aversion parameter, ρ acts as an adjustment to account for the extreme values of the composite indicator's factors, with $\rho = 1$ reducing the measure to that of (7.4).

Taking the a subset of values for ρ from Rogge (2018), the full set of cases covered in this chapter for the parameter is $\{-\infty, 1, \infty\}$, with $\rho = \pm\infty$ leading to the use of $\max(\cdot)$ and $\min(\cdot)$ functions, respectively. Both extreme value cases can then be compared to the base $\rho = 1$ case of the parameter.

Results

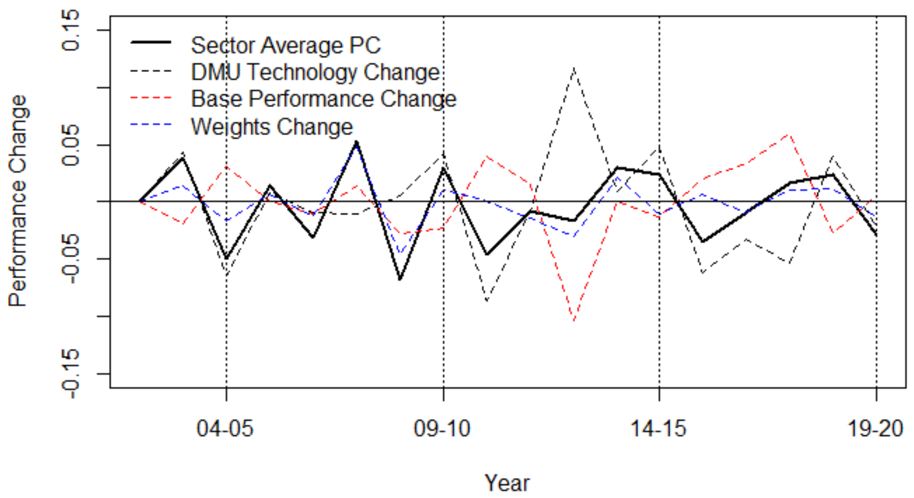
Figure .1: Average Performance Change, 2002/03 - 2019/20



Performance Change over the sample time period, and the range across companies in each year, are illustrated by Figure 7.4. As the graph shows, there is a fairly wide range of values of performance change across the industry, over most of the time period. These wide ranges are bi-directional, in that companies have faced large improvements in composite quality between years - maximally via a 58.8% increase by ANG from 2013/14 to 2014/15, and also large setbacks, with the largest decrease of 37.3% by NWL from 2018/19 to 2019/20. Averaging across the DMUs in each year, the industry average performance change is relatively small, and is centred around a 0% change, with the end of the sample displaying a slight decrease by 2.93% in quality from the previous year.

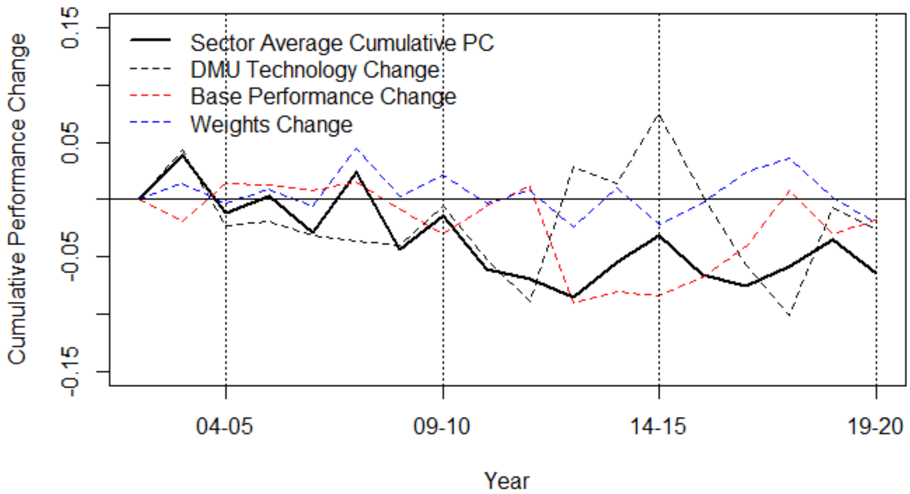
Looking more into what drives the changes between years, Figure 7.5 shows the decomposition of performance change into its three constituent parts: ΔOWN, ΔBP and ΔW.

Figure .2: Performance Change Decomposition, 2002/03 - 2019/20



The most interesting general trend is not tied to one specific part of the overall change. Seemingly, the DMU-specific changes and the industry-wide changes, ΔOWN and ΔBP, counteract each other, implying that any changes made by companies on average are mitigated by changes in the industry's state in that same year. Therefore, though the least impactful in general, the changes

Figure .3: Cumulative Performance Change Decomposition, 2002/03 - 2019/20



in performance due to changes in weights, ΔW, appear to drive the direction of the average industry change over time in most years of the sample.

Figure 7.6 further explains what drives the change in quality over time by showing the Cumulative Performance Change at time t:

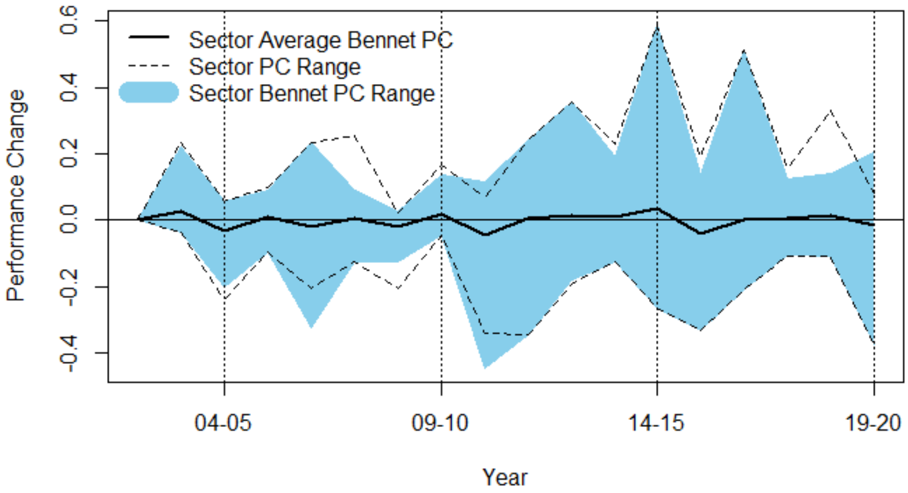
$$PCC_{i,t} = CI_{i,t} - CI_{i,1} \tag{.14}$$

In general, the notion that the changes to factor weights drive the sign of the change in performance still holds, but the cumulative change illustrates that quality has, over time, fallen by 6.39% relative to the starting period, beginning around the 2009/10 time period that began the PR09 regulatory period.

Some comparison with the initial performance change measure and the Bennet performance change can be seen in Figure 7.7, which compares the range of values across the companies between both measures of change.

The Bennet performance change, on average, is smaller in magnitude, owing to its derivation that 'smooths' the change between periods. The range, however, is broadly the same as the original indicator, with a maximum of 58.8% again by ANG from 2013/14 to 2014/15, and a minimum of -44.5% by YKY from

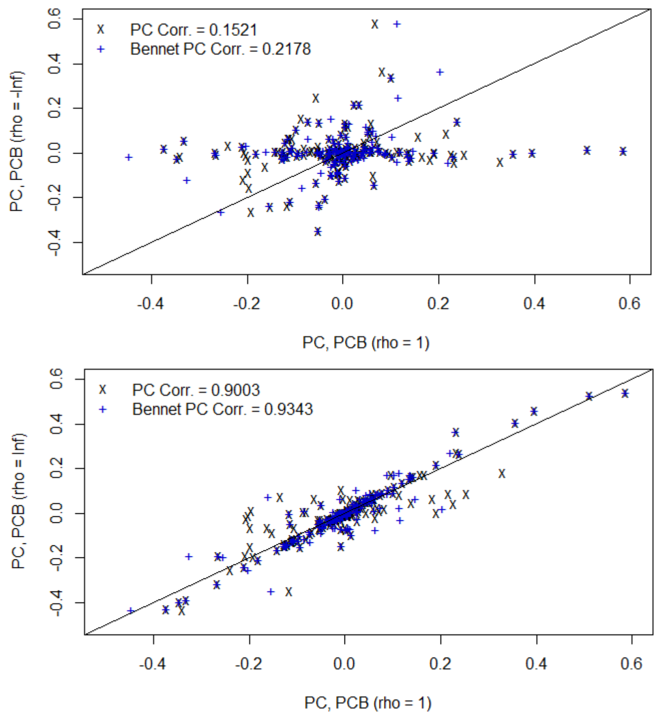
Figure .4: Average Bennet Performance Change, 2002/03 - 2019/20



2009/10 to 2010/11. Mechanically, then, it appears that there isn't too much difference between both versions of performance change, and so there is merit in then utilising the Bennet performance change, if the smoothness of the results between time periods is considered important. This indifference between the two change measures is further compounded by the very strong, positive correlation of 0.9046.

The Rogge (2018) models of performance change further extend the measure of performance change, by evaluating how sensitive the changes in quality are to the industry's minimal values ($\rho = -\infty$) and maximal values ($\rho = \infty$), illustrated by Figure 7.8. Correlating both cases to the initial $\rho = 1$ performance change measure, which in theory is indifferent between minimal and maximal values, it is found that the minima-sensitive changes have a moderate positive correlation of 0.1521, while the maxima-sensitive changes have a very strong positive correlation of 0.9003. Bennet versions of the same measures show slightly stronger correlations for both sensitivities, but don't change the overall interpretation of the results: it seems that, though the initial indicator should prioritise more average values of performance change, in practice the measures of performance change are far more in-line with the behaviour of the maximal

Figure .5: Rogge (2018) Minimum (Top) Sensitivity PC and Maximum (Bottom) Sensitivity PC, 2002/03 - 2019/20



values, rather than the minimal values.

Future Directions

The idea of inter-temporal indices is one that already bears familiarity in the water industry literature, with the indices discussed in Chapter 3 being good examples of exactly these types of measure. Given the similarity of this chapter's performance change measure - it's decomposability, the use of smoothing into an 'ideal' form, and the addition of a sensitivity parameter - to other indices in the research literature, an interesting way to develop the analysis of quality dynamics could be to follow the extensions of the aforementioned productivity and technology change metrics.

Take, for example, the collection of Malmquist index decompositions in Lovell (2003). Letting Inputs x^t and Outputs y^t at a timt t be defined for two time periods, t and $t+1$. Lovell (2003) derives three different expressions of an 'appropriate' Malmquist Index, based on other papers in the productivity index literature. These variants are defined as:

Ray & Desli (1997), $M_{oc} = TE\Delta \times T\Delta \times S\Delta$

5-Part Decomposition, $M_{oc} = TE\Delta \times T\Delta \times S\Delta \times OM\Delta \times IM\Delta$

5-Part with Activity Effect, $M_{oc} = TE\Delta \times T\Delta \times OB\Delta \times IB\Delta \times AE\Delta$

Where the subscripts o and c indicate the index's output orientation and use of benchmarking technologies respectively. $TE\Delta$ refers to Technical Efficiency Change; $T\Delta$ to Technical Change; and $S\Delta$ to the Scale Change Factor. $OM/B\Delta$ and $IM/B\Delta$ refers to Output and Inputs Mixes/Biases respectively, and lastly $AE\Delta = S\Delta \times OM\Delta \times IM\Delta$ is the Activity Effect. Lovell (2003) notes that while in theory the activity effect can be substituted out of the third index definition to create a 7-part decomposition containing the scale change effect, it is unwise in practice due to the difficulties in distinguishing between the effects caused by the mix and bias effects, whose roles are similar.

Another, similar idea is the Bjurek Productivity Index and its decomposition, as in Diewert & Fox (2017). This index follows the principle of Bennet- and Fisher-ideal indices between time periods, but for input and output orientation of the whole index. In this index model, there are Technical Progress, Technical Efficiency Change, and Returns-to-Scale components instead. All of the alternative index models are constructed via Distance Functions, which in practice are estimated by DEA models. Linking this to the composite indicator, though BoD estimation produces the weights of the composite indicator with an analogous model with dummy inputs, it seems passable in theory to build these distance functions with quality factors as the index inputs, and the resultant CI as an output. In the simplest case, performance change is then defined as the geometric equation of Cherchye et. al. (2007b) and, it seems, could then also be linearised as this chapter has done to produce an additive decomposition instead.

As with the previous future directions for extreme weather, another set of information that could be found for the betterment of the dynamic analysis of the composite indicator would come from time-series analysis. In doing so, not only could the behaviour of the dynamic trends be better understood over time, but one could also find convergence measures for the quality indicator, to see if its trajectory tends towards 100% compliance as its predecessors did, or if it diverges away towards an equilibrium that should cause immediate concern at a regulatory level.