Essays in the Measurement of Efficiency for the English and Welsh Water and Sewerage Industry

By

Charlotte Pointon

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Cardiff University

The Economics Section of Cardiff Business School

Cardiff University

November 2014

Declaration

This work has not previously been accepted in substance for any degree and

is not concurrently submitted in candidature for any degree.

Signed ………………………………………………………….. (candidate)

Date …………………………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of …………………………(insert MCh, Md, **MPhil,** PhD etc, as appropriate)

Signed ………………………………………………………….. (candidate)

Date …………………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references.

Signed ………………………………………………………….. (candidate)

Date …………………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………………………….. (candidate)

Date …………………………

Abstract

The English and Welsh water and sewerage industry was privatised in 1989 and is characterised by a series of regional monopolies. The majority of consumers currently have no choice in their supplier. The industry is regulated by Ofwat to guarantee the best value for customers whilst enabling the companies to undertake their activities. The motivation of this thesis is to examine the effectiveness of regulation.

The aim is to examine five research questions. Firstly, has regulation encouraged convergence amongst the efficiency scores? Secondly, have the 1999 and 2004 price reviews been effective in improving efficiency? Thirdly, is there a capex bias in the industry? The final two aims come from a methodological perspective: firstly, to allow for the incorporation of environmental variables within the measurement of efficiency and secondly, to incorporate the long asset life of capital by incorporating capital as an intertemporal factor of production.

Data Envelopment Analysis (DEA) is employed to measure efficiency which is a non-parametric technique that creates a linear frontier over the data. Convergence is examined by drawing from the growth literature to examine β- and σ-convergence. A three-stage DEA model is applied to examine the influence of environmental variables and to obtain an environmental adjusted DEA efficiency score. Finally, the intertemporal nature of capital is incorporated through a dynamic DEA model.

This thesis reports that whilst regulation has produced limited improvements in the average efficiency, regulation has been effective in encouraging the least efficient firms to catch up with the frontier companies. Ofwat's tightening of the price review in 1999 has produced significant improvements in efficiency, whereas the 2004 price review was relatively lax and had no significant influence. Finally, the thesis highlights that the current regulatory framework induces a preference towards capital expenditure which can have implications on the consumer's bill.

Acknowledgements

Firstly I would like to thank my supervisor, Kent Matthews for all of his help and advice in conducting this research. I would like to thank Paul Edwards and Mike Davis.

I have benefitted from attending Wageningen University Theory and practice of efficiency and productivity measurement summer school. I have benefitted from comments received at the 12th International DEA conference (Kuala Lumpur 2014), Quantitative Economics Doctorate (Bielefeld 2014) as well as from colleagues during the course and PhD workshops at Cardiff Business School.

I would like to thank all of my colleagues at Cardiff Business School, especially Jenna Pandeli, Lucy Minford, Mariyani Husairi and Chole Tarrabain.

I would like to thank my external examiner, Sir Ian Byatt and my internal examiner Patrick Minford for a challenging and enjoyable debate with regards to the level of inefficiency within the water industry and the presence of a capex bias.

Finally I would like to thank my family and Tom Beale for all of their support and encouragement.

Contents

List of Tables

List of Figures

1. Motivation and Research Questions

1.1 Motivation

The English and Welsh water and sewerage industry provides an essential service to households and non-households. The industry is characterised by a series of regional natural monopolies. The majority of consumers currently have no choice in their supplier. The industry was privatised in 1989 to encourage investment within an industry with deteriorating quality. Privatisation created ten local Water and Sewerage Companies (WaSCs) and 29 Water only Companies (WoCs). Littlechild (1988) highlighted that the firms have unrivalled monopoly power due to the capital intensive nature of the industry and the lack of substitutability of the services offered. Firms may exploit their monopoly power by increasing prices or reducing quality. Littlechild (1988) highlighted that, due to the 'monopoly par excellence', privatisation would require a permanent regulatory body and so Ofwat (Office of Water Services) was formed. The role of Ofwat is to protect the interest of customers by ensuring that water companies carry out their functions effectively whilst ensuring that they can finance their functions. Ofwat also promotes efficiency and facilitates the introduction of competition.

The industry is regulated under price-cap regulation based on $RPI + K$, where RPI is the retail price index and K is a company specific allowed increase in prices. The K factor is decomposed into $RPI + Q - X$ where X is an efficiency component to reduce prices and Q is a positive component to allow for high prices for quality improvements.

Littlechild (1988) highlights that the success of the $RPI + K$ system is dependent upon choosing the 'right level' of K. K reflects the quality expenditure and the scope of productivity improvements, but the potential productivity improvements are unknown to the regulator and therefore have to be estimated. If K is too large then the regulated companies will earn excess profits, which will undermine the support of price-cap regulation and it is not in the best interest of customers. On the other hand, if K is too small companies may not be able to finance their operations.

As the industry is characterised by a series of natural monopolies and the majority of consumers currently have no choice of suppliers, it is essential that Ofwat acts in the best interest of customers to encourage efficiency and to ensure that monopolies do not abuse their power. This thesis is motivated from two perspectives. Firstly it examines whether regulation has been effective in encouraging technical change and convergence. Secondly, the thesis is motivated from a methodological viewpoint to develop the measurement of efficiency, which is an intrinsic component of the price review.

The remainder of this chapter outlines the five research questions which this thesis aims to address. Section 1.3 will outline the contributions of the research to the current literature and finally, section 1.4 provides a route map of the thesis.

1.2 Research Questions

To examine the effectiveness of regulation and the measurement of efficiency this study aims to answer five main research questions which are outlined overleaf.

Research Question 1: Did the 1999 and 2004 price review improved efficiency?

This research question arises to examine whether the regulation has been effective in improving efficiency.

Research Question 2: Did the English and Welsh water and sewerage industry exhibit convergence in terms of efficiency performance in the period 1997- 2011?

The objective of this question is to examine whether the regulatory framework has been effective in encouraging convergence within the efficiency scores towards the frontier.

Research Question 3: How can the measure of efficiency using DEA incorporate non-discriminatory factors to allow differences in the local environment?

This question arises from a methodological perspective to incorporate environmental variables within the measurement of efficiency using DEA, which makes the implicit assumption of homogeneity.

Research Question 4: How should long life and indivisible capital stock be treated in the measurement of efficiency?

This question arises to address the incorporation of capital within the measurement of efficiency. The question leads to treatment of capital as an intertemporal and a quasi-fixed factor of production though a dynamic DEA model.

Research Question 5: Is there a capex bias within the industry?

This question arises to examine whether there is a preference towards capital expenditure (capex) instead of operating expenditure (opex) in the industry, which is highlighted by Ofwat.

1.3 Contributions of the Research

The aim of this thesis is to examine whether regulation has been effective in encouraging improvements in efficiency and technical change. At the top level this thesis examines whether the 1999 and 2004 price reviews have significantly improved efficiency. The 1999 price review was the first and only price review to date to assign an industry average negative K factor. On the other hand, the 2004 review imposed a K factor greater than the initial factor at privatisation, which was considered as lax. This thesis finds that the regulatory tightening in 1999 significantly improved efficiency but that the 2004 price review has had no impact.

The second contribution is to statistically determine if efficiency has converged within the English and Welsh water and sewerage industry. To determine whether relative efficiency has converged, beta- and sigma-convergence (β- and σ-convergence) is estimated. The results find convergence within the industry for variable costs, with the initial least efficient firms improving at a faster rate than the most efficient firms. The results for total costs indicate that the least efficient firms are growing at a faster rate; however, the dispersion of efficiency does not significantly decrease. The chapter finds limited evidence that the 2004 price review increased the speed of convergence.

To account for firm heterogeneities amongst firms, fixed effects are incorporated to measure conditional convergence. The measurement of conditional convergence highlights the differences within companies' operating conditions through differences in company specific steady state efficiency scores. The chapter highlights the importance of controlling for non-discretionary variables within the measurement of efficiency.

To account for these heterogeneities, the third contribution is to incorporate differences within the firms' operating environments within the measurement of efficiency using DEA. The basic DEA model assumes homogeneity and therefore assumes that all firms operate within the same environment and compares the level of inputs and outputs. Environmental variables are incorporated through a three-stage approach introduced by Fried et al (1999). The approach allows for the incorporation of several environmental variables whilst determining the impact of the variables upon the excess levels of inputs. The model adjusts the inputs for the influence of the environmental variables to obtain an efficiency score when controlling for differences within firms' operating characteristics. The results indicate a significant difference in the efficiency scores once controlling for the differences in firms' operating characteristics with a substantial number of companies appearing to be relatively efficient due to operating within a favourable operating environment.

The fourth contribution of the thesis is to model capital as an intertemporal factor of production through dynamic DEA. The industry is characterised by long-life assets whose useful life expand over several periods and cannot be adjusted to their optimal value instantaneously. The model takes into account decisions made today to influence future production. The thesis highlights the need to incorporate the intertemporal effects as the optimal capital value reported by the static model is underestimated because model does not take into account future production.

The final contribution examines the presence of a capex bias, which is the preference of capex solutions instead of opex solutions through examining the over-utilisation of capital and variable inputs. Dynamic DEA reports a persistent over-utilisation of capital and an improvement in the efficiency of variable inputs over the period. The capex bias may be present due to the Averch-Johnson effect if the rate of return on capital is greater than the cost of capital or due to the nature of the industry building for future demand.

1.4 Dissertation Structure

The dissertation is organised as follows:

Chapter 2 provides the contextual background for the English and Welsh water and sewerage industry. It provides an overview of the restructuring of the English and Welsh water and sewerage industry, from thousands of bodies undertaking water and sewerage in the 1940s, to privatisation and the current structure. The chapter outlines the motivation for privatisation and the introduction of economic regulation. An outline of the price-cap regulation is provided and a detailed outline of the determination of the allowed price increase.

Chapter 3 provides a comprehensive overview of the theory of efficiency and the different techniques which can be applied for its measurement. The thesis examines efficiency through the measurement of DEA and therefore the chapter provides a detailed outline of the measurement of efficiency using DEA.

Chapter 4 outlines the data used within the thesis for the measurement of efficiency for the English and Welsh water and sewerage industry. The data permits efficiency to be examined at different levels within the structure of the industry and the chapter provides a detail discussion of the choice of specification to measure efficiency. A review of the choice of variables used for the measurement of efficiency both within the English and Welsh water and sewerage industry and other water industries is drawn upon to inform the choice of input, output and environmental variables used within the thesis. Detailed definitions of the variables are provided alongside some descriptive statics.

Chapter 5 estimates the presence of β- and σ- convergence of the WaSC efficiency scores over the period examined. β-convergence is present if those companies with an initial low level of efficiency grow relatively faster. σ-convergence examines whether the dispersion of efficiency has reduced over time. The chapter incorporates heterogeneities amongst firms through a fixed effects model and examines whether the 2004 price review influenced the speed of convergence.

Chapter 6 incorporates environmental variables within the measurement of efficiency through a three-stage approach to incorporate the influence of the differences in the operating environment of the WaSCs. The second stage regresses the environmental variables upon the over-utilisation of inputs to determine their influence. Alongside environmental variables, regulatory dummy variables are included to determine the impact of the 1999 and 2004 price reviews.

Chapter 7 measures dynamic DEA by incorporating the intertemporal nature of capital within the estimation of efficiency as an input into today's production and an output into tomorrow's production process. Capital is incorporated as a quasi-fixed input which is held fixed for the estimation of static efficiency. This, therefore, allows overall efficiency to be decomposed into a static and dynamic component. The chapter examines the over-utilisation of capital and examines the issue of the perceived capex bias.

Finally, chapter 8 summarises the key aspects and findings of the research and outlines the main conclusions derived.

2. The Water and Sewage industry in England and Wales: Past, Present, Future

2.1 Introduction

The nature of the industry, which is characterised by a series of regulated regional monopolies, has motivated this study to examine whether the regulatory regime has been effective in encouraging technical change and convergence. This chapter will provide the contextual background of the English and Welsh water and sewerage industry, outlining the background of the industry and the regulatory tools for determining the price limits.

The industry was privatised in 1989 creating a series of local monopolies whereby households and non-households using less than five million litres of water a year in England and fifty million litres a year in Wales currently have no choice in their supplier. The firms are natural monopolies due to the capital intensive nature of the industry and lack of substitutability of the services offered. To ensure that firms do not exert their monopoly power the industry is subject to economic and quality regulation. Economic regulation is undertaken by Ofwat to act as a proxy for competition and to ensure firms are competitive whilst being able to finance their functions. The industry is regulated under price cap regulation, firms are allowed to increase their prices by $RPI + K$. RPI is a measure for the rate of inflation and K is a company specific composite term reflecting the scope for efficiency gains and higher costs for investment to improve quality.

Firstly, this chapter traces the evolution of industry, from its early consolidation with thousands of bodies to the restructuring of the industry creating ten Regional Water Authorities (RWA), through to privatisation. Section 2 will provide a brief outline of the motivations for the privatisation of the industry and the floatation of the industry. Section 3 will provide an outline of the current structure of the industry and the introduction of price cap regulation. The tools to determine the level of K within the periodic review and their mechanisms to encourage economic efficiency are outlined in section 4. Finally, section 5 examines some important features of the 2014 price review for the elimination of the capex bias and the facilitation of the introduction of competition within the industry.

2.2 History of the industry pre-privatisation

The English and Welsh water and sewerage industry was developed in the nineteenth century as a mixture of municipals and small private undertakings. In 1945, there were more than 1,000 bodies involved in the supply of water and around 1,400 bodies responsible for sewerage and sewage disposal (Ofwat, 2008). Due to the subtantial number of bodies involved, water resources planning was highly localised with little co-ordination at either a regional or national level. Ofwat (2008) state that the consolidation was required to provide better planning, to control pollution and meet the increasing demands for water. The water act in 1973 established ten RWA's in 1974 which were responsbile for the whole water cycle within their catchment area; both water and sewerage activities. The 1973 Water Act transferred the control of investment from local authorities to the central government. The RWAs operated under a cost recovery base, with capital investment expenditure requirements meet by the central government and from revenues. Alongside the RWAs there were 29 statutory Water Only Companies (WoCs) which operated under private ownership. Within the areas where WoCs operated the RWAs undertook the sewerage activities whilst the WoCs undertook water activities. The RWAs were substantially larger than the WoCs, the RWAs covered around 75% of the country (Saal and Parker, 2001).

The industry was privatised under the 1989 Water Act. Beesley (1994) states that the water and sewerage industry was privatised against a legacy of neglect. During the 1980s the quality of service deteriorated and in general the industry suffered from heavy underinvestment (Hunt and Lynk, 1995). Ofwat (2008) state that the 1985 river quality survey showed that for the first time since surveys were undertaken in 1958 the river quality deteriorated. Some 903km of the 40,000km of rivers surveyed reported a net deterioration in quality. In 1988, 742 out of 6,407 sewerage treatment works failed their discharge permit requirements. As a result of this lack of investment, a series of European Commision (EC) Directives were breached due to pollution incidents. Ofwat (2008) state that the government recognised the need for significant capital investment as a result of the decision by the EC to start prosecution proceedings against the government for non-compliance with two EC directives in the mid 1980s.

The government had to increase investment through either increased borrowings, increased taxation or through a progamme of privitization. Ofwat (2008) state that the government was relcutant to increase borrowing or taxation and privatisation was considered as a policy to promote efficiency following the privatisation of British Telecom and British Gas in 1984 and 1986 respectively.

The 1989 Water Act provided the mechanism for privatisation within the industry. The ten RWAs became publicly quoted as Water and Sewerage Companies (WaSCs). The WoCs were re-established as normal public limited liability companies. To ensure that the flotation of the privatisation of the WaSCs would attract investors, the government cancelled all of the long term debt owned by the water and sewerage companies at a total cost of £4.9bn (in 1989 prices) and provided a cash injection of £1.5bn known as the "green dowry" (Saal and Parker, 2001). In addition provisions were made for capital tax allowance of £7.7bn to ensure that companies were not disadvantaged with companies who had built up capital allowances over time (Ofwat, 2008).

Sawkins (2001) highlight that the private sector successor of the RWAs face an 'unrivalled' degree of monopoly power due the ownership and control of water and sewerage networks and lack of subsituability of the services they offer. The ownership and control of water and sewerage networks cause high barriers to entry in the industry due to the large networking costs involved in the distribution of water. The water mains and sewers are highly capital intensive and for this reason it is not feasible to have multiple networks throughout England and Wales. Alongside the physical infeasibility is it also economically infeasible as the value of service for water or sewerage activities is very low incomparison to the cost of infrastructure and distributing water and sewerage.

Littlechild (1988) highlights that monopoly power may be exerted by increasing prices, reducing quality or quality of service, allowing for environmental deterioration or allowing for efficiency to decline. Littlechild (1986) report that due to 'the natural monopoly par excellence' privatisation would require a permanent regulatory settlement. The 1989 Water Act provided the Secretary of State the responsibility of drinking water quality (Drinking Water Inspectorate, DWI). The National Rivers Authority (NRA) now the Environment Agency (EA) and National Resources Wales (NRW) would manage pollution and the environment and finally the water act defined the duties of the Director General of Water Services as the economic regulator, Ofwat.

2.3 Current structure of the industry

Within the post-privatised struture of the industry there are two types of firms: WaSCs and WoCs. WaSCs undertake both water and sewerage activities whilst WoCs only undertake water activities. WaSCs and WoCs are vertically integrated companies; for water services a single company extracts, treates, distributes and retails the water, whereas for sewerage activities a given company collects, treats and disposes of the sewage. The WaSCs provide all services within their geographic area and the WoCs provide the clean water operations only; the collection and treatment of dirty water is undertaken by a WaSC. The firms are therefore characterised as natural monopolies where only one firm deliveres either water or sewerage activities or both within a specific region. As previously highlighted natural monopolies arise due to high barriers to entry into the industry; due to the large networking cost involved in the distribution of water. To ensure firms do not abuse their monopoly power the industry is regulated by Ofwat.

There are two well-rehearsed forms of price control: price-cap regulation or rate of return regulation. Rate of return regulation determines a fair return on investment; the Rate of Return (ROR), otherwise known as the cost of capital by Ofwat. The revenue requirement (RR) shown in equation 2.1 is the amount each company needs to collect to finance its operating expenditure and capital programme whilst earning a return on previous capital investment through the rate of return and paying tax. The ROR is multiplied by the rate base to obtain the return on capital for investors. The rate should be high enough to attract investment whilst ensuring a fair price for customers. Once the revenue requirement is decided the regulator determines a tariff structure to recover aggregate costs.

$$Revenue\ Requirement = Total\ Costs = Variable\ Costs + ROR \times Rate\ Base \quad (2.1)$$

On the other hand, price-cap regulation determines the allowed percentage increase in prices under the $RPI - X$ framework of Littlechild (1983) and Beesley and Littlechild (1989). A regulated company has an annual price cap based on the previous year's allowance plus an inflation allowance (RPI) less a company specific efficiency target (X). The annual percentage difference between the revenue requirement and the base year revenue expected from customers is the price limit. The main distinction between the two approaches is that prices are fixed under price-cap regulation and prices change as costs change under rate of return regulation.

Littlechild (1983) and Beesley and Littlechild (1989) highlight the advantages of $RPI - X$ compared to rate of return regulation. Firstly, prices under $RPI - X$ are fixed, and therefore firms have greater incentives for production efficiencies as they are able to retain any profits they earn during the period. Once the price is fixed, companies aim to maximise profits by minimising costs. At a future date the price limit is revised, which is expected to fall due to the efficiency savings. Helm and Rajah (1994) state that the efficiency savings are likely to yield long-term benefits for the customers if the price review period is long enough and the regulator refrains from intervening. Although rate of return regulation is less effective in providing incentives, it does allow for the regulator to adjust prices more quickly and therefore prices are cost-reflective. Secondly, firms experience an incentive to over-invest in capital under rate of return regulation known as the Averch-Johnson effect.

Averch and Johnson (1962) demonstrate that firms which are regulated by rate of return regulation have an incentive to deviate from the cost minimising proportion of inputs if the allowed rate of return is greater than the cost of capital, specifically with regards to over-capitalisation. The disproportionate level of inputs arises when firms only earn the return on capital and not on operating expenditure and therefore prefer capital based solutions. Overall, Beesley and Littlechild (1989) highlight that operating and capital expenditure would be lower under a price cap regulation than under an annual rate of return regulation. However, as a result of the incentive to minimise costs, Cowan (2002) highlights that firms using price-cap regulation have an incentive to cut quality. To ensure that companies are minimising costs whilst maintaining environmental and drinking water quality standards the industry is regulated by the DWI, EA and NRW.

The privatisation programme in the UK in the 1980s was underpinned by price-cap regulation using the $RPI - X$ methodology of Littlechild (1983) and Beesley and Littlechild (1989). Ofwat extended the price-cap regulation of $RPI - X$ to $RPI + K$ in which the company specific K factor is decomposed into two elements a positive Q factor to allow for price increases due to higher costs from tightening quality standards by the DWI, EA and NRW and a negative X factor for efficiency change. Ofwat undertakes a price review every five years, known as the price review, to determine the level of price increase for each company.

As well as encouraging efficiency, Ofwat has to ensure that firms are able to finance the their functions, mainly through a reasonable return on capital which is unique to the water sector[1]. They are also responsible for the interest of customers and to facilitate, but not promote, effective competition.

The K factor is determined through the revenue requirement which is the sum of the operating expenditure, capital charges and return on capital. The return on capital is calculated as the product of the cost of capital determined by Ofwat and the rate base, the Regulatory Capital Value (RCV). The RCV is a regulatory tool which permits a rate of return on the investments made by the company's owners. The RCV is the value placed on companies 200 days after privatisation and is rolled forward based on the amount of capital expenditure in each period less the depreciation charges.

To determine the scope of efficiency improvements (X) Ofwat employs yardstick competition. Yardstick competition, also known as comparative efficiency, was formulised by Shleifer (1985) in which the price of a regulated company is dependent on the costs of firms in the same industry. The paper suggests the use of regression analysis to observe cost differences between the companies operating within heterogeneous operating environments. Shleifer (1985) argued

[1] Helm and Jenkinson (2003, p161)

that yardstick competition would reveal the cost-minimising potential of each firm and secondly the comparison of costs of similar firms would provide a set of 'yardsticks' of performance which could infer any firm's attainable cost levels.

Littlechild's (1986) report to the Department of Environment highlights that the privatisation of the industry creates ten monopolies, therefore providing the opportunity to make regulation more effective by making comparisons of the companies. The set of 'yardsticks' are difficult to determine due to information asymmetries between the regulator and the companies and accounting for heterogeneities amongst companies. Ofwat undertakes comparative efficiency through a series of functional level regression, which are outlined in section 2.7.2.

Within the measurement of comparative efficiency an important feature is the number of available comparators. To obtain valid estimates of the heterogeneities within the industry sufficient degress of freedom are required. Therefore, a lack of comparators can undermine the robustness of yardstick competition. Cowan (2003) states that the boundaries of the WoCs are rather arbitary and mergers between them might be expected to produce efficiency gains. However, under section 32 of the Water Industry Act 1991, mergers between water companies must be assessed by the Competition and Markets Authority (CMA) previously the Competition Commission (CC) where the turnover of each water enterprise exceeds £10m. In practice, any merger within the industry would be referred to the CC (Ofwat, 2008).

The CMA examines the trade off between whether the reduction in the number of comparators within the industry will effect Ofwat's ability to undertake yardstick competition and the potential efficiency gains as a result of the merger. The MMC will allow the merger if they believe there is potential for considerable efficiency gains which are passed onto the customers through lower prices and the impact on Ofwat's ability to undertake yardstick competition is minimal. The number of WoCs within the industry has decreased from 29 after privatisation to eight, and these have been due to a series of mergers either between two WoCs or a WaSC and a WoC. The mergers and the dates are shown in table 2.1. There have been no mergers of two WaSCs within the industry. In 1996, the MMC rejected the bid of Wessex Water and Severn Trent for the acquisition of South West Water as the price reductions would not be enough to compensate the loss of the comparator (Cowan, 2003)

Table 2.1: Mergers and acquisitions in the English and Welsh water and sewerage industry

Date	*New Company*	*Companies Involved*
March 1993	North East Water	Newcastle and Gateshead Sunderland and South Shields
September 1993	Severn Trent	Severn Trent Water East Worcester
March 1994	Essex and Suffolk	Essex Water Suffolk Water
March 1994	Three Valleys	Colne Valley Lee Valley Rickmansworth
April 1994	South East Water	Eastbourne Mid Sussex West Kent
July 1994	Bournemouth and West Hants	Bournemouth & District West Hampshire Water
April 1996	Sutton and East Surrey Water	East Surrey Water Sutton and District Water
April 1996	Northumbrian	Northumbrian North East
April 1997	Dee Valley	Chester Waterworks Wrexham Water
January 1999	South East Water	Mid Southern South East
April 2000	Anglian	Harlepool Anglian
April 2000	Yorkshire	Yorkshire York Waterworks
April 2000	Northumbrian	Northumbrian Essex and Suffolk
October 2000	Three Valleys	Three Valleys North Surrey
December 2007	South East	South East Mid Kent
April 2013	South Staffordshire	South Staffordshire Cambridge

2.4 Regulation

The $RPI + Q - X$ price cap was driven by the large on-going requirements for companies to improve the quality of service. The initial K factors after privatisation were set by the Secretary of State for a period of 10 years. Upon privatisation the weighted industry average allowed increase in prices was 3.7% above the rate of inflation. Sawkins (1995) states that several finance directors within the industry believed the initial K factor was generous for political reasons; to ensure the success of privatisation. Within the first two years of privatisation companies' profits were exceeding forecasts, so Ofwat announced that it would review price limits for all companies in 1994 (Ofwat, 2008). The 1994 price review under the control of Ofwat instead of the government introduced new tighter price caps in line with the publically announced criteria (Sawkins, 1995). Since 1994, Ofwat has undertaken the review for the price limits, known as price reviews, every 5 years[2]. Table 2.2 shows the K factors for the final determinations in 1989, 1994, 1999, 2004 and 2009 for the industry average and the average for WaSCs and WoCs. Since 1989, the K factor has led to an increase in prices with the exception of the 1999 price review. The increase in capital expenditure and resulting operating expenditure has offset the scope for efficiency savings. The K factor has decreased until the 2004 price review; the industry average K factor in 2004 was higher than the initial value set at privatisation. Regulation tightened once again in 2009 saw a substantial fall in the K factor, where the companies' average price increase was 0.5% higher than RPI.

Companies can appeal to the Competition and Market Authorities (CMA) if they believe the final determination does not allow for their firm to finance their functions. Companies can also

[2] Prices are set for the year starting the 1st of April.

appeal with regards to interim adjustments to the K factor and amendments to company licenses.

Table 2.2: Final Determination K factors

Price Review	Allowed K Factors
Price Limits 1989 (1990/91–1999/00)	Weight Industry average of 3.7. WASCs had a weighted average of 3.9 and 1.9 for WOCs
PR94 (1995/96–2004/05)	Weighted average K factor of 0.9 for 10 years 1.0 for WASCs and -0.4 for WOCs. For the first 5 periods of the review 1996/96–1999/00 K factor 1.4 for the industry and 1.5 for WASCs and 0.6 for WOCs
PR99 (2000/01–2004/05)	Industry Weighted average -2.1. The WASC industry average of -2.0 and -2.8 for WOCs
PR04 (2005/06–2009/10)	Industry weighted average for 4.2. WASCs average of 4.3 and 3.1 for WOCs
PR09 (2010/11–2014/15)	Industry weighted average of 0.5. WASC average 0.5 and 0.3 for WOCs

2.5 Environmental Regulation

The industry is regulated by the DWI for drinking water quality and NRW and EA for environmental quality. The K factor in the majority of the price reviews has been positive, therefore indicating that the increased expenditure for quality improvements has outweighed the scope for efficiency improvements. Since privatization the English and Welsh water and sewerage industry has delivered £98 billion of investment (Ofwat, 2012). This investment has improved standards of customer service, increased the stability of supply and delivered improvement to drinking water quality and the environment. The main drivers of capital expenditure were the adoption of the Urban Wastewater Treatment Directive (UWWTD) which increased quality standards for the minimum standard of wastewater treatment and disposal of sludge. The DWI imposes increasing tight quality standards within the industry. There have been other quality drivers such as the Habitats Directive, Bathing Water Directives. Although firms have to comply with quality standards, Ofwat assesses company's investment plans to ensure that the company's capital expenditure for quality is efficient.

2.6 Determining Required Revenue

Ofwat's determination of prices is set by a five year price cap which is determined by the price review. Ofwat's determination is based on a building block approach to determining the "required revenue" whereby there is individual assessment of:

- Operating Cost
- Capital charges (i.e. depreciation)
- Return on capital

The annual percentage difference between the revenue requirement and the base year revenue expected from customers is the price limit. The following sections will consider the assessment of operating and capital expenditure within the price review.

2.6.1 Capital Expenditure

The level of capital expenditure is influenced by the quality standards imposed by the DWI, EA and NRW. To control the level of capital expenditure companies provide investment plans; Asset Management Plan (AMP). Ofwat challenges the AMP to ensure that firms are providing quality improvements at the best value for customers. Customers pay for capital expenditure over the life of the asset through capital charges. The capital charges are the sum of the Current Cost Depreciation (CCD) and Infrastructure Renewal Charge (IRC). CCD is the depreciation of non-infrastructure assets which are above ground assets based upon the asset life. IRC is an annualised cost of maintaining underground assets charged to the profit and loss account. The IRC is a fifteen year average of the infrastructure renewal expenditure.

Companies finance previous capital investment through a return on capital. The return on capital is comprised of the company's RCV and the cost of capital determined by Ofwat. A fair return on capital is required to attract investment within the industry; the cost of capital has to be low enough to avoid firms earning windfall gains whilst high enough to attract investors.

Ofwat determines the cost of capital by applying the Capital Asset Pricing Model (CAPM) which is crossed checked with the Dividend Growth Model (DGM)[3] (Ofwat, 2009) alongside discussions with firms, other regulators, consultants and shareholders. The CAPM considers the cost of debt and equity, risk-free rate and the equity beta value which measures the

[3] The 1994 price review used both the CAPM and DGM model.

company's exposure to systematic risk. To reflect the limited access to capital Ofwat includes a small company premium. The assumed cost of capital over price reviews is shown in table 2.3. The cost of capital reduced significantly in the 1999 price review, which contributed towards the negative K factor in the price review.

Table 2.3: Cost of Capital

Price Review	Cost of Capital-Post Tax	Small Company Premium
1994	5%–6%	0.75%
1999	4.25%–5.25%	0.4%–0.75%
2004	4.2%–5.3%	0.75%
2009	4.5%	0.1%–0.4%

2.7 The assessment of operating and capital expenditure

The level of operating and capital expenditure companies are allowed is subject to an efficiency challenge by Ofwat. Ofwat analyses the scope for efficiency savings to determine the X factor within the price setting to stimulate companies to reduce costs without reducing the level of service to customers and the environment. The methodology for determining the efficient level of operating and capital expenditure until and including 2009 price review is outlined. This is followed by a brief assessment of the approach used in 2014 price review to overcome the capex bias and to facilitate the introduction of competition.

The approach for determining efficient expenditure until and for the year 2009 analyses efficiency savings for operating and capital expenditure separately. Ofwat (1994) states that they find no convincing evidence that high operating expenditure can be explained by low capital expenditure and vice versa. Ofwat analyses the scope of efficiency savings for opex and capex into two components: firstly, a component for continuing efficiency which assumes that all companies can reduce costs by a given percentage, and secondly, a catch-up efficiency assumption based on firms relative efficiency to encourage inefficient firms to catch up to the frontier companies.

2.7.1 Continuing Efficiency

The continuing efficiency component is based on the director's judgement of the amount in which the frontier company can improve. Table 2.4 reports the continuing efficiency percentage for operating expenditure, capital maintenance and capital enhancement assumed within the price review. Ofwat (2004) highlights that the efficiency targets have reduced over the price reviews which are as a result of the reduced scope for overall efficiency gains.

Table 2.4: - Continuing efficiency assumptions

Price Review	Opex Continuing Efficiency	Capital Maintenance continuing efficiency assumption	Capital Enhancement continuing efficiency assumption
1994	2% a year 1995–00 1% a year 2000-05	1%	1%
1999	1.4% a year	1.4%	2.1%
2004	Water 0.3% Sewerage 0.5%	0.5% Water 0.6% Sewerage	0.74% Water 0.88% Sewerage
2009	0.25%	0.4%	0.4%

2.7.2 Catch-up Factor- Econometric Models

The catch-up efficiency assumption is derived through the application of yardstick competition to determine the company's relative efficiency to the frontier company. Opex efficiency is examined through a series of functional level econometric models. Capital maintenance efficiency is examined through both econometric models and a cost base challenge. Efficient capital enhancement expenditure is determined through cost base models. The catch-up factor assigns larger efficiency challenges to those firms which are deemed as inefficient to encourage catch-up to the frontier. Ofwat's 1994 assessment of efficiency was analysed separately for water and sewerage activities at both the activity and functional level through a series of single year cross-sectional regression and unit cost models. The models have been refined over the reviews and the capital maintenance models were used until 2004 whilst the opex models were utilised until the 2009 review.

Operating Cost Efficiency Models

For water, four econometric models are used for four functional areas which are:

1. Water distribution
2. Resources and treatment
3. Power
4. Business activities

For waste, the functional break down is:

1. Network
2. Business activities
3. Large Sewage Treatment Works (LSTW)

4. Small Sewage Treatment Works (SSTW)
5. Sludge disposal

These are modelled by regression with the exception of SSTW and sludge disposal which are analysed by unit cost analysis.

Capital Maintenance Efficiency Models

Water Service Models:

- Water resources and treatment
- Water distribution infrastructure
- Water distribution non-infrastructure
- Water management and general

Sewerage Service Models:

- Sewerage infrastructure
- Sewerage non-infrastructure
- Sewage treatment
- Sludge treatment and disposal
- Sewerage management and general

Within the econometric models actual expenditure is regressed on a series of explanatory variables to predict the efficient level expenditure for a given firm. The set of explanatory variables includes the cost drivers, such as the amount of distribution input alongside differences in the operating characteristics. The unit cost model derives the predicted costs based upon the industry average unit cost. The actual modelled costs and predicted costs are summed to the activity level to obtain the relative efficiency for water and sewerage activities. The modelled costs exclude third party services, service charges, local authority rates and exceptional items which are deemed as uncontrollable by Ofwat. Post modelling adjustments are made to account for Company Specific Factors (CSF) which are expected to increase costs and are considered as material[4]. Average efficiency is calculated as the ratio of modelled costs divided by predicted costs. To obtain a measure from the frontier, Ofwat applies Corrected Ordinary Least Squares (COLS), which shifts the regression downwards to the frontier company. The relative efficiency of a firm is therefore calculated as the distance of the firm's actual costs to the frontier. Ofwat determines the frontier company based on the companies' relative efficiency alongside other criteria such as the firm's size and whether they operate under any special characteristics to ensure that it is a suitable benchmark. To determine the efficiency challenge for operating and capital expenditure firms are banded A–E based on their distance from the chosen frontier company and firms close a proportion of the distance.

Table 2.5: PR04 Catch-up Factor Banding and Efficiency calculation

Band		Half Bands		Assumed catch-up (five-year total) %	Mid Point x catch-up assumption	Annual Percentage
		Frontier or better		0	0	0
A	0-5%	Lower	0-5%	60% of 2.5	1.5	0.3
		Upper	5-10%	60% of 7.5	4.5	0.9
B	5-15%	Lower	10-15%	60% of 12.5	7.5	1.5
		Upper	15%-20%	60% of 17.5	10.5	2.1
C	15-25%	Lower	20%-25%	60% of 22.5	13.5	2.7
		Upper	25%-30%	60% of 27.5	16.5	3.3
D	25%-35%	Lower	30%-35%	60% of 32.5	19.5	3.9
		Upper	35%-40%	60% of 37.5	22.5	4.5
E	5%-45%	Lower	40%-45%	60% of 42.5	25.5	5.1

[4] The qualifying threshold for CSF's to be accounted for is 1% of total modelled operating expenditure for each service. Ofwat (2008) state small claims that are not material will be offset by other small benefits.

Table 2.5 shows the banding and efficiency assumptions for opex at the 2004 price review. Ofwat then divides the bands into half-bands and takes the mid-point of each half-band as representative of all companies within the band. The efficiency challenge is determined based on closing a proportion of the efficiency gap to the frontier company. The proportion of the gap in which firms have to close is based on Ofwat's assumption of the catch-up factor. In the 2004 price review firms, efficiency targets for opex were based on the catch-up assumption of closing 60% of the gap to the frontier company for base operating expenditure to determine the annual percentage reduction in opex which is shown in table 2.5. The same methodology was applied for capital maintenance, although the banding and catch-up factor percentages differ. The catch-up assumptions for each price review are reported in table 2.6. It is important to note that from 2004 onwards, Ofwat reduced the residuals by 10% and 20% for water and sewerage services respectively to account for underlying errors in the models.

Table 2.6: Comparative Efficiency Catch-up

Price Review	Operating Expenditure	Capital Maintenance	Capital Enhancement
1994	25-35%		
1999	60%	Between 40% and 50%	75%
2004	60%	40% Econometric Benchmark 50% Cost Base Benchmark	75%
2009	60%	N/a	N/a

2.7.3 Catch-up Factor- Cost Base

In additional to the application of econometric models Ofwat analyse firms' relative capital maintenance and capital enhancement efficiency through a cost base assessment. The cost base is a set of capital unit cost estimates for standardised projects or units of work typical for the water industry; standard costs. Companies provide audited estimates for the cost base and companies are compared in order to gauge their relative efficiency. Benchmark firms are chosen for standard costs or groups of standard costs based on their relative efficiency alongside other factors to ensure that the firm is a suitable benchmark. A company's efficiency challenge is calculated based upon closing a proportion of the gap between the benchmark company and the company's standard costs. The proportion in which companies have to close the gap is known as the catch-up factor, displayed in table 2.7. The cost base was introduced in the 1994 price review and continued to be used until the 2004 price review.

Table 2.7: Cost Base Catch-up Factor

Price Review	Capital Maintenance	Capital Enhancement
1994	50%	50%
1999	50%	75%
2004	40% For Econometric Models 50% Cost Base Models	75%

2.7.4 Capital Incentive Scheme

The assessment of firms' capital expenditure in the 2009 price review was based on menu regulation; the Capital Incentive Scheme (CIS), rather than econometrics and unit cost analysis. The motivation was to encourage truth telling and to incentivise companies to put forward challenging and efficient business plans. Ofwat establishes an independent baseline of what an efficient firm should spend and firms put forward their capital programmes. The ratio of the company's bid and Ofwat's independent base is the CIS ratio, and depending on the ratio, firms are given incentives to reduce costs below their bid. The matrix in figure 2.1 shows the incentive rates; the lower the CIS ratio (the company's bid is lower than Ofwat's) the more attractive the incentive rates are for the given firm to deliver the given capex programme at a lower cost. The varying incentive rates encourage firms to submit realistic and challenging bids. The incentive rate is decomposed into an efficiency incentive and an additional income. The additional income is a fixed amount which depends on the CIS ratio. The greater the challenge firms set themselves, the higher the additional income, irrespective of whether they meet the challenge. If the baseline is above 100, the business plan is greater than Ofwat's baseline. This additional income is negative and therefore penalises those companies whose baseline is higher than Ofwat's assessment. The efficiency incentive is a symmetric treatment of over and under spend of actual expenditure relative to the allowed expenditure. If a firm under-spends it keeps a proportion of the amount but if it overspends it is penalised by a proportion of this additional cost.

Figure 2.1: CIS Matrix

CIS ratio (company: baseline)	80	85	90	95	100	105	110	115	120	130
Efficiency Incentive	45.00%	41.25%	37.50%	33.75%	30.00%	27.50%	25.00%	22.50%	20.00%	15.00%
Allowed Expenditure	95.00	96.25	97.50	98.75	100.00	101.25	102.50	103.75	105.00	107.5
Additional Income	1.00	0.89	0.69	0.39	0.00	-0.41	-0.88	-1.41	-2.00	-3.38
Actual Expenditure										
70	12.25	11.72	11.00	10.09	9.00	8.19	7.25	6.19	5.00	2.25
80	7.75	7.59	7.25	6.72	6.00	5.44	4.75	3.94	3.00	0.75
85	5.50	5.53	5.38	5.03	4.50	4.06	3.50	2.81	2.00	0.00
90	3.25	3.47	3.50	3.34	3.00	2.69	2.25	1.69	1.00	-0.75
95	1.00	1.41	1.63	1.66	1.50	1.31	1.00	0.56	0.00	-1.50
100	-1.25	-0.66	-0.25	-0.03	-0.00	-0.06	-0.25	-0.56	-1.00	-2.25
105	-3.50	-2.72	-2.13	-1.72	-1.50	-1.44	-1.50	-1.69	-2.00	-3.00
110	-5.75	-4.78	-4.00	-3.41	-3.00	-2.81	-2.75	-2.81	-3.00	-3.75
115	-8.00	-6.84	-5.88	-5.09	-4.50	-4.19	-4.00	-3.94	-4.00	-4.50
120	-10.25	-8.91	-7.75	-6.78	-6.00	-5.56	-5.25	-5.06	-5.00	-5.25
130	-14.75	-13.03	-11.50	-10.16	-9.00	-8.31	-7.75	-7.31	-7.00	-6.75
140	-19.25	-17.16	-15.25	-13.53	-12.00	-11.06	-10.25	-9.56	-9.00	-8.25

Source: Ofwat (2009)

2.8 Industry going forward

2.8.1 Capex Bias

Efficiency within previous price reviews has been examined for opex and capex separately. In the 2014 price review, Ofwat undertook a Total Operating Expenditure (totex) approach, examining opex and capex together in order to eliminate a perceived capex bias. The perceived capex bias is the view that WaSCs and WoCs have an inappropriate preference for expenditure on capital expenditure over day-to-day operating expenditure. If the capex bias does exist this is important for the regulator as companies may not be spending customers' money in the most efficient way. Incentives are used within the industry to encourage companies to generate outcomes that customers and society needs, want and are willing to pay for and to do so efficiently (Ofwat, 2011a).

Ofwat (2011b) have highlighted some potential drivers of the capex bias. The first potential driver is the difference between the strength of financial incentives between operating and capital expenditure. The second is the return on capex; capital expenditure is remunerated through the RCV which earns a rate of return whereas opex is recovered from the customer within the year pounds for pound. The Averch-Johnson effect shows incentives to over-capitalise if the rate of return is greater than the cost of capital.

There are other potential drivers that are considered including financing and ownership where the RCV is symbolised as company growth. To overcome the potential capex bias, within the 2014 price review Ofwat assess efficient expenditure for operating expenditure and capital expenditure together, known as totex, to equalise the incentive rates. To overcome the potential bias generated by companies earning a return on capital, Ofwat have introduced a Pay As You Go (PAYG) Rate, whereby firms earn a return on a fixed proportion of totex.

2.8.2 Retail Competition

Moving forward, the aim of the regulatory programme is to encourage the introduction of competition for retail activities for all non-household consumers from 2017. Ofwat has facilitated the introduction of downstream retail competition through introducing separate price controls for retail and wholesale elements of the production process. Entrants can apply for either a retail licence, which allows the licensee to purchase a wholesale supply from an appointed company and supply the premises of its customers, or a combined supply licence which enables the licensee to introduce water into a supply system and supply the customers.

With regards to the wholesale price limits, Ofwat intends to set separate price controls for water and sewerage activities. The price controls differed previously as Ofwat proposed to set controls that limit each company's total wholesale revenue rather than charges. This gives the companies more responsibility and accountability for the wholesale charges they will have to publish going forward.

The proposed retail price limit is set using the industry Average Cost To Serve (ACTS) with a glide path down to the lowest ACTS within the industry. The ACTS is calculated as the total retail costs divided by the number of retail customers. Ofwat has also introduced other incentives to incentivise companies to increase the amount of water traded.

2.9 Summary

This chapter has outlined the contextual framework of the English and Welsh water and sewerage industry. The water and sewerage industry has developed from thousands of bodies undertaking water and sewerage to the consolidation of the industry into ten RWAs and 29 WoCs.

The privatisation of the industry attracted much needed capital investment to improve deteriorating quality standards. The privatisation of the natural monopolies resulted in the introduction of the regulatory framework for both prices and quality standards. Ofwat sets prices under price-cap regulation based on $RPI + Q - X$. The level of efficient expenditure is determined through yardstick competition and menu regulation. The approach is designed to encourage both continuing efficiency improvement and catch-up to the frontier.

If regulation has been an effective tool the industry should experience productivity improvement and convergence towards the frontier. The privatisation and regulation of the English and Welsh water and sewerage industry has inspired a large body of literature that examines the effectiveness of regulation and privatisation. This study will examine whether the industry has converged and will examine the effectiveness of the 1999 and 2004 price reviews. Finally the study will examine the presence of the capex bias. The following chapter will outline the methodology employed for the measurement of efficiency.

3. Economic Efficiency: The Theoretical Perspective

3.1 Introduction

The main contribution of this chapter is to outline the theoretical framework and the different techniques for the measurement of productivity and efficiency. Chapter 2 provided the contextual framework and the motivation for measuring performance. The measurement of the performance of a firm is a relative concept examined either across an industry or a period of time. Performance is measured by either productivity or efficiency; this chapter will define the two concepts and extensively review the measurement of efficiency. Within the measurement of efficiency the unit of analysis is referred to as a DMU, Decision Making Unit. This term allows for flexibility, whereby a DMU may refer to firm or a set of departments of a firm.

Koopman (1951) defined a producer as efficient if it maximises outputs given inputs. This chapter outlines the economic background of efficiency and different techniques for the measurement of efficiency. Technical efficiency examines the relationship between inputs and outputs. Economic efficiency is outlined estimating either a cost or revenue. The measurement of economic efficiency requires the availability of input prices and behavioural assumptions with regards to cost minimisation or revenue maximisation. Economic efficiency can be decomposed into technical and allocative efficiency. Allocative efficiency examines the correct combinations and inputs and outputs given input and output prices.

Section 3.6 outlines several methodologies for the measurement of efficiency; a cost function, SFA (Stochastic Frontier Analysis) and DEA. The thesis employs three extensions on the basic DEA model to measure efficiency. The technology of DEA is based on the seminal work of Farrell (1957) and elaborated by Charnes et al (1978) and Banker et al (1984). DEA is employed to estimate convergence, the influence of environmental variables and dynamic efficiency. An extensive introduction of DEA is outlined in section 3.6.5.

3.2 Theory of Efficiency

The concept of efficiency is derived from the basis of the microeconomic theory of the firm. The conventional neoclassical theory treats the firm as a black box which transforms resources into saleable goods. The transformation of inputs into outputs is described by a production function or production possibility set. The conventional neoclassical theory of the firm assumes that a firm is operating in a perfectly competitive environment. Firms seek to maximise profits by maximising revenue and minimising costs simultaneously. In the long run, the competitive equilibrium leads to all firms earning normal profits. In the short run it is possible for some firms to make abnormal profits. The existence of abnormal profits will attract other firms to enter the market, which drives down prices until all firms are making normal profits.

In the long run, classical microeconomic theory assumes that any deviation from the optimal position is due to statistical noise. Given the same inputs if the level of outputs produced differ between two firms this leads to the conclusion that some relevant inputs that are not equal for both firms have been neglected (Ray, 1988). Kumbhakar and Lovell (2000) notes that these arguments are strong and although firms attempt to optimize they do not always succeed. In response to this, techniques have been developed where the producer's behaviour is unchanged but success is not guaranteed. Within these techniques, the deviation from the maximum attainable profit may not just be due to noise; instead it may also be due to efficiency.

Koopman (1951) defined a producer as efficient if it maximises outputs given inputs. Efficiency is measured as the ratio of the observed output to the maximum attainable output given the level of inputs, or the ratio of observable inputs to the minimum attainable input level to produce a given output. The optimal level of inputs and outputs are defined by the production possibility technology. This section will expand on the definition of efficiency outlining the definition of Debreu (1951), Farrell (1957) and Koopman (1951) and outline the definition of production possibility technology.

Debreu (1951) and Farrell (1957) define the measure of technical efficiency known as Farrell measure as:

"*one minus the maximum equiproportionate reduction in all inputs that still allows the production of given outputs, a value of one indicates technical efficiency and a score less than unity indicates the severity of technical inefficiency.*"

Koopman (1951) define technical efficiency as:

"*a producer is technically efficient if an increase in an output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least on other input or a reduction in at least one output.*"

The Farrell measure of efficiency examines the radial contraction of inputs to the production possibility set or the radial expansion of outputs to the production possibility set. Koopman's (1951) definition is stricter, measuring the reduction of inputs to the production possibility set; this incorporates both radial and non-radial reduction of inputs or expansion of outputs. The difference between the measures will be extended upon in the examination of slacks in section 3.6.5.4. When examining the measurement of efficiency, the Farrell definition is utilised unless stated otherwise.

The measurement of efficiency compares the actual level of inputs or outputs relative to the optimal level indicated by the respective frontier. The frontier can be denoted by the production function for a one output multiple input case and can be extended to multiple outputs through the Production Possibility Set (PPS). A production function defines the relationship between inputs and outputs. For a given level of inputs, the production function defines the maximum attainable level of output that can be produced. Mas-Colell et al (1995) assume a firm has J inputs to produce a single output. The technical possibilities of the firm can be summarised using the production function:

$$q = f(x) \qquad (3.1)$$

Where q represents the output and x represents an J by 1 vector of inputs; $x = (x_1, x_2, \dots, x_J)'$ for this part of the analysis, we assume that the inputs are controllable by the firm. The properties of the production function outlined by Mas-Coelell et al (1995) will be discussed briefly as they underpin the economic analysis of measuring efficiency. The properties are:

- Non-negativity – The production function is finite, non-negative and is a real number.
- Weak Essentiality – The production of output is impossible without one input
- Nondecreasing in x – Additional units of inputs will not decrease the output produced. More formally if $x^0 \geq x^1$ then $f(x^0) \geq f(x^1)$
- Concave in x – A linear combination of x^0 and x^1 will produce an output that is no less than the same linear combination of $f(x^0)$ and $f(x^1)$. Formally $f(\theta x^0 + (1 - \theta)x^1) \geq \theta f(x^0) + (1 - \theta) f(x^1)$ where $0 \leq \theta \leq 1$. Concavity implies that all marginal products are non-increasing.

Note that these properties are not exhaustive, nor are they universally maintained. Mas-Colell et al (1995) assume that producers are on the frontier and therefore, the producer is efficient.

The production function examines the one output and multiple inputs case. This can be extended to multiple inputs and multiple outputs through a production possibility set (PPS). The PPS describes all input and output combinations that are technologically feasible (Varian, 1992).

The production possibility set can be outlined following Fare and Primont (1995). We can consider an input vector $x = \left(x_1, \dots, x_J\right)$ used to produce an output vector $y = (y_1, \dots, y_R)$ in a technology involving N production units. The technology set is defined as:

$$T = \{(x, y)\ |x\ can\ produce\ y\} \qquad (3.2)$$

The set contains all input-output vectors (x, y) such that x can produce q.

The production technology T can be represented as either an input or output set. The input set $L(y)$consists of all input vectors x that can produce a given output vector y.

$$L(y) = \{x|(x, y)\epsilon\ T\} \qquad (3.3)$$

Given the basic assumption of the production technology, the following properties of the input sets can be derived:

1. $L(y)$ is closed for all y;
2. $L(y)$ is convex for all y;
3. Inputs are said to be weakly disposable if $x\ \epsilon\ L(y)$ then, for all $\lambda \geq 1$, $\lambda x\ \epsilon\ L(y)$; and
4. Inputs are said to be strongly disposable if $x\ \epsilon\ L(y)$ and if $x^* \geq x$ then $x^* \epsilon\ L(y)$.

The output set $P(x)$, also known as the production possibility set, is the subset of all output vectors y that can be produced using the input vector x.

$$P(x) = \{y|(x,y)\epsilon\, T\} \qquad (3.4)$$

The properties of the output can be summarised for each x the output set $P(x)$ is assumed to satisfy:

1. $0\ \epsilon\ P(x)$; nothing can be produced from a given set of inputs (i.e., inaction is possible);
2. Non-zero output levels cannot be produced from zero level of inputs;
3. $P(x)$ satisfies strong disposability of outputs: if $q\ \epsilon\ P(x)$ and $q^* \leq q$ then $q^*\ \epsilon\ P(x)$;
4. $P(x)$ satisfies strong disposability of inputs: if q can be produced from x, then q can be produced from any $x^* \geq x$;
5. $P(x)$ is closed;
6. $P(x)$ is bounded; and
7. $P(x)$ is convex.

Given the production possibility set, Farrell (1957) measures input-orientation technical efficiency by examining the radial contraction of inputs given the amount of outputs whilst remaining within the feasible set. The Farrell input technical efficiency measure of DMU n is defined in equation (3.5) by Fare and Lovell (1978) and Fare et al (1985).

$$\theta_n = \min\{\theta|\theta x\ \epsilon\ L(y)\} \qquad (3.5)$$

θ is the proportional reduction of inputs whilst remaining within the input set, which is the measurement of Farrell input-orientation efficiency. Technical efficiency can be measured under output-orientation, which considers the radial expansion of outputs holding inputs constant whilst remaining within the feasible set. The Farrell output technical efficiency for DMU n can be defined in equation 3.6.

$$\phi_n = \max\{\phi|\phi y\ \epsilon\ P(x)\} \qquad (3.6)$$

ϕ is the proportional expansion of outputs whilst remaining within the output set, which is the measurement of Farrell output-orientation efficiency.

3.3 Economic Efficiency

Technical efficiency has been outlined, which examines the use of inputs and outputs relative to the production possibility set. However, producers have to make decisions given input and output prices. Economic efficiency can be measured through the utilisation of price data and the specification an economic objective. Economic efficiency can measure cost efficiency and revenue efficiency. Cost efficiency makes the assumption that firms aim to minimise costs given input prices. Cost efficiency is measured as the ratio of minimal feasible costs and actual costs.

Revenue efficiency assumes firms aim to maximise revenue given output prices. Revenue efficiency is measure as the ratio of maximum feasible revenue and actual revenue. Farrell (1957) proposed that economic efficiency can be divided into two components, the first being technical efficiency as discussed above while the second component is allocative efficiency. Allocative efficiency for input orientation examines the combination of inputs which produces a given quantity of outputs at the lowest cost. For output-orientation, allocative efficiency determines the combination of outputs to maximise revenue.

3.3.1 Cost Function

Economic efficiency under input-orientation can be measured through the estimation of a cost function, whereby firms aim to minimise costs. The cost minimisation problem of a firm can be written in equation 3.8 where $T(y, x)$ is the production technology.

$$c(w, y) = \min_{x} w'x \quad such\ that\ T(y, x) = 0 \qquad (3.8)$$

Here $w = (w_1, w_2, \dots, w_n)'$ is a vector of input prices. The problem searches over all technically feasible input-output combinations to find the input quantities to produce the output vector y at the lowest cost. The cost function satisfies the following properties (Coelli et al, (2005)

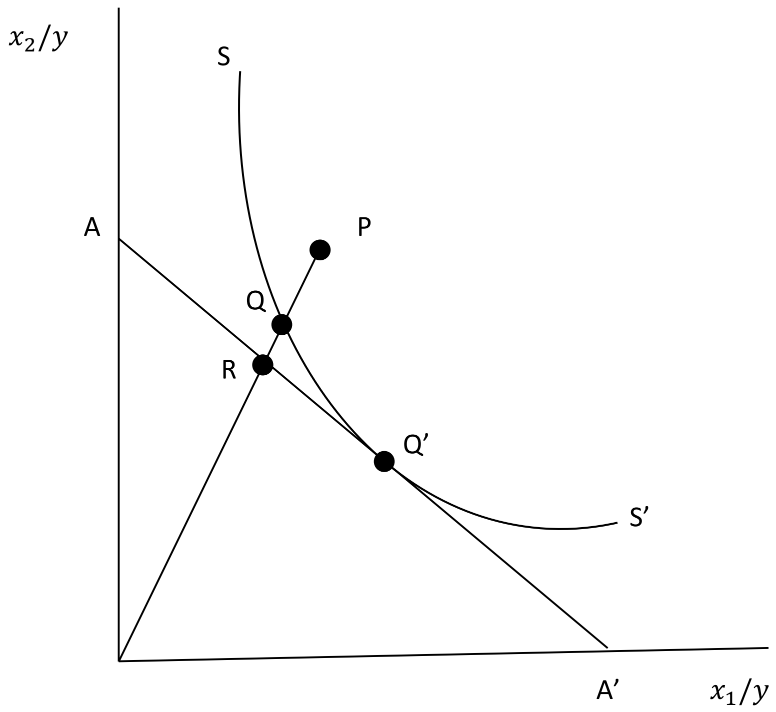
1. *Non-negativity*: Costs can never be negative
2. *Non-decreasing in w*: An increase in input prices will not decrease costs. If $w^0 \geq w^1$ then $c(w^0, y) \geq c(w^1, y)$
3. *Non-decreasing in q:* It costs more to produce more output. If $y^0 \geq y^1$ then $c(w, y^0) \geq c(w, y^1)$
4. *Homogeneity:* Multiplying all input prices by an amount $k > 0$ will lead to a k-fold increase in costs. $c(kw, y) = kc(w, y)$ for $k > 0$.
5. *Concave in w:* $c(\theta w^0 + (1-\theta)w^1, y) \geq \theta c(w^0, q) + (1-\theta)c(w^1, y)$ for all $0 \leq \theta \leq 1$.

Farrell (1957) demonstrate the cost minimisation problem by assuming that a firm has two inputs x_1 and x_2, produces one output y and assumes constant returns to scale (CRS), depicted in figure 3.1. If a firm exhibits CRS any proportional increase or decrease in inputs, the outputs will increase or decrease respectively in the same proportion. SS' shows the firm's isoquant, the optimal combination of inputs to produce a given amount of output which is assumed to be known. Farrell (1957) assumes that the firm uses inputs at point p; therefore it is technically inefficient as the firm could use fewer inputs to produce the same level of outputs.

Technical efficiency is measured by the amount in which inputs can be proportionally contracted and still produce the same level is output, measured by the distance QP. The technical efficiency score of the firm is calculated by ratio OQ/OP, which takes a value between 0 and 1. If a firm lies on the frontier it is technically efficient and therefore has an efficiency score of 1. Alongside the measure of technical efficiency, one can measure the extent in which firms use the correct combination of inputs given their input prices; allocative efficiency. Farrell (1957) notes that the line AA' shows the ratio of prices, and given the input prices the cost minimising point of production is Q'.

Allocative efficiency measures the degree of inefficiency from using the wrong combination of inputs, measured by the ratio OR/OQ. The allocative efficiency score also takes a value between 0 and 1, where a score of 1 indicates that the firm is utilising the input resources in the optimal combinations. Comparing the technical efficient level of inputs Q and the cost minimising level of inputs Q', to achieve allocative efficiency input x_1 should be increased and x_2 should be decreased. Therefore input x_1 is underutilised whereas x_2 is over-utilised. Cost efficiency (CE) is calculated as the ratio of actual costs and optimal costs given the input prices. This is measured as wx^*/wx, where w are the input prices, x is the actual inputs and x^* is the optimal inputs at point Q'. Alternatively, cost efficiency is calculated as the ratio OR/OP. The product of technical efficiency (TE) and allocative efficiency (AE) is equal to cost efficiency (CE), $CE = TE \times AE$. Farrell (1957) does highlight that the measure is not "entirely conclusive" as the level of technical efficiency may change as the combination of inputs alter.

Fig. 3.1: Input-Orientation



Source: Farrell (1957)

3.3.2 Revenue Function

Economic efficiency under output-orientation is measured through a revenue function whereby firms aim to maximise revenue given the level of inputs. The revenue maximisation problem can be written as:

$$r(p, x) = \max_{y} p'y \quad such\ that\ T(y, x) = 0 \qquad (3.9)$$

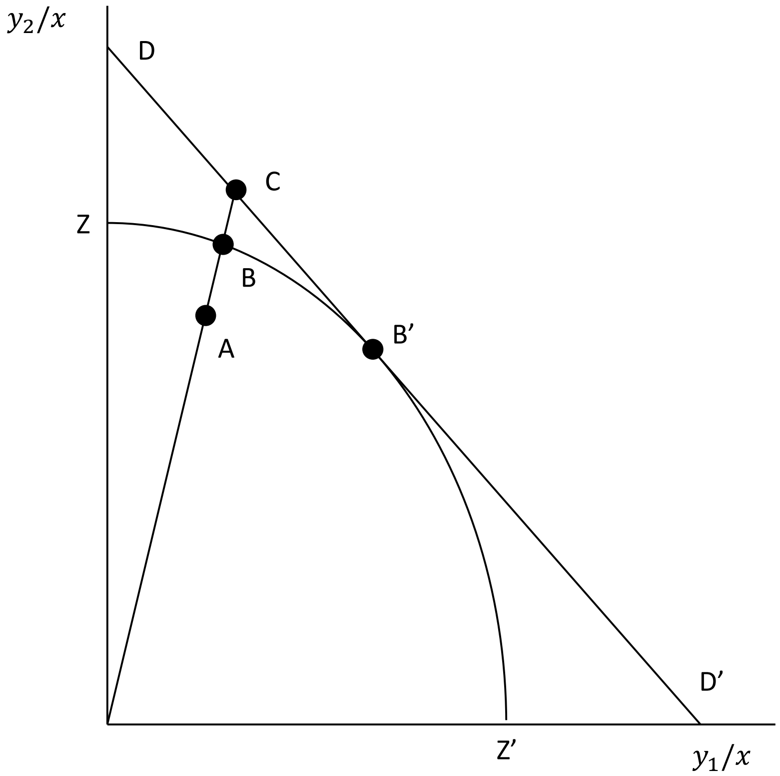
Where $p = (p_1, p_2, \dots, p_M)'$ is a vector of output prices over which the firm has no influence and $T(y, x)$ is the production technology. The revenue function satisfies the following properties:

1. *Non-negativity*: Revenue can never be negative

2. *Non-decreasing in* p: An increase in output prices will not decrease revenue. If $p^0 \geq p^1$ then $r(p^0, x) \geq r(p^1, x)$
3. *Non-decreasing in x:* If $x^0 \geq x^1$ then $(p, x^0) \geq r(p, x^1)$: An increase in inputs producers more output.
4. *Homogeneity:* Multiplying all output prices by an amount $k > 0$ will lead to a k-fold increase in revenue. $r(kp, x) = kr(p, x)$ for $k > 0$.
5. *Convex in* p*:* $r(\theta p^0 + (1 - \theta)p^1, x) \geq \theta r(p^0, x) + (1 - \theta)r(p^1, x)$ for all $0 \leq \theta \leq 1$.

Farrell (1957) demonstrates the measurement of efficiency applying a revenue function in figure 3.2 if we assume that a firm is producing two outputs y_1 and y_2 from a single input x. Assuming again for completeness that the firm has constant returns to scale, we can depict the technology by a single production possibility curve (ZZ') which shows the optimal combination of outputs given inputs. If the firm produces at point A it is technically inefficient, as it could produce additional outputs given the level of inputs. The technical efficiency of the firm A is measured by OA/OB; this is the radial expansion of outputs, given inputs to the production possibility curve. DD' shows the output price ratio, given inputs the firm maximises revenue at B'. Allocative efficiency is the degree of inefficiency from producing the wrong combination of outputs, measured by OB/OC. Comparing the technical efficient level of outputs B and the revenue maximising level of output B', to obtain allocative efficiency the output level y_1 should be increased and y_2 should be decreased. Finally, we can calculate the revenue efficiency (RE) as the ratio of actual revenue and maximum revenue py^*/py where p is the price of outputs, y is the actual outputs and $y^*$is the optimal outputs at point B', or measured by the ratio OA/OC. Revenue efficiency can be decomposed into technical and allocative efficiency $RE = TE \times AE$.

Fig. 3.2: Output-Orientation

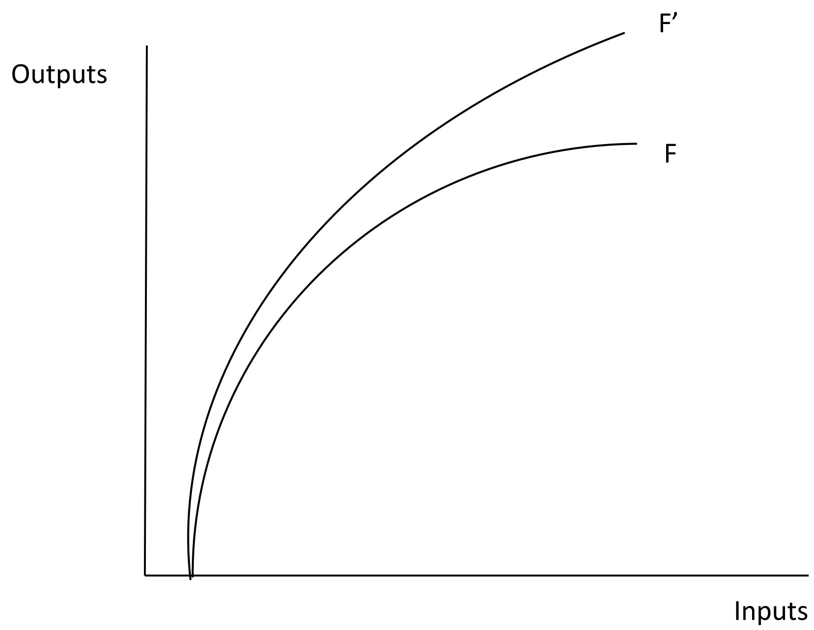


Source: Coelli et al (2005)

3.4 Technical Change

Coelli et al (2005) note that technology changes within an industry can be considered when efficiency is measured over time. This can be shown through the movement of the production function. When technology within the industry is increased, the industry-wide production function will be shifted outwards this is shown in figure 3.3. Alternatively, the industry can experience technological regression whereby the frontier shifts inwards. If panel data is used, which measures variation across time and firms, efficiency changes can be broken down into individual firm changes and also changes in technology within the industry which is explored in section 3.6.5.8.

Fig. 3.3: Technology Shift



Source: Coelli et al (2005)

3.5 Efficiency and Productivity

The performance of a firm can be measured by the concepts of efficiency and productivity. As previously defined, the efficiency of a producer is calculated as the ratio of observed output and the maximum potential output for the given inputs, known as output-orientation. On the other hand efficiency can be examined under input-orientation which examines the ratio of the observed inputs and the minimum potential inputs to produce a given amount of outputs (Fried et al, 2008). The optimal values are determined either through the production possibility set, by measuring technical efficiency or through an economic (cost, revenue or profit) function, measuring economic efficiency.

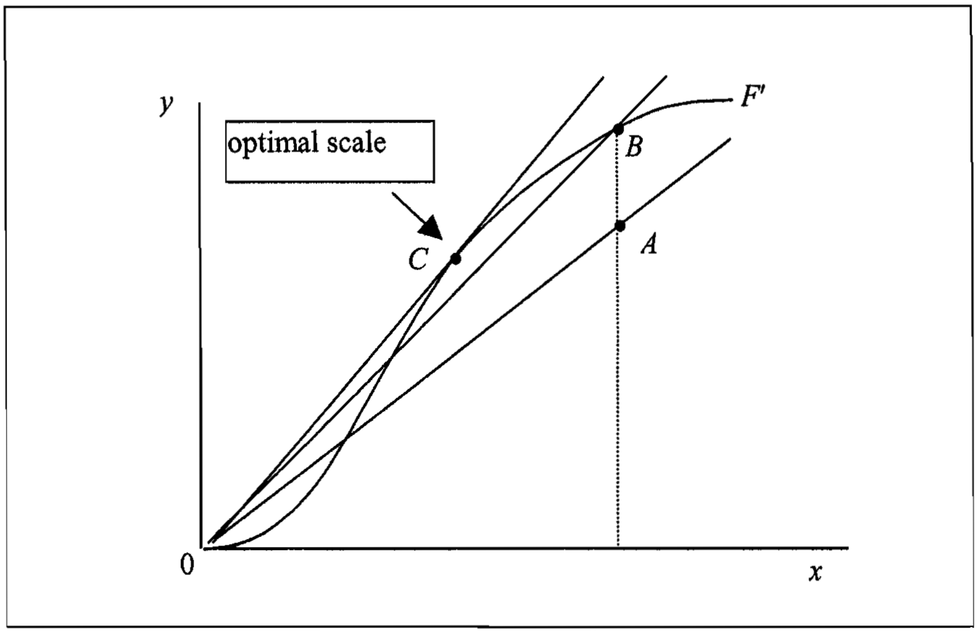
Productivity is calculated as the ratio of inputs and outputs produced by a firm:

$$Productivity = Outputs/Inputs \qquad (3.10)$$

The higher the value of the productivity measure, the more productive a firm is; the firm is more productive in converting inputs into outputs. A firm can increase the measure of productivity through either decreasing the number of inputs for a given number of outputs or increasing the level of output given inputs. When a firm uses numerous inputs or produces numerous outputs, aggregation techniques have to be employed to measure productivity. Productivity can be calculated including all factor inputs, denoted as total factor productivity or alternatively calculations of productivity may only take into consideration some of the factors of production, denoted as partial factor productivity e.g. labour productivity. The difference between efficiency and productivity is the former examines the ratio relative to the optimal amount.

Coelli et al (2005) note that the terms productivity and efficiency are usually used interchangeably. In order to distinguish the difference between them they consider a simple production process where there is one output and one input. In figure 3.4, FF' is the production frontier, Coelli et al (2005) consider points B and C which are both technically efficient points. If we draw a ray from the origin going through both points we can calculate the productivity ratio. The ray will show the ratio y/x or outputs/inputs; if the firm moves from point B to point C its productivity is increased. The ray going through point C is tangential to the production function and this is the maximum possible productivity. The movement from point B to C exploits scale economies, which is the increased efficiency due to firms producing at their optimal size. This shows that a firm can be technically efficient but not necessarily at the maximum obtainable productivity level. This section has outlined two concepts for the measurement of the performance of a firm. Efficiency is the main concept to be examined within this study.

Figure 3.4: Productivity vs. Efficiency



3.6 The Measurement of Efficiency

The measurement of efficiency involves the comparison of the firm's actual performance with the optimal performance located on the relevant frontier. Within most industries the true frontier is unknown; therefore an empirical estimate is required, known as the best practice frontier (Fried et al, 2008). The following section will outline the several techniques to measure efficiency. The techniques can be characterised as parametric and non-parametric. DEA is representative of a non-parametric frontier and a cost function and SFA are parametric approaches. This chapter will outline the concepts of a cost function, SFA and DEA, for the measurement of efficiency. The thesis employs three extensions of DEA to measure convergence, the inclusion of environmental variables and dynamic DEA. The chapter will therefore provide a detailed outline of the basic DEA model.

3.6.1 Parametric deterministic frontier

The production function or isoquant has been assumed to be known. In some engineering or physical production processes it is possible to determine the exact functional form of the production function. However, in most industries the exact production function is unknown. An approximation of the production or cost function is specified through an algebraic (functional) form between the dependent and explanatory variables. Through the estimation of the functional form, efficiency can be analysed through the deviation of actual costs from the optimal costs.

Cobb-Douglas Function Form

The Cobb-Douglas functional form of the production function is widely used to represent the relationship of inputs and outputs, developed by Cobb and Douglas (1928). The Cobb-Douglas form is:

$$TC = \alpha_0 \prod_{r=1}^{R} Y_r^{\alpha_r} \prod_{j=1}^{J} W_j^{\beta_j} \qquad (3.11)$$

Taking logarithms

$$\ln TC = \ln \alpha_0 + \sum_{r=1}^{R} \alpha_r \ln Y_r + \sum_{j=1}^{J} \beta_j \ln W_j \qquad (3.12)$$

Where, TC is total costs, Y_r is the *rth* output, W_j is the price of the *jth* input and α_0, $\alpha_i$and β_j are parameters to be estimated. The Cobb-Douglas cost function is homogeneous of degree one in input prices if $\sum_{j=1}^{J} \beta_j = 1$. Linear homogeneity implies a proportional increase of all input prices and results in the same proportional increase in costs. Linear homogeneity restriction is imposed within the estimation of a cost function. The restriction of the Cobb-Douglas cost function is that the first order approximation exhibits constant elasticity of scale. To overcome this disadvantage, more flexible functional forms have been developed, such as the translog functional form.

Translog Functional Form

Christensen et al (1973) developed the translog (transcendental logarithmic) production function, which is a second-order Taylor expansion as a local approximation of the "true" underlying production function. The translog cost function is specified as:

$$\ln TC = \delta + \sum_j \alpha_j \ln W_j + \sum_r \chi_r \ln Y_r + \frac{1}{2}\sum_j \sum_v \gamma_{\mathrm{jv}} \ln W_j \ln W_v + \sum_j \sum_r \kappa_{jr} \ln W_j \ln Y_r + \frac{1}{2}\sum_r \sum_z \xi_{r,z} \ln Y_r \ln Y_z \qquad (3.13)$$

Where TC is total costs, $Y_r$is the *rth* output, W_j is the price of the *jth* input. To impose linear homogeneity in input prices, the following restrictions are imposed on the parameters:

$$\sum_j \alpha_j = 1\,, \quad \sum_j \gamma_{j,v} = 0\,, \sum_j \kappa_{j,y} = 0 \qquad (3.14)$$

In addition, symmetry restrictions are imposed on the second-order parameters:

$$\xi_{r,z} = \xi_{z,r},\ \gamma_{\mathrm{jv}} = \gamma_{\mathrm{vj}},\ \kappa_{jr} = \kappa_{rj} \qquad (3.15)$$

The tranlog cost function allows for variable returns to scale and the estimation of the typical U-shaped average cost curve. The translog cost function has been applied within the empirical literature within the English and Welsh water and sewerage industry by Saal and Parker (2000), Saal and Reid (2004) and Bottasso and Conti (2009) to examine efficiency and the presence of economies of scale.

Quadratic Cost Function

The translog cost function does not easily allow for the incorporation of zero values. To allow the researcher to incorporate zero values for sewerage for WoCs Saal et al (2011) employ a quadratic cost function which is specified as:

$$TC_n = \alpha_0 + \sum_{r=1}^{R} \beta_n \ Y_r + \sum_{j=1}^{J} \gamma_j \ W_j + \frac{1}{2}\sum_{r=1}^{R}\sum_{m=1}^{R} \rho_r \ Y_r Y_m + \frac{1}{2}\sum_{j=1}^{J}\sum_{l=1}^{J} \lambda_{jl} W_j W_l + \sum_{r=1}^{R}\sum_{j=1}^{J} \theta_{rj} Y_r W_j \qquad (3.16)$$

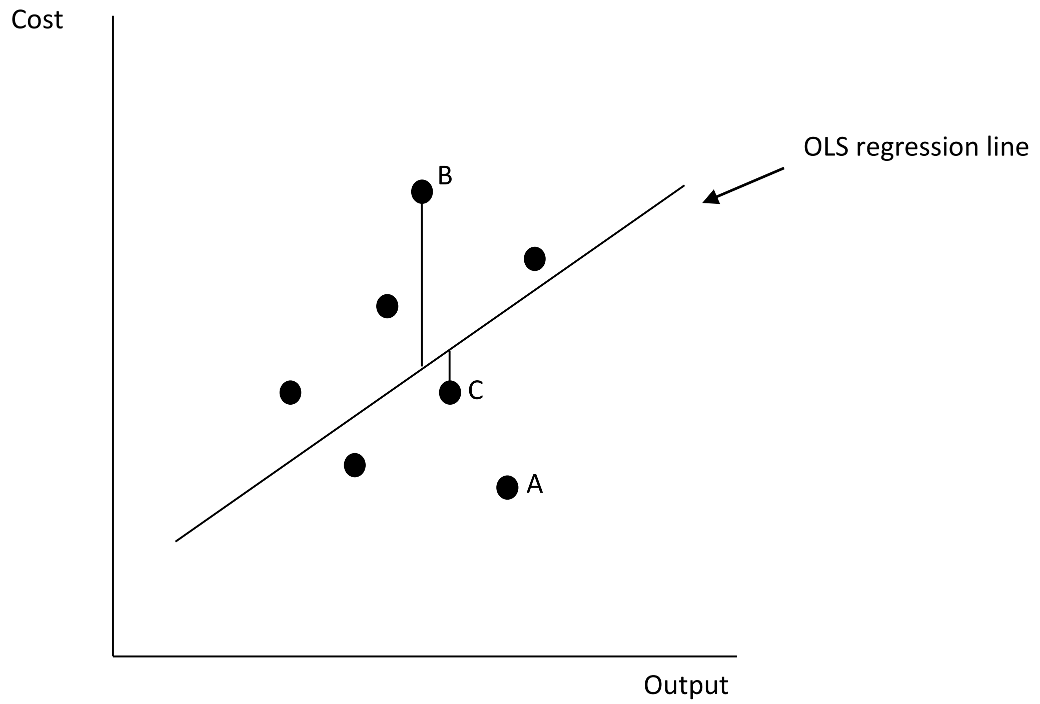
Where TC is the total cost, Y_r is the quantity of output r and W_j is the price of input j. Other functional forms have been applied within the measurement of costs functions. Diewert (1971) proposed the generalised Leontief cost function.

The functional form can be estimated using Ordinary Least Squares (OLS), Seemingly Unrelated Regression (SUR) or Stochastic Frontier Analysis (SFA).

3.6.2 Concept of Regression Analysis for Measuring Efficiency

The functional forms for the cost functions can be estimated through OLS or Seemingly Unrelated Regression (SUR). The first technique considered for measuring efficiency is OLS. OLS is applied by Ofwat within their measurement of efficiency through a series of linear regressions of explanatory variables on costs. OLS is a parametric technique which aims to measure efficiency through the estimation of the average cost functions through a linear regression. In respect to efficiency analysis, OLS aims to identify the relationships between the industry's costs and their associated cost drivers. As OLS measures an average cost function, this means that some companies will be more efficient whilst others less efficient than the average. The degree of efficiency is measured through the residuals, which show the difference between actual costs and estimated costs. The technique of using OLS for measuring efficiency and standard OLS differ in the assumption of the residuals. In traditional OLS it is assumed that the residuals are due to statistical noise, whereas in this case the residuals are assumed to be due to inefficiencies.

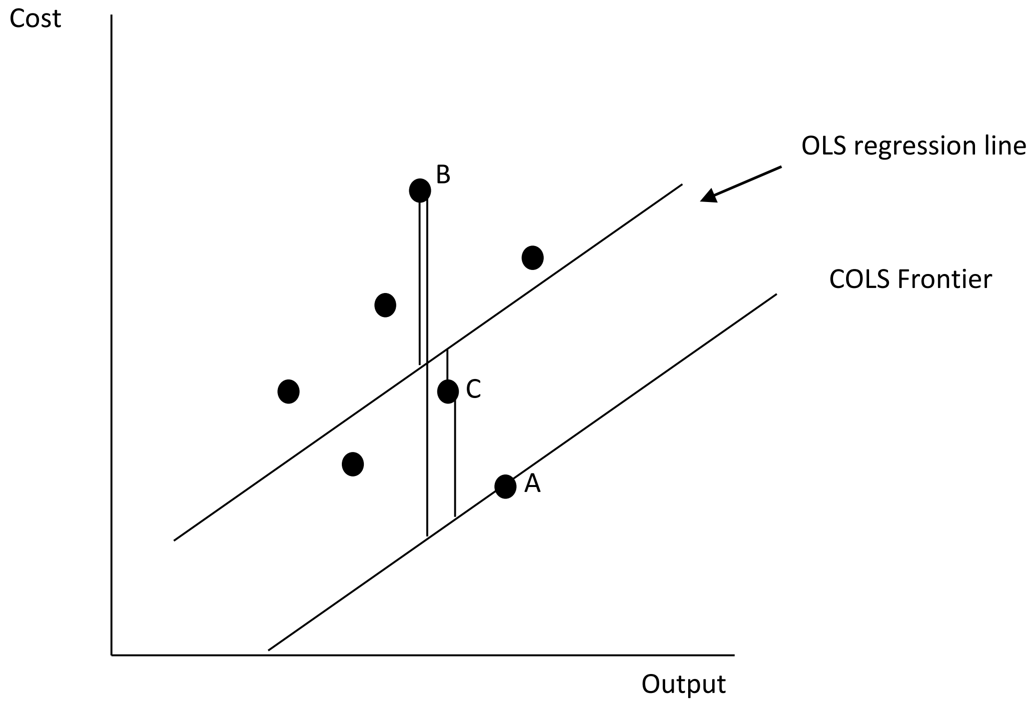
Fig. 3.5: OLS efficiency measurement



Source: Authors Illustration

In figure 3.5, the OLS regression line is the estimated costs function. Any firm that is above the regression line is classified as inefficient, as for a given output level their actual costs are higher than the estimated costs. A firm below the regression line is classified as efficient, as for a given level of output their actual costs are lower than the estimated costs. The level of efficiency is shown by the vertical line going up to the regression line. The standard benchmarking technique is to benchmark against the industry best, whereas OLS benchmarks against the industry average. A benchmark company within the industry can be established by employing Corrected Ordinary Least Squares (COLS) shown in figure 3.6. COLS shifts the regression line downwards until it intersects with the most efficient firm within the industry, which is assumed to be 100% efficient. In this case, firm A is the most efficient, therefore all firms that are not on the frontier are considered as inefficient.

Fig. 3.6: Correct Ordinary Least Squares



Source: Authors Illustration

If we compare the results from OLS and COLS, we can see that COLS always produces larger inefficiency scores. Kumbhakar and Lovell (2000) noted that the simplicity of COLS comes with problems, COLS is obtained by a parallel shift of the OLS regression. This implies that the structure of the best practice production technology is the same as the industry average. In practice the structure of the production technology is likely to differ between the best in the industry and the average, and therefore COLS does not bound the data as closely as possible since the parallel shift is required.

3.6.3 Measuring Cost Function Using Seemingly Unrelated Regression

The deterministic cost function is typically estimated alongside the cost share equations which are derived from the translog cost function. The estimation of both the translog cost function and cost share equation allows for additional information to be utilised. As the errors between the translog and costs share equations are correlated a more efficient estimation can be obtained using Zellner (1962) Seemingly Unrelated Regression (SUR). A translog cost function with one output (Y) and three inputs is shown in equation (3.17) where W is a 1 by 3 vector of input prices.

$$\ln TC = \alpha_1 + \beta_r \ln Y + \sum_{j=1}^{3} \gamma_j \ln W_j + \frac{1}{2}\rho_{rr} \ln Y \ln Y + \frac{1}{2}\sum_{j=1}^{3}\sum_{l=1}^{3} \lambda_{\mathrm{lv}} \ln W_j \ln W_l$$

$$+ \sum_{j=1}^{3} \theta_{jr} \ln W_j \ln Y \qquad (3.17)$$

The cost share equation is obtained by taking the derivative of the cost function with respect to input prices $\frac{\partial \ln C}{\partial \ln W_j} = \frac{\partial C}{\partial W_j}\frac{W_j}{C}$. Shephard's lemma states that the quantity demanded of the jth input is $X_j = \frac{\partial C}{\partial W_j}$, substituting the result $\frac{\partial \ln C}{\partial \ln W_j} = \frac{W_j X_j}{C} = S_j$ where X_j is the quantity of j used and S_j is the cost share for input j. The cost share equations follow:

$$s_1 = \frac{\partial \ln C}{\partial \ln W_1} = \gamma_1 + \sum\nolimits_{j=1}^{3} \lambda_{1j} \ln W_j + \theta_{r,1} \ln Y + \varepsilon_1$$

$$s_2 = \frac{\partial \ln C}{\partial \ln W_2} = \gamma_2 + \sum\nolimits_{j=1}^{3} \lambda_{2j} \ln W_j + \theta_{r,2} \ln Y + \varepsilon_2$$

$$s_3 = \frac{\partial \ln C}{\partial \ln W_3} = \gamma_3 + \sum\nolimits_{j=1}^{3} \lambda_{3j} \ln W_j + \theta_{r,3} \ln Y + \varepsilon_3 \qquad (3.18)$$

The cost shares sum to a value of 1; therefore one of the equations must be deleted to ensure linear independence and to avoid singularity of the error covariance matrix. The level of efficiency is the error term, although this has a two-sided distribution and a true measure of efficiency should have a one-sided distribution.

3.6.4 Stochastic Frontier Analysis

Stochastic frontier analysis (SFA) proposed by Aigner et al (1977), Battese and Corra (1977) and Meeusen and Van Den Broeck (1977) is a frontier approach which extends upon the estimation of a cost function to allow for the deviation from the frontier that is not under the control of the DMU, the incorporation of noise alongside the measurement of efficiency. SFA decomposes the error term into two sections: firstly the traditionally assumed symmetric random noise component, and secondly a one-sided inefficient component. The composite of the error terms therefore are skewed negatively for the case of the production function and will exhibit a negative mean. The SFA production frontier is written as:

$$\ln y_n = x_n'\beta + v_n - u_n \qquad (3.19)$$

Where y_n is a vector of outputs for firm $n = 1,..,N$, x_n is a vector of inputs and v_n is a two-sided error term component and u_n is a non-negative disturbance, which represents technical efficiency. If a DMU is technically efficient $(u_n) = 0$, the production function takes the form:

$$\ln y_n = x_n'\beta + v_i \qquad (3.20)$$

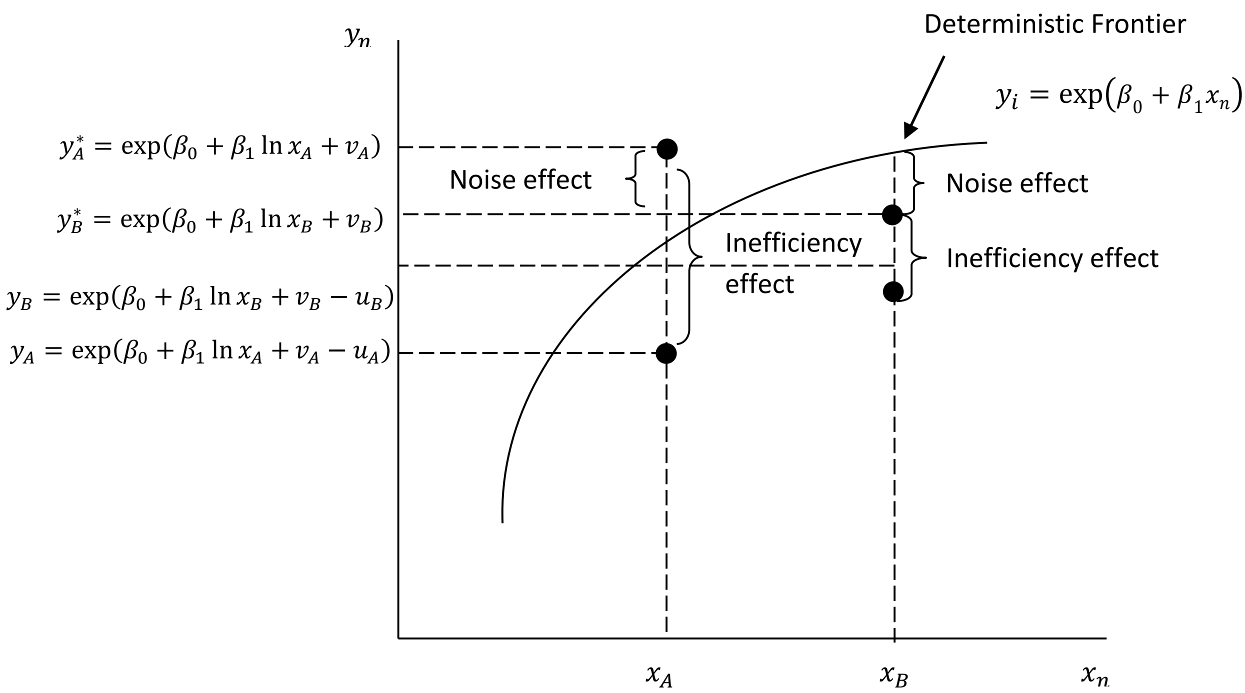
Coelli et al (2005) note that the frontier can be drawn in figure 3.7 where they plot the inputs and outputs of two firms and they note the production frontier exhibits diminishing returns to scale. Inputs are shown on the horizontal axis and outputs on the vertical. The graph shows that firm A uses the input level x_A to produce output y_A and firm B uses x_B to produce y_B.

$$y_A^* = \exp(\beta_0 + \beta_1 lnx_A + v_A) \quad and$$

$$y_B^* = \exp(\beta_0 + \beta_1 lnx_B + v_B) \qquad (3.21)$$

The frontier output in equation 3.21 for firm A, which excludes inefficiency lies above the deterministic production frontier due to a positive noise effect whereas firm B lies below due to a negative noise effect. The observed output of firms A and B; y_A and y_B lies below the frontier as the sum of the inefficiency term and the noise term is negative.

Figure 3.7: Stochastic Frontier Analysis



Source: Coelli et al (2005)

Coelli et al (2005) note that the unobserved frontier; which includes noise is evenly distributed above and below the observed deterministic frontier. The observed output tends to lie below the deterministic frontier due to the inefficiency component. Input-oriented technical efficiency for multiple inputs and outputs can be measured through an SFA input distance function introduced proposed by Lovell et al (1994).

Cost efficiency can be measured through the estimation of a cost function if input prices and total cost data are available. The cost frontier allows the extension of allowing for variable inputs and quazi-fixed inputs. Kumbhakar and Lovell (2000) note that the cost function requires behaviour assumptions. The SFA cost function takes the form:

$$\ln C_n = \ln C(y_n, w_n) + v_n + u_n \qquad (3.22)$$

Here, C_n is the observed total costs, y_n is a vector of outputs, w_n is an input price view, v_n is a two-sided noise component and u_n is a nonnegative disturbance which represents an individual firm's deviation from the efficient cost frontier; the firm's cost efficiency.

SFA is estimated by maximum likelihood (ML) and requires additional distributional assumptions for both of the error terms. The two-sided error term v_i is independent and identically normally distributed with mean 0 and variance σ_v^2, $v_n \sim iidN(0, \sigma_v^2)$. Several distributional assumptions have been outlined for the distribution of the one-sided managerial inefficiency error term: half-normal, truncated normal, exponential and gamma. The most common distributional assumption is half normal distribution proposed by Aigner et al (1977) and Mester (1993). The assumption of half normal distribution imposes the restriction that most firms are clustered near efficient and that higher degrees of inefficiency are increasingly unlikely (Berger, 1993). Stevenson (1980) and Greene (1980a,b) consider the normal-gamma distribution, which is a more flexible distribution but it can be difficult to separate the inefficiency from the random error.

3.6.5 Data Envelopment Analysis

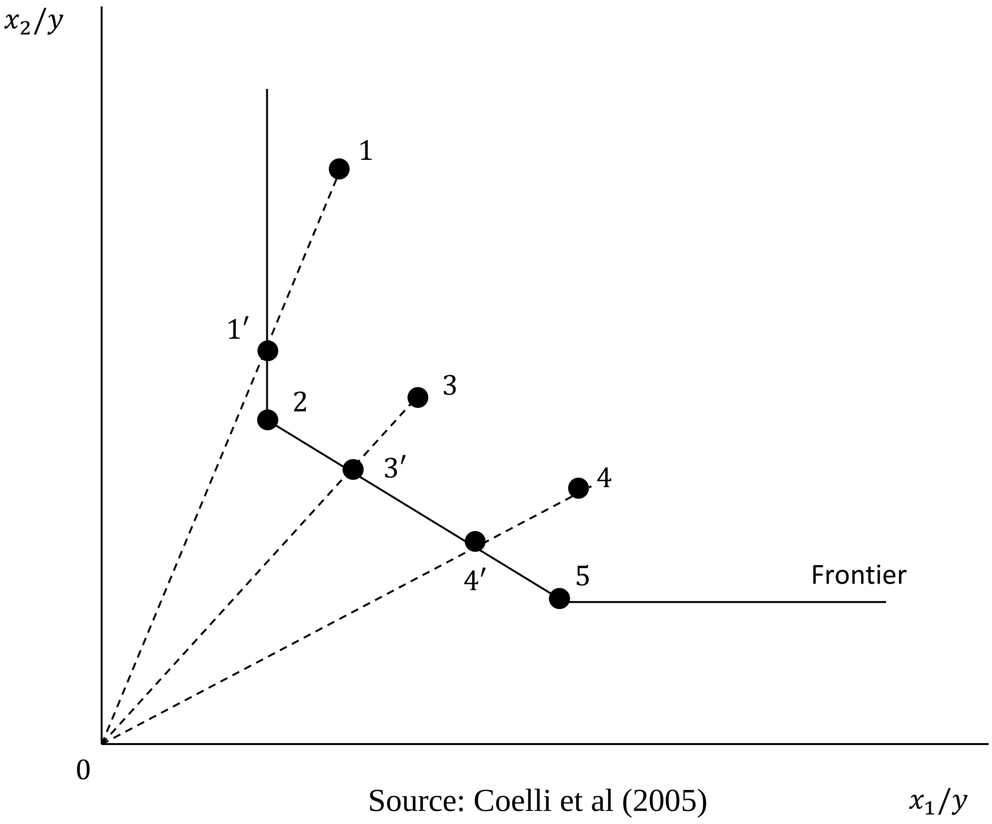
DEA is a non-parametric methodology which employs linear programming to construct a non-parametric piecewise surface (or frontier) over the data. DEA was developed by Charnes, Cooper and Rhodes (1978) (CCR). Efficiency is calculated as the distance relative of the DMU to the frontier. The level of efficiency is determined by a comparison to a single reference DMU or a convex combination of other referent DMUs located on the efficient frontier (Charnes et al, 1994). One of the advantages of DEA is it does not require any assumptions with regards to the functional form. Instead of specifying a functional form for the cost or production function, inputs and outputs are linked through a PPS in equation 3.2. DEA is advantageous by incorporating multiple inputs and outputs which are not required to be analogous (Cooper et al, 2006). Within the same model, inputs in monetary value can be incorporated alongside inputs in physical units.

On the other hand DEA has a series of limitations. Firstly, DEA measures relative efficiency not absolute efficiency, therefore efficiency is relative to the peers within the reference set, not a theoretical frontier. DEA assumes that the data is measurement free (Mester, 1996). DEA makes the implicit assumption of homogeneity (Golany and Roll, 1989). Homogeneity assumes that all firms undertake the same activities and operate in the same environment. Several extensions within the literature have been developed to overcome several of these limitations; the incorporation of environmental variables and bootstrapping techniques to incorporate measurement noise.

3.6.5.1 Input-Orientation

DEA can be measured using either an input or output oriented approach. The input-orientation approach as described by Farrell (1957) examines the proportional reduction of inputs given the amount of outputs whilst remaining within the feasible set. To illustrate input-orientation DEA, Coelli et al (2005) assume a case with two inputs and one output. They plot the input-output ratios for each input and plot those against each other, then the envelope can then be distinguished by creating a convex linear hull around the data. The data points can be enveloped within the region enclosed by the frontier with the horizontal line passing through 5 and the vertical line through 2. This is the piecewise linear production possibility set. In the case for input-orientation all points have to be above the frontier, except those firms that are efficient and lie on the frontier.

Fig. 3.8: Input-Orientation DEA



Source: Coelli et al (2005)

In this case, firm 2 and 5 are efficient as they are on the frontier. The remainder of the DMUs are inefficient as they could reduce their inputs and remain within the feasible set. If we consider firm 1, its efficiency score θ is calculated by the radial contraction of inputs to the frontier measured by the ratio $01'/01$, where the point 1' is the point that crosses the frontier when considering the ray from the origin to point 1. θ takes a value between 0 and 1, where 1 represents that the firm is efficient. The radial contraction of inputs to be efficient is given by θ.

DEA allows for the measurement of efficiency alongside the assessment of the target inputs or outputs and the determination of a firm's peer group. Charnes et al. (1994) state that the level of efficiency is determined by a comparison to a single reference DMU or a convex combination of other referent DMUs located on the efficient frontier, known as the peer group. The peer group is the set of points on the frontier in which the line connecting the origin and the point of consideration intersects. Therefore, for point 1 firm number 2 is its peer group, whereas the peer group for firm 3 is 2 and 5. This shows that for point 3 to be at its optimal efficiency, the firm's inputs weights should be a combination of the levels of inputs used by firm 2 and 5.

DEA uses mathematical linear programming to find the set of weights for each firm $\lambda_1, \lambda_2, \ldots, \lambda_n$ that maximises efficiency subject to constraints. Using Coelli et al (2005) notation, if we assume there are data on J inputs and R outputs for each N firm. For notation we say x_n and y_n are the input and output vectors respectively. The J^*N input matrix X and the R^*N output matrix Y represent all the data for the N firms. The problem involves finding the optimal weights to maximise the firm's efficiency subject to the constraint that the efficiency score must be less than or equal to one. Each firm is assigned a best set of weights and the values can vary from firm to firm.

The mathematical linear programming solution for input-orientation under CRS introduced by Charnes et al (1978) (CCR) is shown equation 3.23 which creates an envelope of the data.

$$min_{\theta,\lambda}\ \theta$$

$$st \qquad -y_n + Y\lambda \geq 0,$$

$$\theta x_n - X\lambda \geq 0,$$

$$\lambda \geq 0 \qquad\qquad (3.23)$$

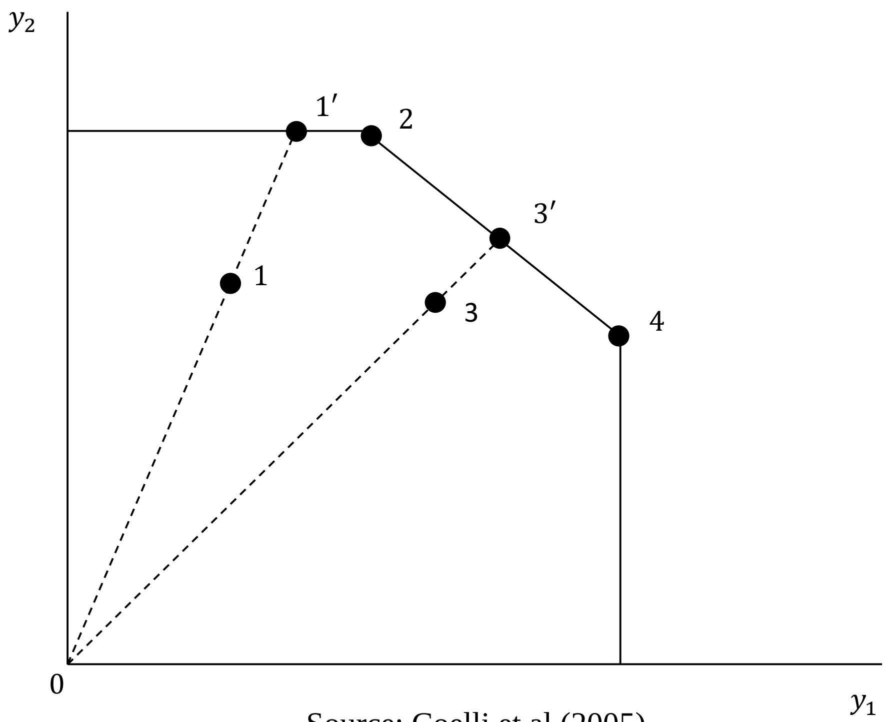
λ is an $N^{*}1$ is a vector of constants θ is a scalar which is the efficiency score for the nth firm. Farrell (1957) measure of efficiency θ takes a value between 0 and 1, where a value of 1 indicates a point on the frontier and indicates that the firm is technically efficient. The problem is solve N times until each firm has a value of θ. Coelli et al (2005) denote that in the problem above, the nth firms seeks to radially contract the input vector x whilst remaining in the feasible set. This is done by choosing optimal weights for each firm; we note that the weights may be different for each firm within the data. The feasible set here is determined by the observed decision making units within the sample and is a piecewise linear isoquant. The linear contraction assigns weights to inputs and outputs ($X\lambda, Q\lambda$), which is a projection point but remains within the feasible set (Coelli et al, 2005).

The optimal value of θ is independent of the units in which the inputs and outputs are measured provided that the units are the same for each firm. If we multiply each unit of input by a constant the solution obtained will be the same, known as the unit invariance property.

3.6.5.2 Output-Orientation

Output-orientation measures technical efficiency as proportional expansion of outputs to the frontier whilst holding input levels the same. Figure 3.9 depicts the production possibility for two outputs. Firms operating within the PPS are inefficient as the level of output is lower than the maximum attainable set. When using output-orientation, efficiency is calculated as the radial expansion in outputs to the frontier measured by the ratio $O1'/O1$.

Fig. 3.9: Output-Orientation DEA



Source: Coelli et al (2005)

Output-orientation DEA is calculated using the linear programme in figure 3.24 which maximises ϕ and λ.

$$max_{\phi\lambda}\ \phi$$

$$st \quad -\phi y_n + Y\lambda \geq 0$$

$$x_n - X\lambda \geq 0$$

$$\lambda \geq 0 \qquad (3.24)$$

Here $1 \leq \phi < \infty$ and ϕ is the proportional increase in outputs to achieve the highest level of efficiency for the nth firm when all inputs are held constant. Therefore, $1/\phi$ is the technical efficiency score which is bounded between 0 and 1.

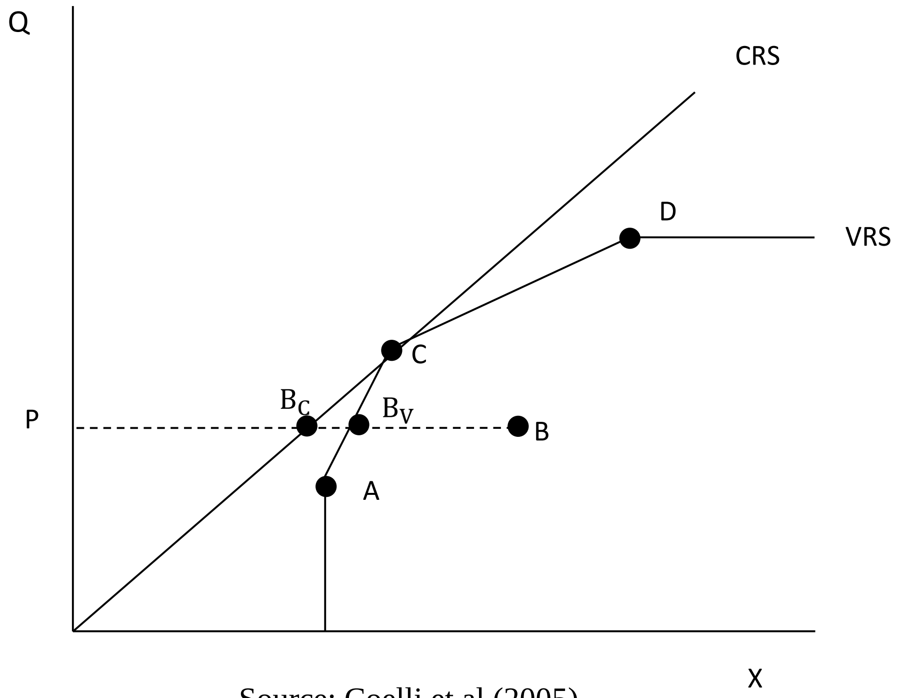
The decision of whether efficiency is examined under input or output-orientation depends on if input or outputs are easier to vary within the chosen industry. Input-orientation is assumed for the measurement of efficiency for the English and Welsh water and sewerage industry by Thanassoulis (2000a,b), Cubbin and Tzanidakis (1998) and Erbetta and Cave (2007) as the demand level faced by suppliers is exogenous, and therefore inputs are easier to vary.

3.6.5.3 Variable Returns to Scale

The model considered above is assumed to have Constant Returns to Scale (CRS), which assumes that firms are producing at their optimal level. A firm is operating under CRS if a proportional increase in inputs results in the same proportional increase in outputs. Increasing Returns to Scale (IRS) indicates that outputs increase by more than the proportional increase in inputs whereas Decreasing Returns to Scale (DRS) is the situation when outputs increase by proportionally less than the increase in inputs. Coelli et al (2005) highlight that CRS is only an appropriate assumption when all firms are operating at an optimal scale. However, it can be argued that some firms may not be working at their optimal level due the presence of imperfect competition, government regulations, constraints on finance etcetera and therefore may be subject to scale inefficiencies. The use of VRS does not assume that the firm is operating at its optimal size and therefore scale efficiencies are not taken into consideration when determining the level of technical efficiency. The measure of efficiency using CRS combines both technical and scale efficiencies, therefore the difference between the efficiency measured under VRS and CRS is attributed to scale efficiencies. Scale efficiency measures the impact of the scale size on productivity, either the firm is too large and is operating under DRS or the firm is too small and is operating under IRS. The use of VRS has been suggested by many authors including Afrait (1972), Fare et al (1983) and Banker et al (1984).

The use of VRS can be shown graphically in figure 3.10 when considering a firm that produces one output using one input. The graph shows both the CRS and the VRS frontiers. If we consider point B, under CRS the technical efficiency ratio of this firm is give by PB_c/PB, whereas under VRS the technical efficiency is PB_v/PB. Therefore the difference in these ratios is due to scale efficiencies which is PB_c/PB_v. As stated previously, the CRS ratio can be decomposed into technical efficiency and scale efficiency. Therefore $TE_{CRS} = TE_{VRS} \times SE$, where TE_{CRS} and TE_{VRS} denotes technical efficiency under CRS and VRS respectively, and SE denotes scale efficiency. SE takes a value between 0 and 1, where a value of 1 indicates that the DMU is scale efficient, whereby $TE_{CRS} = TE_{VRS}$.

Fig. 3.10: CRS and VRS



Source: Coelli et al (2005)

Banker, Charnes and Cooper (1984), known as BCC note that CRS linear programming can easily be transformed to allow for VRS by adding the constraint: $I1'\lambda = 1$ to equation 3.1 to produce the equation 3.25.

$$min_{\theta,\lambda}\ \theta$$

$$st \qquad -y_n + Y\lambda \geq 0$$

$$\theta x_n - X\lambda \geq 0$$

$$I1'\lambda = 1$$

$$\lambda \geq 0 \qquad (3.25)$$

Here, $I1$ is an $N \times 1$ vector of ones. Coelli et al (2005) note that this approach creates a convex envelope which is tighter than that produced by CRS. VRS produces technical efficiency scores that are greater or equal to those produced by CRS. The constraint $I1'\lambda = 1$ is the convexity constraint. The convexity constraint prevents any interpolation points constructed by the DMU being scaled up or down to form a reference point as scaling is not possible under VRS (Thanassoulis, 2001). The VRS assumption ensures that firms are only benchmarked against those that are of a similar size, where under CRS a firm may be benchmarked against a firm which is substantially smaller or larger. If the firm is smaller, then the λ weights are greater than 1, and if the firm is larger, then the λ weights are less than 1.

Tone (2001) provides of four sets of properties in which an efficiency measurement should satisfy:

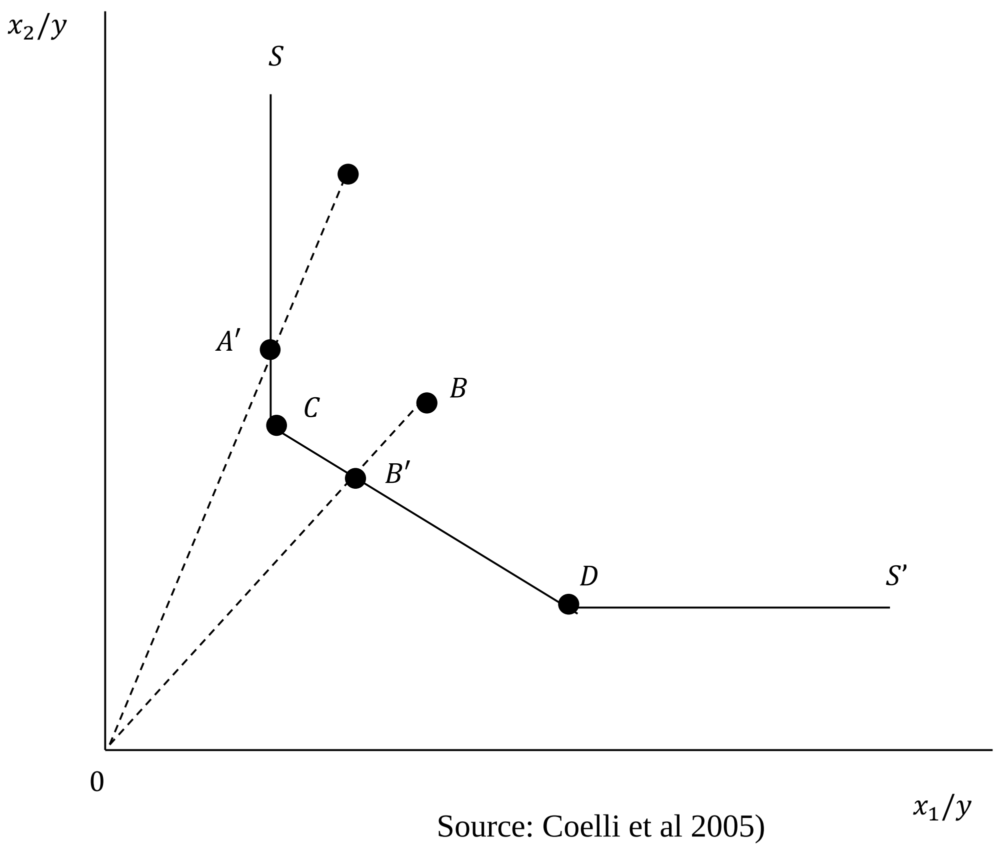
- *Unit Invariance:* The measure should be invariant with respect to the units of data
- *Monotone:* The measure should be monotone decreasing in each slack in input and output
- *Translation Invariant:* The measure should be invariant under parallel translation of the coordinate system applied (Ali and Sieford, 1990; Pastor, 1996)
- *Reference set invariant:* The measure should be determined only by consulting the reference set of the DMU concerned

The CCR and BCC DEA models satisfy the properties of being monotone and reference set invariance. The CCR and BCC radial measure of efficiency are unit invariant, therefore multiplying the inputs by a constant does not alter the efficiency score. However, any input slacks, which will be discussed next, are not unit invariant. The CCR and BCC model are not translation invariant, therefore adding a constant to an input will skew the efficiency scores.

3.6.5.4 Slacks

The Farrell (1957) measure of technical efficiency examines the radial contraction of inputs, whereas Koopman (1951) provides a more strict definition of technical efficiency whereby a firm is technically efficient if it operates on the frontier and exhibits zero slacks. The problem of slacks arises due to the parts of the frontier that run parallel to the axis. This can be shown in figure 3.11 obtained from Coelli et al (2005) where we have the piecewise linear function which is made up of point C and D. If we consider firm A, which has an efficiency ratio of $0A'/0A$. However it can be questioned whether point A' is an efficient point as the amount of input x_2 used can be reduced whilst the same level of output is being produced. This is known as an input slack. Output slacks can also occur when multiple inputs and multiple outputs are considered within the model.

Fig. 3.11: Input Slacks



Source: Coelli et al 2005)

Slacks can be incorporated within the radial measure of efficiency developed by Banker et al (1984) for input-orientation and variable returns to scale. In equation 3.26, s_i^- and s_r^+ represent the input and output slacks respectively and ε is a non-Archimedean small positive number. The model identifies the DMUs radial efficiency θ alongside the non-radial efficiency or slacks $\left(\sum_{j=1}^{J} s_i^- + \sum_{r=1}^{R} s_r^+\right)$.

$$Min\ \theta + \varepsilon\left(\sum_{j=1}^{J} s_i^- + \sum_{r=1}^{R} s_r^+\right)$$

$$st \quad \sum_{n=1}^{N} \lambda_n x_{jn} + s_n^- = \theta x_{n0}$$

$$\sum_{n=1}^{N} \lambda_n y_{rn} + s_r^+ = y_{r0}$$

$$\sum_{n=1}^{N} \lambda_n = 1$$

$$\lambda_n \geq 0; \quad s_n^- \geq 0; \quad s_r^+ \geq 0$$

$$j = 1, \dots, J \quad r = 1, \dots, R \quad n = 1, \dots, N \qquad (3.26)$$

One of the desirable properties is that the DEA model is unit invariant, therefore the efficiency scores are independent of the unit of measurement. However, Lovell and Pastor (1995) highlight that the Banker et al (1984) model is unit invariant with regards to the radial component although the slack component is not. The slack component therefore depends on the unit of measurement, multiplying the inputs by a fixed constant will change the slacks.

Charnes et al (1985) developed the additive model of DEA to incorporate both input and output slacks. The dual of the Charnes et al (1985) is shown in equation 3.27 where s_{j0}^{-} and s_{r0}^{+} represent the input and output slacks for the DMU under evaluation. A DMU is efficient if $s_{j0}^{-*} = s_{r0}^{+} = 0$, therefore the firm exhibits zero input and output slacks.

$$\max Z = \sum_{j=1}^{J} s_{j0}^{-} + \sum_{r=1}^{R} s_{r0}^{+}$$

$$st \quad \sum_{n=1}^{N} \lambda_n x_{jn} + s_{j0}^{-} = x_{j0}$$

$$\sum_{n=1}^{N} \lambda_n y_{rn} - s_{r0}^{+} = y_{r0}$$

$$\sum_{n=1}^{N} \lambda_n = 1$$

$$\lambda_n \geq 0$$

$$s_{i0}^{-}, s_{r0}^{+} \geq 0$$

$$j = 1, \dots, J \quad r = 1, \dots, R \quad n = 1, \dots, N \qquad (3.27)$$

The additive model determines the inefficiency for each input and output and therefore can discriminate between efficient and inefficient DMUs. However, it does not provide an efficiency score of θ which takes a value between zero and one (Tone, 2001). The CCR and BCC radial efficiency model is not translation invariant; however, the additive model has the desirable property that the results are translation invariant. Tone (2001) develops a slack based model, which examines the reduction of input and output slacks. The model provides an efficiency score for radial and non-radial efficiency which is unit invariant and monotone decreasing with respect to input excess and output shortfall. Within this thesis dynamic DEA measures cost efficiency. Radial efficiency is employed for the measurement of β-convergence and 3-stage DEA to incorporate environmental variables whilst not violating the unit invariance property.

3.6.5.5 Dimensionality

DEA is referred to as a deterministic technique for the measurement of efficiency. The efficiency score is measured relative to an estimate of the true (but unknown) production function. Estimates are obtained from a finite sample, therefore the measurement of efficiency are sensitive to sampling variation (Kneip et al, 1998). A well-known potential issue with regards to DEA is the 'curse of dimensionality' and the rate of convergence[5]. Banker (1993) shows consistency of the input efficiency estimator $\hat{\theta}_{VRS}$ with inputs $J = 1$ and outputs $R \geq 1$. Korostelev et al (1995) analysed the rate of convergence of the estimated frontier to the true frontier for the one input case. Kneip et al (1998) examined the rate of convergence for the multivariate case where J and R are greater than one:

$$\hat{\theta}_{VRS}(x,y) - \theta(x,y) = O_p\left(n^{-\frac{1}{J+R}}\right) \qquad (3.28)$$

Simar and Wilson (2008) note that the convergence rate for CRS has not been established yet, but imagine that its convergence rate would be faster if the production possibility set exhibits constant returns to scale. The rate of convergence depends on the dimensionality of the problem; the number of inputs and outputs. Simar and Wilson (2008) state that for a given sample size, as the number of inputs and outputs increase the number of DMUs lying on the

[5] Convergence here refers to the statistical properties of the DEA estimator and its rate of convergence to the true frontier.

frontier will increase. As a result of the well-known ‘curse of dimensionality’, several authors have stated rules of thumb for the number of DMUs required for a given number of inputs and outputs. Golany and Roll (1989) establish a rule of thumb that the number of units should be at least twice the number of inputs plus outputs. Bowlin (1998) establishes a rule of thumb of three times the number of inputs plus outputs. However, given these rules Wheelock and Wilson (2003) and Wilson (2004) have found cases for several thousand DMU and nearly all the DMUs lie on the frontier. The rule of thumb will be used as a guide alongside the number of DMUs which make up the frontier to determine whether there is a dimensionality issue.

3.6.5.6 Non-Discretionary Variables

WaSCs operate under different operating characteristics which are outside of managerial control, known as non-discretionary or environmental variables. DEA makes the implicit assumption that DMUs are homogeneous. Ray (1988) states that for a given level of inputs, conceptually there is a maximum amount of output that can be produced given the level of technology. Ray asks what causes a firm to produce less than the maximum. Within the analysis of efficiency, the difference between the actual output and maximum attainable output is attributed to managerial inefficiency. However, other factors may be present which influence the amount of outputs that can be produced which are not under control by the firms; these are defined as non-discretionary variables. If the true production function takes the form:

$$y_t = g(x_t, z_t) \qquad (3.29)$$

Here, y_t is the level of outputs, x_t the level of inputs and z_t is a vector of non-discretionary variables. If non-discretionary variables are not incorporated within the analysis then the production function is specified as:

$$y_t = p(x_t) \qquad (3.30)$$

Production function (3.30) implicitly makes the assumption that firms operate within homogeneous environments and that non-discretionary variables do not impact the production function. If two firms t and s have identical bundles of controllable inputs $(x_t = x_s)$ but produce different levels of outputs $(y_t \neq y_s)$. The difference in the output level is interpreted as differences in efficiency. However if the true production function is (3.29) then the level of outputs should not be equal unless the environmental variables are equal $(z_t = z_s)$. If $(z_t \neq z_s)$ and the true production function (3.29) is measured as (3.30) the efficiency scores will be biased depending upon the relationship of the environmental variables. The exclusion of non-discretionary environmental variables can make those firms that operate in a favourable environment appear relatively efficient whilst those operating in an unfavourable environment appear relatively inefficient.

Implicit within the measurement of efficiency, DEA assumes that DMUs are homogeneous (Golany and Roll, 1989). Homogeneity assumes that the DMUs undertake the same activities, have access to the same inputs and finally operate within similar environments. Theoretically, it is important to ensure that non-discretionary variables which influence the production function are incorporated within the analysis. As DEA assumes homogeneity, the following section outlines one approach to incorporate non-discretionary variables within the measurement of efficiency through DEA.

Quality can be incorporated within the measurement of efficiency by a quality-adjusted measure of output. Saal and Parker (2000) employ a quality-adjusted measure of output for the English and Welsh water and sewerage industry due the large changes in quality standards imposed since privatisation. A quality-adjusted measure of output (Y_Q) is calculated by multiplying the level of output (Y) by a quality index (Q) which takes a value between 0 and 1 as shown in equation 3.31.

$$Y_Q = Y \times Q \qquad (3.31)$$

The application of a quality-adjusted measure of output implicitly makes the assumption that an improvement in the measure of quality requires additional inputs. A quality-adjusted measure of output is applied within the measurement of β-convergence in chapter 5. Chapter 6 provides a more in-depth discussion of the inclusion of environmental variables for the measurement of efficiency. These advanced techniques are not employed within the measurement of β-convergence due to the issue of dimensionality[6].

3.6.5.7 Bootstrap DEA

DEA is a deterministic frontier, and therefore does not take into account noise, and all deviations from the frontier are attributed to inefficiency. DEA does not account for random noise such as measurement error. Instead, it assumes that there is no noise within the data, and given this Coelli et al (2005) state that the DEA frontier is biased downwards and therefore the efficiency scores are biased upwards. Simar and Wilson (1998,2000a) use bootstrapping to provide a statistical foundation for DEA models and to adjust for the non-parametric bias. Coelli et al (2005) denote bootstrapping in its simplest form, which involves generating thousands of ‘pseudo samples’ from the observed data set. ‘Pseudo estimates’ are then obtained from the ‘pseudo samples’. An empirical distribution is then derived from these pseudo estimates. The distribution is used as an approximation of the true underlying sampling

[6] The measurement of β-convergence requires separate frontier for each time period. The three-stage DEA employs a meta-frontier, pooling all DMUs for all time periods under a common frontier, improving the dimensionality issue.

distribution of the estimator. DEA bootstrapping can cause biases and inconsistency problems due to the one-sided nature of the inefficiency distribution. Simar and Wilson (2000a) provide a solution to this problem in which they propose estimating a bias-corrected, non-parametric kernel estimate of the density of the inefficiency score and the drawing the pseudo samples from this density. Coelli et al (2005) also state that if the data for a whole industry or population is available, and assumed to be "noise-free", then there is no point in using bootstrapping as the DEA frontier obtained must be the true frontier. They also note that bootstrapping is a good technique to analyse the sensitivity of DEA. The quality of the bootstrap approximation depends upon the number of bootstraps and the sample size, as both tend towards infinity the approximation becomes exact (Simar and Wilson, 2000b). Simar and Wilson (2000b) through Monte Carlo simulations report that the performance of bootstrap DEA for a one input one output VRS model is low in small samples[7]. Bootstrap bias adjusted DEA efficiency scores are not examined within this study to avoid biases generated by the small sample size.

3.6.5.8 Malmquist Productivity index

Efficiency can be examined over time through the application of a Malmquist productivity index. This technique can be employed using DEA and SFA techniques. The concept of the Malmquist total factor productivity (TFP) index is drawn from the work of Fare et al (1994). The Malmquist TFP index measures the productivity change over two periods by calculating the ratio of the distance from the frontier for the two times periods relative to the industry frontier. We can consider an input-orientation approach where period t technology is the

[7] This width of the estimated 95 percent confidence intervals decreases as the sample size increases, becoming small when the sample size equals 400

reference frontier and the Malmquist TFP index between period s and the base period t is denoted by:

$$m_0^t(y_s, x_s, y_t, x_t) = \frac{d_0^t(y_t, x_t)}{d_0^t(y_s, x_s)} \qquad (3.32)$$

We note that $d_0^t(x_s, y_s)$ is the distance period and s is away from period $t.$ If $t = s$ then the efficiency score obtained will be the same as using a one-period data set. Fare et al (1994) note that the TFP calculation can be separated into efficiency change and technical change denoted as:

$$Efficiency\ Change = \frac{d_0^t(y_t, x_t)}{d_0^s(y_s, x_s)} \qquad (3.33)$$

$$and$$

$$Technical\ Change = [\frac{d_0^s(y_t, x_t)}{d_0^t(y_t, x_t)} * \frac{d_0^t(y_s, x_s)}{d_0^t(y_s, x_s)}]^{\frac{1}{2}} \qquad (3.34)$$

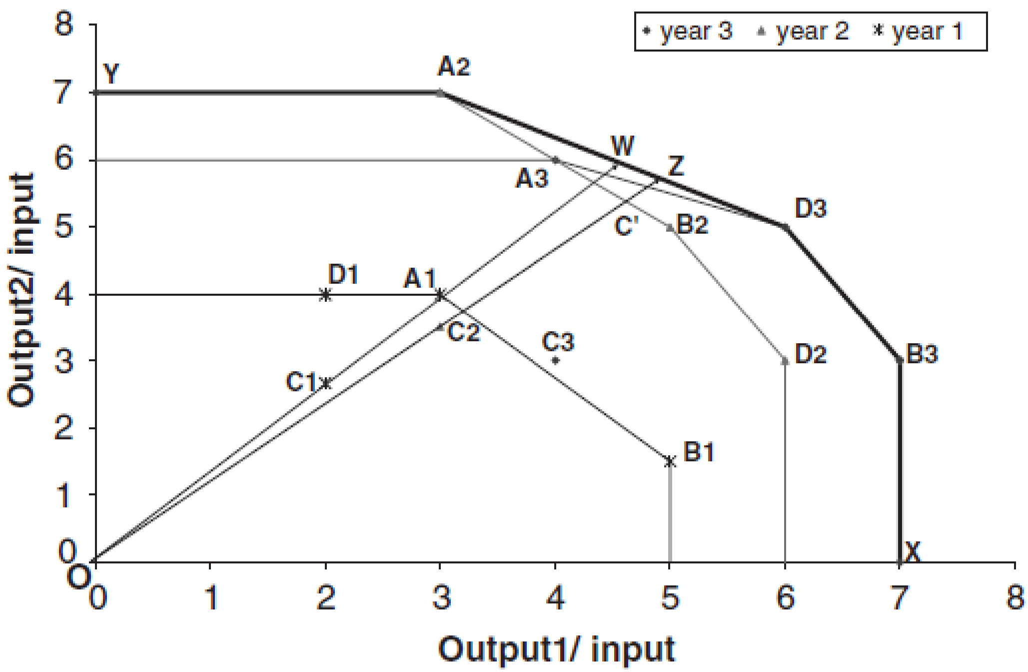
Efficiency change measures improvement in the efficiency score. Technical change measures the extent to which the firm catches up with new technology. Fare et al (1994) introduced a Malmquist index under VRS to incorporate scale economics, but Ray and Desli (1997) questioned the validity of the model to incorporate technical change alongside scale change. Scale change is calculated by comparing the CRS and VRS technology. Coelli et al (2005) highlight that the main criticism is that if there is scale change, then the true production technology must exhibit VRS.

3.6.5.9 Panel Data

In the presence of panel data there are several approaches that can be used to evaluate efficiency: separate frontier, meta-frontier or window analysis. Separate frontiers can be estimated for each time period and this allows for technical progress and regress. The evolution of the efficiency over time can be examined through the Malmquist index.

On the other hand, an alternative approach is to pool the data and estimate a Meta-frontier. This approach makes the assumption of unvarying best-practice technology. Figure 3.12 depicts three frontiers for different time periods. The points $A1, B1$ and $D1$ make up the frontier in period 1. The bold frontier is the meta frontier which envelopes all of the DMUs for all time periods. The measurement of efficiency for $C1$ against the frontier for period 1 is $\frac{OC1}{OA1}$, where against the meta-frontier the efficiency score is $\frac{OC1}{OW}$. It can be seen that under technological progress, where the frontier is expanding over periods, the measurement of efficiency in year 1 under the meta-frontier will be lower in comparison to the separate frontier. This approach is frequently utilised to overcome the issue of dimensionality but treats each firm for each time period as a separate DMU.

Figure 3.12: Meta-frontier



Source: Portela et al (2011)

An intermediate approach known as window analysis estimates a sequence of overlapping pooled panels which consist of a few time periods of arbitrary length. Figure 3.13 depicts window analysis for eight time periods. The length of the window is a pre-specified length and the window is rolled forwards for each period.

Figure 3.13: Window Analysis

	Time Period							
	1	2	3	4	5	6	7	8
Window 1	x	x	x	x				
Window 2		x	x	x	x			
Window 3			x	x	x	x		
Window 4				x	x	x	x	
Window 5					x	x	x	x

The advantage of this approach is to allow higher degrees of freedom whilst allowing for the technological process and regress. One of the drawbacks of the approach is that the beginning and end periods are not tested as frequently as the other periods.

3.7 Conclusion

This chapter has introduced the theory of efficiency and outlined several alternative approaches to measuring efficiency. The approaches can be divided into those which are parametric (OLS, SUR, SFA) and those which are non-parametric (DEA). SFA has the advantage of allowing the measurement of efficiency whilst incorporating noise within the estimation. On the other hand, DEA has the advantage of not requiring assumptions with regards to the functional form. The chapter goes on to provide a comprehensive review of DEA theory. Basic DEA models for input and output-orientation and constant and variable returns-to-scale considerations are outlined. Detailed methodology for the measurement of β-convergence, three-stage DEA and dynamic DEA is outlined within their respective chapters. The following chapter will outline the data used for the study and examine the industry specification for measuring efficiency within the English and Welsh water and sewerage industry.

4. 'Garbage in, garbage out': The data

4.1 Introduction

One of the key considerations within empirical economics is the choice of data. No matter how powerful the statistical technique tool, "garbage in equals garbage out" (Coelli,et al. 2005). The previous chapter outlined the theory of efficiency and the measurement of efficiency through the application of DEA. The choice of data employed for the measurement of efficiency is an essential component for the measurement of efficiency. This chapter outlines the data used for the measurement of efficiency within the English and Welsh water and sewerage industry.

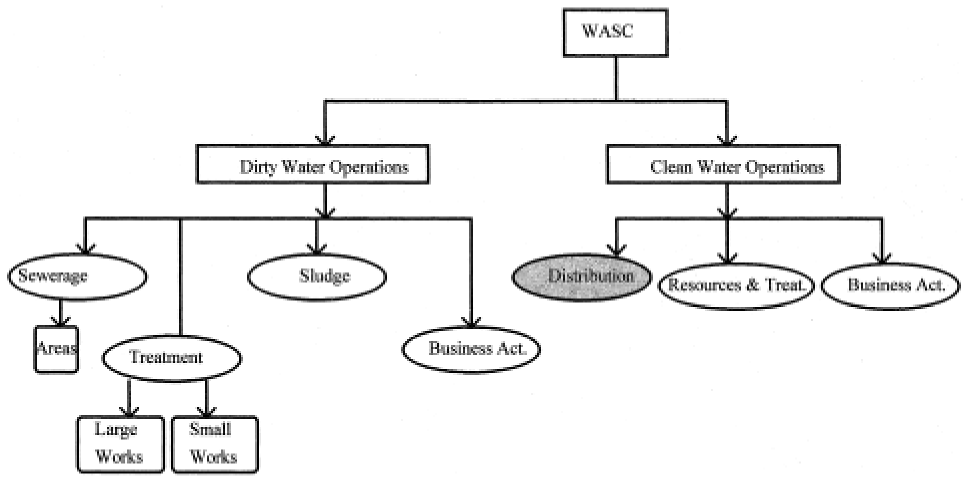
There are several key questions to guarantee an appropriate choice and reliability of the data. What variables should we use? Do the variables capture the production process? How reliable is the data? Are there measurement errors or outliers? This chapter addresses these questions, firstly by considering the choice of industry specification. The choice of input and output variables are discussed drawing from Ofwat's choice of variables and those applied within the academic literature. The data is obtained from published data from the industry June Return (JR), company's limited accounts and the DWI annual reports. Finally, the chapter will present a description of the data used within this study of efficiency.

4.2 Specification

The English and Welsh water and sewerage industry is characterised by two distinctive types of firms, WaSCs and WoCs which undertake two distinctive activities: water and sewerage. WaSCs undertake both water and sewerage activities whereas WoCs only undertake water activities. The breakdown of the WaSC's activities is displayed in figure 4.1. The figure depicts the two separate activities undertaken by WaSCs: water and sewerage and their function activities; distribution, resource and treatment and business activities for water services. Accounting separation allows for efficiency to be analysed at the functional level for water and sewerage. There are several different specification issues that need to be considered. Firstly, should water and sewerage activities be analysed together or separately? Secondly, if water and sewerage activities are modelled separately, one has to consider whether efficiency should be evaluated at the activity level or at the functional level of distribution, treatment etc. Finally, should WaSCs and WoCs be analysed under a common frontier? The following section will outline the approach undertaken by Ofwat to evaluate efficiency within the price review alongside a review of the methodologies applied within the academic literature to determine the preferred industry specification to model efficiency.

Within Ofwat's assessment of relative opex efficiency up until and including the 2009 price review, Ofwat examined water and sewerage activities separately, which allowed for WaSCs and WoCs to be evaluated under a common frontier for water activities. The modelling of water and sewerage activities separately does not allow for the incorporation of any cost interactions between water and sewerage activities. As water and sewerage activities are evaluated separately, efficiency can either be examined at the activity level or at the functional level. Ofwat examines efficiency at the functional activity level and sums the predicted and actual costs to the activity level to obtain a measure of efficiency for water and sewerage activities. This approach allows for an in-depth and detailed analysis of each functional activity, enabling the incorporation of detailed environmental variables such as the type of treatment work.

Figure 4.1: WaSC Activity Breakdown



Source: Thanassoulis (2000a)

Following Ofwat's specification, several authors have modelled water and sewerage activities separately. Cubbin and Tzanidakis (1998), Thanassoulis (2000a,b), Bottasso and Conti (2003), Saal and Parker (2005), Bottasso and Conti (2009) and Portela et al (2011) have examined water only activities and Thanassoulis (2002) has studied sewerage activities. Modelling water and sewerage activities separately allows for efficiency to be examined at either the activity or functional level. Thannassoulis (2000a,b) examine the efficiency of WaSCs at the functional level for distribution activities. This approach allows for relatively complex activities to be modelled using a variety of inputs, outputs and explanatory variables. Meanwhile Bottasso and Conti (2003), Saal and Parker (2005), Bottasso and Conti (2009) and Portela et al (2011) examine water activities at the activity level.

Saal et al (2011) estimate a four output quadratic cost function and state that it would be preferable to model each activity at its functional level, but notes that it is infeasible for two reasons. Firstly, there are difficulties with identifying meaningful outputs at each level and secondly even if the data was available they highlight that it would be economically intractable to estimate a total cost function for all functional activities. Saal et al (2011) use a four output model to estimate the presence of vertical economics and report significant vertical integration economies between network and abstraction and treatment activities for both water and sewerage activities. The results conclude that there are substantial cost savings for undertaking all components of the value chain for water activities and sewerage activities.

Stone and Webster (2004) through the estimation of a translog cost function reveal overall diseconomies of scope between four outputs: water delivered, equivalent population served and the number of water and sewerage properties. The measurement of vertical economies of scope between the physical volume of outputs and the connected properties reveal significant diseconomies of scope for sewerage and zero economies or diseconomies for water activities. There are significant economies of integration between water treatment and sewage treatment. The evaluation of efficiency at the functional level does not take into account the cost interactions between functional activities.

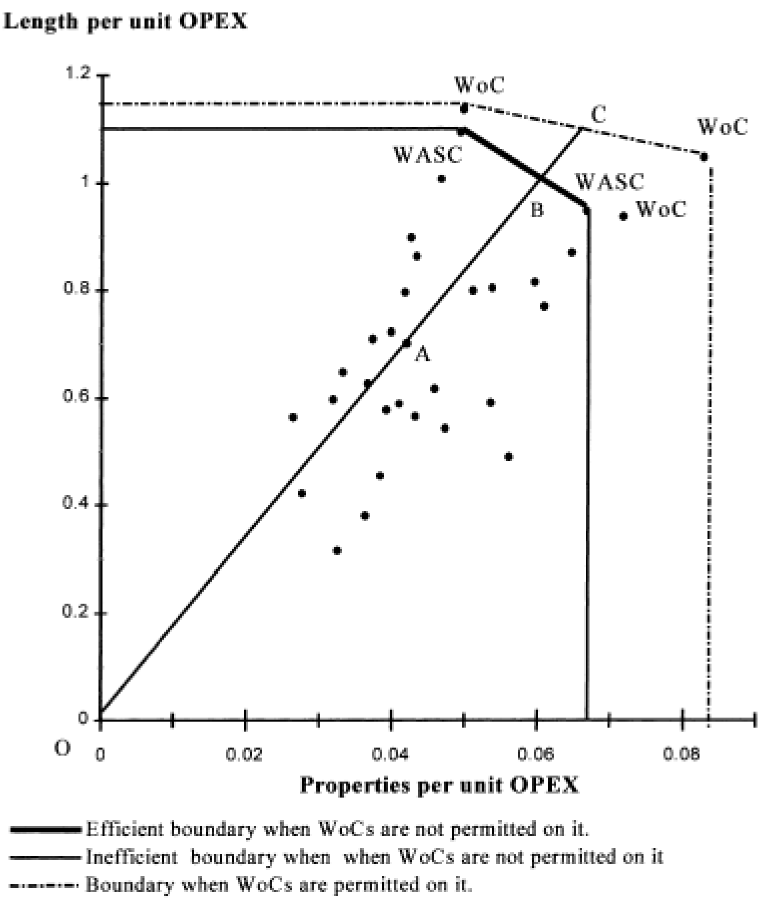
Water and sewerage activities can either be modelled separately following the methodology from Ofwat or can be modelled under a joint specification. A joint specification for water and sewerage activities for WaSCs allows for the incorporation of economies of scope. Saal and Parker (2000) estimate the separability of inputs and non-jointness through the estimation of a multiple output translog cost function for water and sewerage activities for the WaSCs. The test of non-jointness measures whether the cost of producing several outputs would be the same if they were produced jointly or separately. The results are not statistically significant and therefore imply neither economies nor diseconomies of scope. Alongside the measure of non-jointness, Saal and Parker (2000) test the assumption of the separability of inputs between water and sewerage activities. The test of separability determines whether it is appropriate to evaluate WaSCs costs using an aggregate measure of output and using a single output cost function, or whether WaSCs should be evaluated under a multiple output cost function. The test of separability is rejected, therefore implying that it is inappropriate to evaluate WaSCs costs without using a multiple output cost function. As well as Saal and Parker (2000), several other studies (Lynk, 1993; Hunt and Lynk, 1995; Stone and Webster, 2004) demonstrate the significant cost interactions between water and sewerage activities. Alongside the issue of incorporating cost interactions between water and sewerage activities, Saal and Parker (2005) highlight that there may be cost allocation problems in modelling water and sewerage activities separately.

Ofwat evaluates WaSCs and WoCs under a common frontier for water activities, this approach is followed by Cubbin and Tzanidakis (1998), Bottasso and Conti (2003), Saal and Parker (2005) and Portela et al (2011) who examine water only activities. For the analysis of a multiple output function for water and sewerage Ashton (2000a), Saal and Parker (2000), Saal and Reid (2004), Erbetta and Cave (2007), Saal et al (2007) and Maziotis et al (2013) only examine WaSCs. WaSCs have only been examined due to the difficulties of incorporating WoCs as zero producers of sewerage within the translog cost specification and within DEA. Bottasso and Conti (2011) and Saal et al (2011) are the only two papers that are known to measure both water and sewerage activities for WaSCs and WoCs. Bottasso and Conti (2011) and Saal et al (2011) overcome this problem by measuring a Box-Cox transformation and quadratic cost function respectively.

However, WaSCs and WoCs are two different types of firms and therefore may have access to different technology, raising the issue of whether WaSCs and WoCs can be pooled under a common frontier. Thanassoulis (2000a,b), Saal and Parker (2005) and Bottasso and Conti (2011) refer to the issue of the poolability of WaSCs and WoCs under a common frontier. Thanassoulis (2000a) measures efficiency at the functional level for water distribution and makes reference to the choice of companies which make up the DEA frontier. Thanassoulis

(2000a) notes that the WaSCs are substantially larger and account for some 75% of the water delivered in England and Wales. Pooling WaSCs and WoCs under a common frontier for the measurement of efficiency using DEA, Thanassoulis (2000a) reports that the WoCs make up the majority of the frontier. Thanassoulis (2000a) illustrates the problem of pooling WaSCs and WoCs in figure 4.2 which shows the two potential frontiers: firstly, the frontier including WoCs, and the secondly, the frontier obtained when excluding the WoCs. The graph shows that the inefficiency of a specific unit when including WoCs into the frontier is higher than when they are excluded. The outer frontier indicates that the production possibilities are higher for WoCs than WaSCs. Thanassoulis (2000a) states that one reason the difference arises is because WaSCs tend to be more complex companies in comparison to WoCs. Thanassoulis (2000a) calculates the correlation coefficient which suggests that once WoCs are removed from the frontier the unit cost is reduced, although the relative rankings remain unchanged. The paper takes ‘a conservative view’ and excludes WoCs from the calculation of the DEA frontier.

Figure 4.2: DEA Frontier WaSCs and WoCs



Source: Thanassoulis (2000)

Saal and Parker (2005) consider the implication of the poolability of WaSCs and WoCs for water only activities using an input distance function. Saal and Parker (2005) estimate the input distance function firstly by pooling WaSCs and WoCs and estimating the input distance function separately for WaSCs and WoCs. Within the pooled specification, a dummy variable is incorporated for WaSCs which indicates that the input requirements for a WaSC are substantially higher than those for a WoC. Saal and Paker (2005) demonstrate that the technology between the two types of firms differs by testing whether the coefficients are significantly different for WaSCs and WoCs. Their results conclude that it is inappropriate to assume that the underlying frontier for WoCs and WaSCs is the same. Saal and Parker (2005) highlight that only water activities are considered and therefore there could be cost allocation problems between water and sewerage activities as well as cost complimentarily issues through ignoring any cost interactions between water and sewerage activities. Overall, the paper highlights that the inappropriate assumption with regards to separability of water and sewerage activities may partly explain why WaSC and WoC frontiers differ from one another.

Portela et al (2011) highlights the issue of poolability by finding significant differences in productivity between WaSCs and WoCs. The paper reports that the average meta-efficiencies are significantly different for WaSCs and WoCs through the application of a Mann-Whitney test. Bottasso and Conti (2003) include dummy variables within their estimation of a cost function for water activities. Bottasso and Conti (2003) and Porela et al (2011) report that WaSCs are associated with cost advantages. On the other hand, as previously highlighted Saal and Parker (2005) report that the input requirements for WaSCs are substantially higher.

Bottasso et al (2011) examine the assumption of poolability and extend upon Saal and Paker (2005) by estimating a general cost function for both water and sewerage activities for WaSCs and WoCs. The model is a general specification of the composite cost function which applies a Box-Cox transformation. Bottasso et al (2011) conduct a likelihood ratio test with the null hypothesis that there is a common parameter vector for WaSCs and WoCs tested against the alternative that the parameters differ across the two sub-samples. Their results reject the null hypothesis and conclude that the preferred specification is to consider separate parameters for the two sub-samples.

Overall, given the discussion the measurement of efficiency within this thesis is examined at the activity level for both water and sewerage activities. This methodology allows for the incorporation of cost interactions between both water and sewerage activities and functional activities. WaSCs are only considered within the analysis, firstly as DEA does not easily allow for the incorporation of WoCs as a zero producer of sewerage, and secondly to avoid biases introduced by pooling WaSCs and WoCs under a common frontier.

4.3 Choice of Inputs and Outputs

This section will outline the process of determining the input and outputs used for measuring efficiency within the English and Welsh water and sewerage industry. The theoretical and methodological implications for the choice of input and output variables for measuring efficiency through DEA will be outlined. A review of the inputs and outputs used for the measurement of efficiency both by Ofwat and the academic literature is presented.

One of the key components within the measurement of efficiency is the choice of input and output variables. Firstly, the choice of variables should be common to all DMUs and represent all activities in which the firms undertake. The input variables should capture all resources utilised and the output variables should capture all outcomes that have a bearing on the type of efficiency being assessed (Coelli et al, 2005). There is a trade-off between the number of inputs and outputs included within the measurement of efficiency. The incorporation of additional variables allows for additional information to be contained within the analysis. However, as previously highlighted within the methodology the inclusion of additional input and output variables can lead to the issue of dimensionality. Dimensionality leads to a slower rate of convergence and leads to upwards bias in the efficiency scores. To avoid the curse of dimensionality, Dyson et al (2001) suggest a rule of thumb of two times the total number of inputs and outputs. Bowlin (1998) mentions the rule of thumb that the number of DMUs should be three times larger than the total number of input and output variables. To avoid the issue of dimensionality with measuring efficiency for the ten WaSCs a maximum of three inputs and outputs can be incorporated within the DEA specification. However, the number of DMUs can be increased by applying a meta-frontier or window analysis discussed in the methodology chapter.

The choice of input and output variables for the English and Welsh water and sewerage industry is led by the level at which the industry is examined, Ofwat's choice of variables within their efficiency analysis and the academic literature.

Ofwat examines efficiency at the functional level and examine opex and capex separately. The 2009 econometric models to measure water opex efficiency by Ofwat apply the following cost drivers: Distribution Input (DI); average pumping head; household billed water; non-household billed water; length of mains; proportion of DI from boreholes; proportion of DI from bulk suppliers. CEPA (2014) produce a series of totex models for the estimation of the baseline in 2014. Three output variables are incorporated: density (number of properties/length of mains), length of mains and usage (potable water/connected properties). The cost drivers applied by Ofwat for their assessment of sewerage opex in the 2009 price review include the area of sewage district, length of sewers, proportion non-residential population, household properties, non-household properties, load received from different bands of sewage works, equivalent population served and type of treatment. Within their assessment of totex expenditure CEPA (2014) use the length of sewers, total load and density as a cost driver.

It is important to note the distinction between Ofwat's methodologies and the measurement of a cost function. The former is used to predict the level of operating and capital expenditure going forwards to determine the price cap, whereas the latter is examining previous costs to determine the past level of efficiency. Several of the variables outlined can be considered as environmental variables instead of input and output variables.

The choice of input and output variables is influenced by the academic literature for the measurement of efficiency for water and sewerage companies from both English and Welsh studies and international studies. Appendix 1 outlines the input and output variables used within a selection of studies.

4.3.1 Inputs

The choice of inputs should represent all the resources utilised for the production of the output variables. The measurement of cost efficiency analyses the optimal use of physical inputs given the input prices and the production possibility set. Technical efficiency examines the radial contraction of inputs given the production possibility set. Technical efficiency DEA does not require the measurement of units of the inputs and outputs to be analogous. Within the same model, DEA allows for one input to be measured as the physical input, such as the number of employees, whereas another input can be incorporated in monetary terms (Cooper et al., 2006). Inputs can be incorporated as either the physical quantity (Garcia-Sanchez, 2006; Saal and Parker, 2005; Saal et al, 2007), the monetary value (Cubbin and Tzanidakis, 1998; Thanassoulis, 2000a; Thanassoulis, 2002; Tupper and Resende, 2004; Portela et al, 2011) or a combination of both the physical output and monetary value (Picazo-Tadeo et al, 2008) under technical inputs. This thesis considers the measurement of cost efficiency, and therefore inputs should relate to the total operating costs incurred by the firm.

In its assessment of efficiency, Ofwat modelled operating and capital expenditure separately for the measurement of efficiency within their 1994, 1999, 2004 and 2009 price reviews. Ofwat "has seen no convincing evidence that relatively high operating expenditure can be explained by relatively low capital expenditure or vice versa" (Ofwat, 1994, p.30). If operating and capital expenditure are incorporated within the same model then it is assumed that there is substitutability between the inputs. Thanassoulis (2000a), Cubbin and Tzanidakis (1998) and Portela et al (2011) follow Ofwat's methodology and examine operating expenditure for the measurement of efficiency within the English and Welsh water and sewerage industry. On the other hand, Saal and Paker (2000), Bottasso and Conti (2003, 2009), Saal and Parker (2005), Saal et al (2011), Maziotis et al (2012a, b), Erbetta and Cave (2007), Saal et al (2007), Saal and Reid (2004), Maziotis et al (2013) and Stone and Webster (2004) incorporate capital either within the measurement of efficiency through a total or variable cost function.

Capital can be included within the measurement of efficiency either as a fully controllable input or as a quasi-fixed input. A total cost function in equation 4.1 assumes that firms have the ability to adjust all inputs in the long term to their optimal level. The cost function can therefore be denoted as the decision of the firm n to minimise costs subject to this production function, where x_n is a vector of inputs, w_n a vector of input prices and $\bar{y}$ is the exogenous output level.

$$\min_{x_n \geq 0} \sum_n w_n x_n \qquad (4.1)$$

$$s.t \;\; f(\bar{y}, x) = 0$$

However, within the variable cost function in equation 4.2, capital is incorporated as a quasi-fixed input, therefore capital is not considered as a control variable and cost minimisation is only related to variable inputs. Capital x^K can therefore be incorporated as a quasi-fixed factor and x_j^v denotes the variable factors of production and firms solve the following cost minimisation problem:

$$\min_{x_n \geq 0} \sum_n w_n^v x_n^v \qquad (4.2)$$

$$s.t \;\; f(\bar{y}, x^v, x^K) = 0, \quad x^k = K$$

Capital has been incorporated as an input into the production of water and sewerage by Saal and Parker (2000) and Saal and Parker (2011). Saal and Parker (2000) state that a total cost function is applied to obtain an overall view of cost efficiency and to eliminate any potential biases introduced by cost allocation issues. However, Saal and Reid (2004) and Bottasso and Conti (2009) incorporate capital as a quasi-fixed input into the production process as technology within the industry is indivisible and associated with a long capital life and therefore is difficult to vary. Stone and Webster (2004) and Bottasso and Conti (2009) estimate a variable cost function for the English and Welsh water and sewerage industry and the coefficient on the quasi-fixed input implies a tendency of overcapitalisation. This is a common finding in the literature of public utilities (Caves et al, 1981 and Cowing and Holtman, 1983). Bottasso and Conti (2009) state that overcapitalisation can be interpreted as the Averch-Johnson effect due to the presence of rate of return regulation alongside the capital intensive nature of the industry and the presence of investment to meet future demand. Stone and Webster (2004) and Bottasso and Conti (2009) state that the presence of overcapitalisation could result in a misspecified total cost function where the assumption is that firms can instantaneously vary the level of capital.

In this study, capital costs are incorporated within the analysis of efficiency as controllable input to allow for the measurement of total cost efficiency. The incorporation of capital within the measurement of efficiency is extended by the measurement of dynamic efficiency in chapter 7.

The most common inputs within the measurement of efficiency within the English and Welsh water and sewerage industry are labour, capital and other (Saal and Parker, 2000; Saal and Parker, 2001; Stone and Webster, 2004; Erbetta and Cave, 2007; Saal et al, 2007; Bottasso et al, 2011; Maziotis et al, 2012a,b; Maziotis, 2013). Other inputs are the remainder of costs that are not associated with labour and capital costs. Other costs are therefore associated with power costs, materials, customer service, scientific services and other business activities. Saal et al (2011) include four inputs within their cost function, decomposing energy into a separate input. Ofwat reports energy expenditure however does not report a measure of energy used. Saal et al (2011) employ an energy price index from the department of energy and climate change, however to allow for varying energy prices, Saal et al (2011) employ separate price indices based on whether companies are small, medium or large. Energy costs have not been considered as a separate input due to the lack of data on energy prices and usage within the industry and the issue of dimensionality. The inputs considered are therefore capital, labour and other.

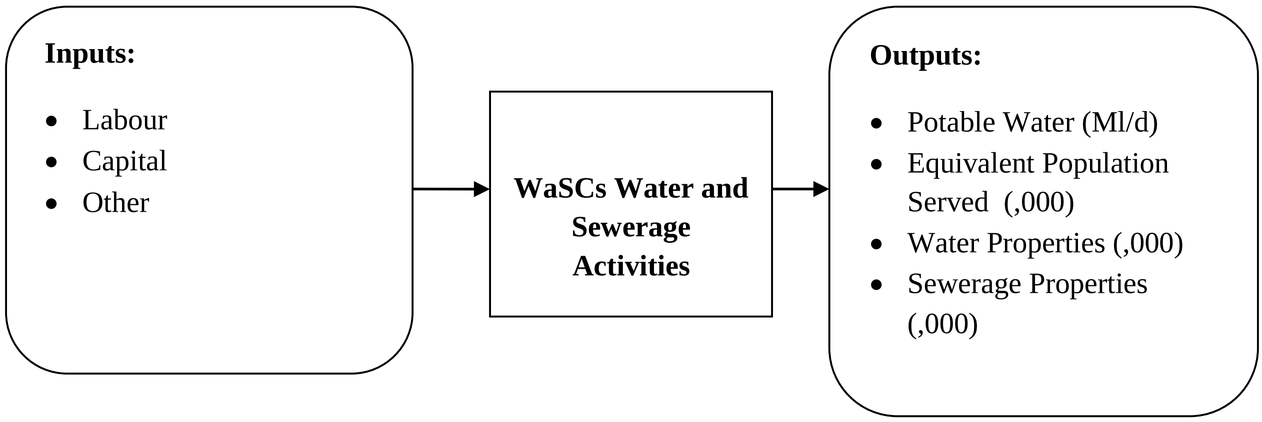
4.3.2 Output Variables

The choice of outputs used within the literature is partly determined upon which level of water and sewerage activities are examined. Cubbin and Tzandakis (1998) determine the output variables by regressing their input variable, opex on a set of potential output variables. Insignificant variables are dropped one by one until they arrive at a “statistically valid and economically meaningful parsimonious specification” (Cubbin and Tzandakis, 1998, pp. 82). The choice of outputs can also be determined based on economic and engineering judgement of the relevant outputs within the industry.

For the measurement of efficiency at the activity level of both water and sewerage activities, Stone and Webster (2004), Erbetta and Cave (2007), Saal et al (2007), Bottasso et al (2011) and Saal et al (2011) use four outputs, a physical measure of both water and sewerage and the number of properties billed for water and sewerage. Garcia and Thomas (2001) and Stone and Webster (2004) highlight the advantage of incorporating the number of billed properties alongside the measure of the physical amount produced. Saal et al (2011) estimate that the long run marginal cost of serving an additional water property is £67.86 and £96.81 for an additional sewerage property. They estimate that producing an additional cubic unit of water costs £0.41 and the long run marginal cost of £48.77 per equivalent person of sewage treatment. This result therefore indicates that serving an additional property is significantly different from the costs of an additional unit of the physical output. To allow for the difference in the input requirements for the number of properties served and the physical output of water and sewerage, both the number of properties and a measure of the physical output is included within the estimation of efficiency.

A summary of the selection of input and output variables is shown are figure 4.3. It should be highlighted that the data for the application of DEA for the measurement of β-Convergence, three-stage DEA and dynamic DEA do differ slightly due to technicalities explained within their relevant chapters. Inputs have not been assigned a unit of measurement in figure 4.3 as chapters 5 and 6 incorporate inputs in monetary terms whereas chapter 7 applies a cost function. The cost function incorporates the quantity of inputs given the inputs prices.

Figure 4.3: Choice of Inputs and Outputs



4.4 Data Description

The data collected is obtained from three sources; the June return published by Ofwat, the companies' limited company accounts and the Drinking Water Inspectorate (DWI) annual reports. The data is audited by an independent reporter for Ofwat; therefore the data is considered as reliable. This section will provide definitions and descriptions of the data utilised within the study. The data requirements vary between the chapters therefore a snap shot of the data will be provided within each empirical chapter.

A panel dataset is available for the periods 1996/97–2010/11 for the WaSCs. Prior to privatisation the industry consisted of ten state owned RWAs and 29 privately owned statutory WoCs. After the privatisation of the English and Welsh water and sewerage industry, the ten RWAs became publicly quoted WaSCs and the WoCs were re-established as public limited liability companies. After privatisation there was a succession of mergers and acquisitions of the WoCs; in 2011 eleven WOCs remained. Table 2.1 outlines the mergers and acquisitions within the industry whilst figure 4.4 shows the companies that are within the panel database. When there is a merger between a WaSC and a WoC, the WaSC remains within the panel due to the size difference between the companies[8].

[8] This is the case for Northumbrian and North East Water, Anglian and Hartlepool Water, Yorkshire and York Waterworks and Northumbrian and Essex & Suffolk.

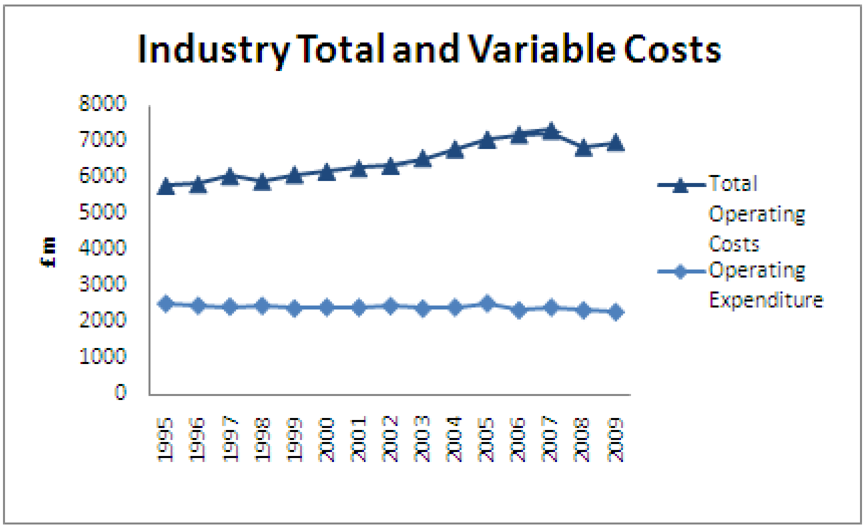
Figure 4.4: Mergers and Acquisitions

	1996-97	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Anglian							x	x	x	x	x	x	x	x	x
Dwr Cymru	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Northumbrian Water	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Severn Trent	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
South West	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Southern	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Thames	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
United Utilities	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Wessex	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Yorkshire Water Services LTD					x		x	x	x	x	x	x	x	x	x
B&W Hants	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Bristol	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Cambridge	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Dee Valley		x	x	x	x	x	x	x	x	x	x	x	x	x	x
Portsmouth	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
South East Water	x	x	x	x		x	x	x	x	x	x	x	x	x	x
South Staffs	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Sutton & E Surrey	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Veolia- Central (Three Valleys)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Veolia- East (Tendring Hundred)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Veolia- South East (Folkstone)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Mid Kent	x	x	x	x	x	x	x	x	x	x	x	x			
Sembcorp Bournemouth Water	x	x	x	x	x										
Essex and Suffolk	x	x	x	x											
Mid Southern	x	x	x	x											
North Surrey Water	x	x	x	x											
Hartlepool Water	x	x	x												
Yorks Waterworks	x	x	x												
Chester	x														
Wrexham	x														

4.4.1 Total Costs

Total costs are calculated as the sum of capital costs and operating expenditure. Total operating expenditure is obtained from the June Return and is calculated net of third party services, exceptional items, doubtful debts, service charges and local authority rates which are deemed as non-controllable costs by Ofwat and are not incorporated within their assessment of efficiency. Exceptional items are by definition atypical. Third party services relate to costs incurred for output produced by other companies. Local authority rates and doubtful debts are considered as non-controllable. High levels of doubtful debts are due to the legal and regulatory decision of prohibiting the shutting off of water and sewerage activities when bills are not paid. Service charges are charged by the EA and NRW for water abstraction. Total costs are decomposed into three inputs; labour, capital and other. The trends in total costs and operating expenditure are depicted in figure 4.5. There is a slightly downwards trend in the operating expenditure of -0.63% a year on average over the period. On the other hand, there is an upwards trend in total operating expenditure of 1.43% a year on average over the period. This is as a result of increasing capital expenditure to meeting quality standards, reducing leakage and increasing demand.

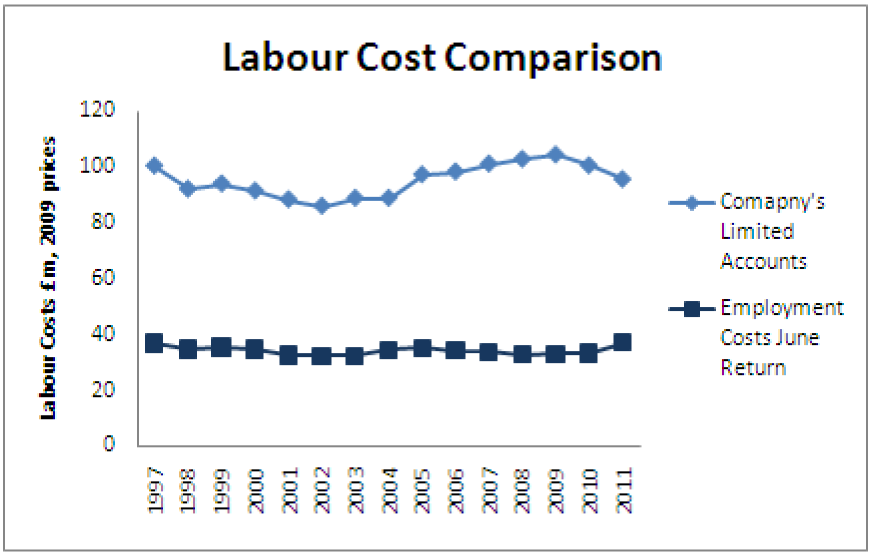
Figure 4.5: Industry Total and Variable Costs



4.4.2 Labour

Within the empirical literature, there have been several methodologies for the measurement of labour cost and the number of full time employees. Bottasso and Conti (2003, 2009) and Erbetta and Cave (2007) obtain labour costs and the number of full time equivalent employees from the June return. The employee costs obtained from the June return relate to direct labour costs. These costs therefore do not incorporate indirect labour costs associated with head office activities. Saal and Parker (2000, 2001), Stone and Webster (2004) and Saal et al (2011) obtain labour costs and the number of full time equivalent employees from the companies' limited company accounts which incorporate both direct and indirect labour costs. Figure 4.6 reports the WaSCs' average staff costs from the companies' limited accounts and the employment costs from the June return for both water and sewerage activities. The graph indicates that there is a substantial difference of around £60m per year between the two measures of labour costs. Direct costs make up a small proportion of total operating costs and a small proportion of companies total labour costs. Using the June return employment costs will incorporate indirect labour costs to other costs which in turn could lead to a misleading result with regards to the allocative efficiency.

Figure 4.6: Labour Cost Comparison



The measurement of staff costs from the companies' limited accounts is not a perfect measure as the accounts may include non-appointed activities. Following Stone and Webster (2004), capitalised staff costs are not excluded due to inconsistencies in the reporting in the company's accounts[9]. However an alternative approach would be to apply a labour wage rate index. This methodology does not enable the price of labour to vary amongst companies.

Overall, although there are several limitations to using the labour costs from the statutory accounts, the data provides a better reflection of the proportion of total costs that relate to the labour. Labour costs and the number of equivalent employees are obtained from the companies' limited accounts for the accounting year ending 31 March[10].

The price of labour is calculated as the ratio of labour costs and the number of full time equivalent employees obtained from the companies' limited company accounts. This methodology allows for the price of labour to vary across time and companies.

4.4.3 Dŵr Cymru

Dŵr Cymru Welsh Water contracted out the operations of their activities from 2001 to 2010. The staff costs and number of equivalent employees reported within their company accounts only related to head office activities. Therefore, in order to obtain a comparable measure of staff costs and the number of employees, the staff costs were obtained from Dŵr Cymru's annual limited company accounts alongside the accounts for the contracted services.

In March 2001 Dŵr Cymru Welsh Water contracted out their operations and maintenance of water and waste water assets, with the exception of its sewerage network, to United Utilities Operational until March 2010. Kelda Water Services undertook the maintenance and operation

[9] Anglian, Dwr Cymru, Severn Trent and Yorkshire do not report capitalised staff costs and instead report own work capitalised. Own work capitalised relates to the capitalisation of other costs such as materials alongside staff costs. The proportion of own work capitalised to staff costs varies substantially over the period causing difficulties in estimating the capitalised staff costs.

[10] Northumbrian 2000, Wessex 2000, Thames 2002, Wessex 2002 are reported for December

of waste water assets in South Wales and Hereford on behalf of Dwr Cymru on the 1st of April 2005 until March 2010.

4.4.4 Capital

4.4.4.1 Capital Stock

Capital stock can either be measured through a monetary measure or a physical measure, which can be proxied by the length of mains. Aubert and Reynaud (2005) highlight that the choice of length of mains as a proxy of capital is imperfect. This is because it does not take into account all of the capital employed and secondly it does not reflect any depreciation of capital.

There are two choices of monetary values for the capital stock which can be applied within the industry; namely the Regulatory Capital Value (RCV) and the Mean Equivalent Asset (MEA). The regulatory capital value measures the financial capital employed by WaSCs and WoCs and is a regulatory tool used by Ofwat for the purpose of setting price limits; the data is only available at a company level. The RCV was originally determined as the value of the company 200 days after privatization. This value is rolled forwards based upon the amount of capital expenditure less the amount of depreciation.

The MEA value is the estimation of the replacement cost of tangible fixed assets, reported for water and sewerage activities. Companies report both the gross and net MEA value; the gross value is the replacement cost of an old asset with a technically up-to-date new asset with the same service capacity. The net MEA value is applied to take into account depreciation, which reflects the remaining service potential of the capital stock.

Stone and Webster (2004) highlight the distinction between the two measures of capital; they note that the RCV represents the financial capital employed whereas the MEA value captures the quantum of capital inputs into production. Within this study, the measure of capital stock used as an input into production is proxied by the MEA value. The RCV value is used within the calculation of capital costs to reflect the level of investment made by companies which earn a rate of return.

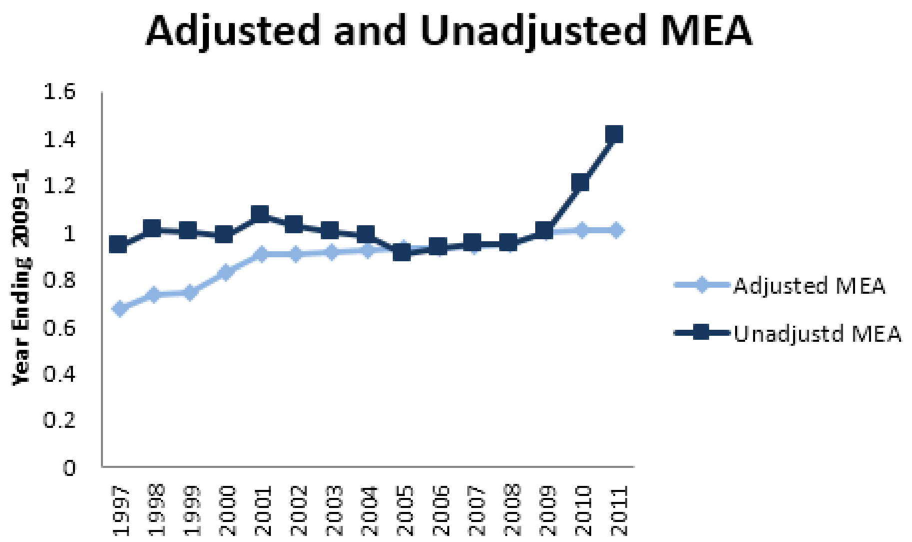
The MEA estimates are calculated based on an initial assessment of the replacement cost of capital and are rolled forwards based on an RPI adjustment, reclassification and an AMP adjustment, plus additions less disposals. The AMP adjustment adjusts the MEA value as a result of any MEA revaluations which take place periodically. The revaluations of the net MEA value results in arbitrary jumps in the measure of capital over time. Saal and Parker (2000, 2001), Saal and Reid (2004) Stone and Webster (2004), Erbetta and Cave (2007) and Saal et al (2011) adjust the MEA value in order to smooth out the series for the revaluations.

The capital stock is adjusted in order to smooth out the series for any revaluations following the methodology employed within the literature. The MEA value for the year ending 2009 is used as a starting point. Saal and Parker (2005) highlight that the most accurate choice of MEA value should embody all previous revaluations, therefore the most recent MEA revaluation is chosen. The 2009 price review was chosen as it was the most recent revaluation within the price review; this was also employed by Saal et al (2011). The MEA value is updated going forwards and backwards based upon the net investments in 2009 prices. Net investment is the sum of disposals, additions, investments and depreciation. Stone and Webster (2004), Saal and Parker (2005), Bottasso and Conti (2009) and Maziotis et al (2012a,b) average the year ending and year beginning MEA value to provide a more accurate representation of the amount of capital stock available to the companies in a given year. The MEA values will be averaged for the measurement of β-convergence and the application of three-stage DEA. The measurement of dynamic DEA will apply the year beginning and year ending value in order to explicitly incorporate the level of investment in the capital stock. This methodology is consistent with the perpetual inventory method.

In the accounting policies, Ofwat uses RPI to adjust the MEA value to the current prices. Saal and Parker (2000) state that the industry argues that an industry specific Capital Cost Index (CCI) prepared for the industry by London Economics is a more accurate movement in capital prices. Stone and Webster (2004), Bottasso and Conti (2009) and Erbetta and Cave (2007) use the Construction Output Price Index (COPI) as data for the CCI was not available. Following this methodology, net investment is deflated to 2009 prices using COPI.

The adjusted and unadjusted MEA values for the average WaSCs are displayed in figure 4.7. The adjusted series is smoother than the unadjusted as a result of the impact of the AMP revaluations being removed. As highlighted by Stone and Webster (2004) the unadjusted series provides a better reflection of the growth of the capital series than the unadjusted series. There is a large increase in the unadjusted MEA value in 2011 this was as a result of large AMP adjustments by WaSC 7 and WaSC 8 for sewerage activities.

Figure 4.7: Adjusted and Unadjusted MEA



4.4.4.2 Capital Costs

Capital costs are defined as the sum of capital charges (Current Cost Depreciation, CCD and Infrastructure Renewal Charges, IRC) and the estimated financing cost of assets employed. CCD is the depreciation of non-infrastructure assets, which are above ground assets based on the asset life. IRC is an annualised cost of maintaining underground assets charged to the profit and loss account. The IRC is a 15 year average of the infrastructure renewal expenditure. The estimated financing cost of assets employed is calculated as the Weighted Average Cost of Capital (WACC) multiplied by the capital rate base. Saal and Parker (2000), Stone and Webster (2004) and Erbetta and Cave (2007) measure the capital rate base using the MEA value. Saal et al (2011) and Cherchye et al (2013) on the other hand measure the cost of capital as the product of the WACC and the RCV. The RCV is applied instead of the MEA value for the calculation of capital costs as it represents the amount spent by the companies instead of the replacement cost of capital.

Within the regulatory control Ofwat determines the cost of capital. The cost of capital should be high enough to attract investors and to allow firms to finance their functions. On the other hand, if the cost of capital is too high firms will earn windfall profits. Ofwat's assumptions on the cost of capital are shown in table 4.1.

Table 4.1: Cost of Capital

Price Review	Cost of Capital-Post Tax	Mid-Range Cost of Capital	Small Company Premium
1994	5%–6%	5.5%	0.75%
1999	4.25%–5.25%	4.75%	0.4%–0.75%
2004	5.1%	5.1%	0.75%
2009	4.5%	4.5%	0.1%–0.4%

There are three measures of the WACC used within the literature. The first is applied by Saal and Parker (2000) and Bottasso and Conti (2009) where the measure is based upon the assumptions made by Ofwat on the cost of capital for the price review. Ofwat makes assumptions with regards to the risk free rate of capital and the risk premiums associated with equity and debt within the water and sewerage industry. Ofwat also makes assumptions with regards to small company premiums. The second methodology is considered by Stone and Webster (2004) who proxy the risk free rate of interest by the nominal rate of ten-year UK gilts instead of applying Ofwat's assumptions with regards to risk free rate of capital. The risk premium assumed by Ofwat as the weighted average of the risk premium associated with water company equity and corporate debt is added. The tax benefit associated with debt financed is subtracted.

The third measure is applied by Saal et al (2011) who follow the methodology applied by Stone and Webster (2004) with regards to proxying the risk free rate of capital by the medium-term UK gilt rate. Stone and Webster (2004) apply a risk premium attached for the common risk premium for both debt and equity. Saal et al (2011) apply Ofwat's assumption with regards to risk premiums and leverage to calculate a cost of equity and a cost of debt. The cost of equity is calculated as a function of the midrange values of Ofwat's assumption on equity beta and equity risk premium including small company premiums and the gilt rate. Similarly the cost of debt is calculated as a function of the midrange values of Ofwat's ex-ante assumptions on the debt premium and the gilt rate. The WACC is calculated using the cost of equity and debt alongside Ofwat's ex-ante assumptions with regards to gearing ratios and effective corporate tax rates.

Cherchye et al (2013) apply the WACC of Ofwat as these are used within the regulatory price review. The paper notes the robustness of the measurement as the methodology is based on discussions with firms, other regulators, consultants and shareholders. The WACC within this thesis takes the mid-range point estimate of the assumptions made by Ofwat within their price reviews reported in table 4.1.

4.4.4.3 Price of Capital

The price of capital is then the cost of capital divided by capital stock. The capital stock is measured by the MEA value.

4.4.5 Other Inputs

Other inputs are calculated as the total operating costs less labour costs and capital costs. Other costs therefore include cost for materials, energy, outsourced services etc. As other costs are a composite of costs there is no direct measure of an input. Within the literature there are two distinct ways of measuring the price and amount of inputs for other costs. Firstly, there is the approach adopted by Bottasso and Conti (2009), whom define the price of other inputs as other costs divided by the length of mains. Secondly Saal and Parker (2000), Ebretta and Cave (2007) and Saal et al (2011) employ a price index for the price of other inputs. Saal and Parker (2000) use RPI, whilst Erbetta and Cave (2007) apply a weighted average of the RPI and a real index of energy for the industrial sector where the weights are determined by the cost share. This approach takes into account the heterogeneous nature of the price index. The price of other inputs follows Saal et al (2011) and Cherchye et al (2013) employing a UK price index from the Office of National Statistics (ONS) for materials and fuels purchased for the purification and distribution of water industry. A measure for the physical amount of other inputs is calculated as other costs divided by the price of other costs. Following Erbetta and Cave (2007) all costs apart from power costs are deflated using RPI to 2009 prices; power is deflated by an energy price index for the industrial sector derived from the Department for Trade and Industry (DTI).

4.4.6 Output Variables

Four output variables are considered for the measurement of efficiency for water and sewerage activities at the activity level: the amount of water delivered, equivalent population served, number of water properties and the number of sewerage properties. Garcia and Thomas (2001) and Stone and Webster (2004) highlighted the advantage of including both the number of properties and the amount of water delivered within a joint specification. The marginal cost of producing an extra mega litre of water is substantially different from connecting an additional property.

The average number of water and sewerage properties served for WaSCs are shown in figure 4.9. The graph shows that the number of properties is relatively stable over time with a slight upwards trend. The number of sewerage properties is higher than the number of water properties as the WaSCs supply the sewerage properties within the WoCs operating areas.

Figure 4.9: Water and Sewerage Properties

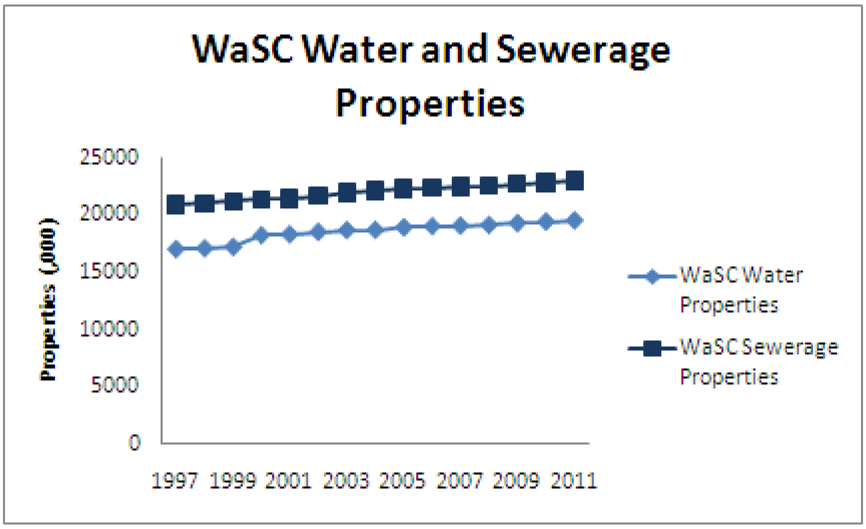
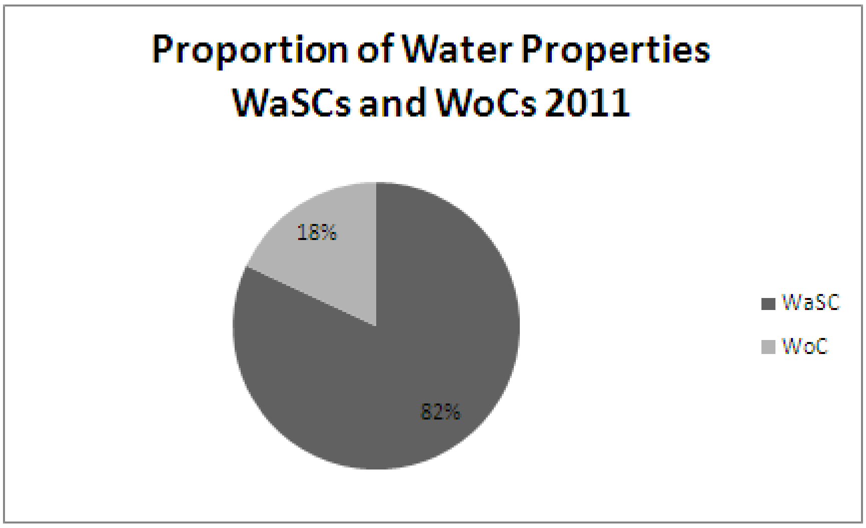


Figure 4.10 shows the proportion of the water properties in 2011 which are served by WaSCs and WoCs. The figure shows that the WaSCs serve the majority of the industry. The proportion of the industry served by WoCs has fallen from 22% in 1997 to 18.2% in 2011 due to a series of mergers.

Figure 4.10:Water Properties- WaSCs and WoCs

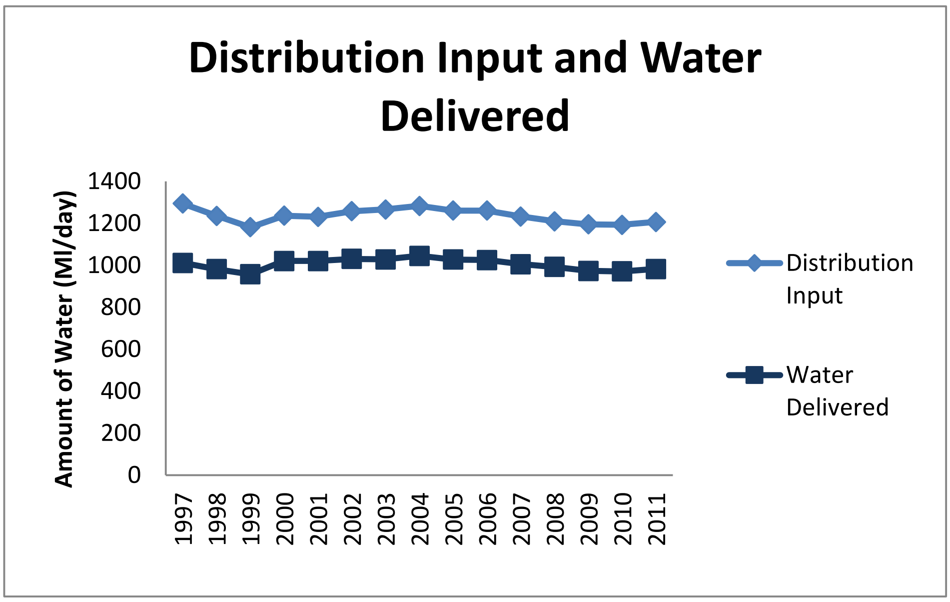


4.4.7 Physical Output

For the measurement of the physical output of water activities two variables are used within the academic literature, either the Distribution Input (DI) or Water Delivered. The latter deducts the level of leakage. Ofwat imposes leakage targets on the WaSCs and WoCs, reducing the level of leakage by a third since its peak in the mid- 1990's (Ofwat, 2012). There are substantial costs involved with the reduction of leakage which should be incorporated into the measurement of efficiency. Saal and Parker (2000) highlight the advantage of using water delivered instead of the distribution input as the latter does not take into account leakage which could bias the results. The majority of authors employ the use of water delivered (Cubbin and Tzanidakis, 1998; Thanassoulis, 2000a; Saal and Parker, 2005; Saal and Reid, 2004; Bottasso and Conti, 2003 and Bottasso and Conti, 2009) on the other hand Saal and Parker (2011) use the distribution input and account for the proportion of distribution losses as an environmental variables within the estimation of their cost function. Erbetta and Cave (2007) used the total volume of water delivered but also examined the impact of leakage, as a proxy of the condition of assets upon the efficiency scores through a second stage regression. By examining the production process the DI is an input into the production process whereas the final desirable output for consumers is the volume of water delivered.

Figure 4.11 depicts the industry average distribution input and water delivered over the period 1996/97 to 2010/11. The difference between the measures is attributed to leakages. There is an average difference over the period of 231ml/day which reduces from 283ml/day to 223ml/day. Overall water delivered is used as an output of the production process to incorporate leakage whilst it is considered as the output for customers.

Figure 4.11: WaSC Distribution Input and Water Delivered



The physical amount of sewerage is proxied by the equivalent population served. The equivalent population served is an estimate of the capacity of the sewerage treatment works based on the assumption that one period is equivalent to 60g of biochemical oxygen demand. The equivalent population served is applied by Saal and Reid (2004), Stone and Webster (2004), Saal et al (2007), Bottasso et al (2011) and Saal et al (2011) use the equivalent population served. Meanwhile Erbetta and Cave (2007) and Cherchye et al (2013) use the physical measure of waste water returned. The equivalent population served is utilised by Ofwat within their assessment of efficiency.

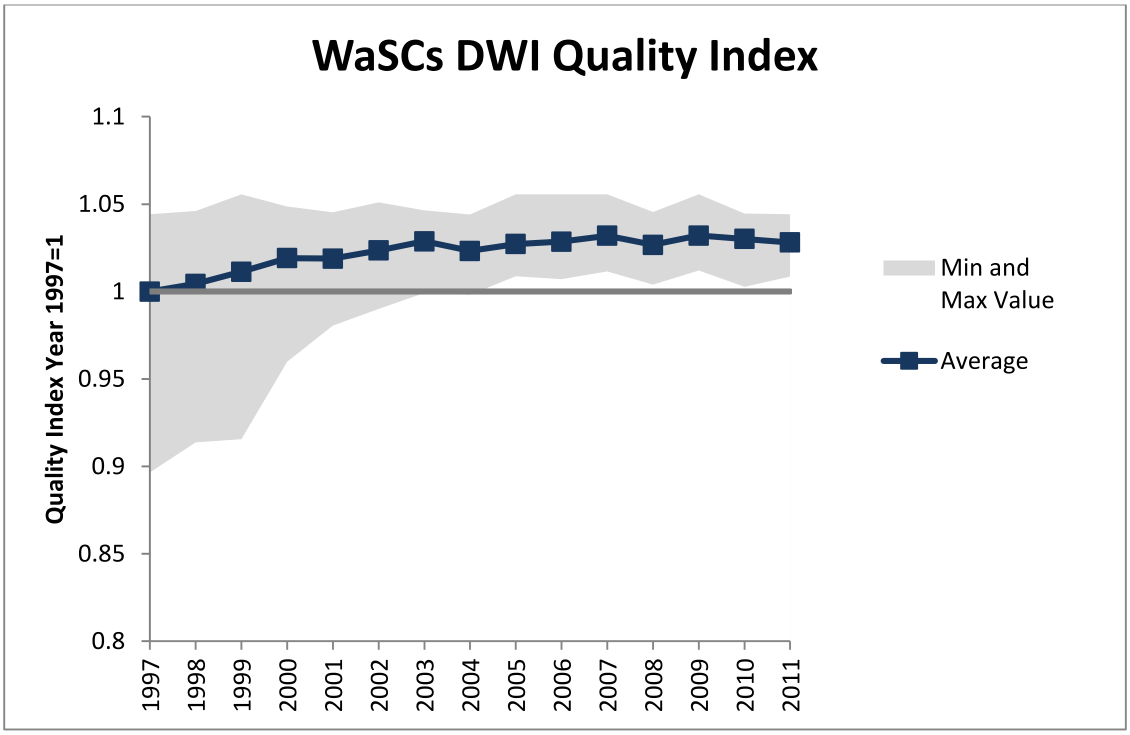
4.4.8 Environmental Variables

The methodology chapter highlighted the need to control for the impact of environmental and operating characteristics beyond the control of management. Saal and Parker (2000) were the first to introduce quality into the measurement of efficiency. Appendix 1 outlines the environmental variables which have been used for the measurement of efficiency for the English and Welsh water and sewerage industry. Given the choice of environmental variables used within the literature, six have been incorporated for the measurement of efficiency within this study.

Saal and Parker (2000) highlight the need to control for changes in quality of water and sewerage output due to improved quality standards. Water quality is taken into account using a quality index which is defined as the ratio of the average percentage of each WaSCs' water supply zones that are compliant with key water quality indicators as defined by the Drinking

Water Inspectorate (DWI[11]). This methodology follows that of Saal and Parker (2000), Erbetta and Cave (2007) and Bottasso and Conti (2009). The average quality index for the WaSCs alongside the minimum and maximum values is shown in figure 4.12. A value greater than 1 implies that there has been a quality improvement relative to the average quality in 1997, whereas a number less than 1 implies that the quality is lower than the average in 1997. The graph indicates an improvement in the average quality throughout the period and a large increase in the minimum compliance until 2003.

Figure 4.12: WaSC DWI Quality Index

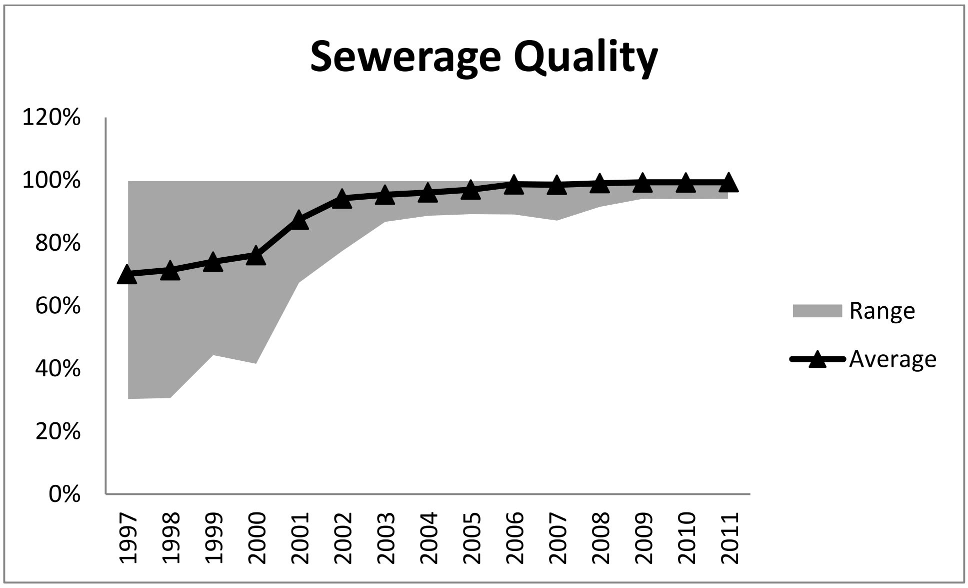


Sewerage quality is accounted for by calculating the proportion of the total load receiving secondary treatment. Figure 4.13 depicts the range of the sewerage quality index alongside the

[11] The average of several key indicators are considered: taste, odour, nitrate, aluminium, iron, lead and pesticides

average. The graph shows that there is a substantial improvement with regards to sewerage quality from 1997 to 2003.

Figure 4.13: Sewerage Quality



The density of water and sewerage activities is considered to account for the differences in costs between operating in a rural and urban area. Bottasso and Conti (2003) and Erbetta and Cave (2007) report that those firms operating under a higher water network density operate under a favourable environment. On the other, hand Saal and Reid (2004) report that higher water density results in higher opex costs, and therefore urban areas are unfavourable. The evidence for sewerage is mixed with several papers reporting that operating under an urban area is favourable whilst others report that it is unfavourable. Erbetta and Cave (2007) do not find a statistically significant influence of sewerage density on inputs slacks. Saal and Reid (2004) report that a higher sewerage density reduces opex costs, therefore operating in an urban area is favourable. Tupper and Resende (2004) regress sewerage density amongst other variables on the reciprocal of the efficiency score. They highlight that the expected influence is negative however report a positive influence. This indicates that operating within a high density area is unfavourable, although the impact is insignificant.

Saal and Parker (2005) and Saal et al (2011) include squared terms for density for water activities and both water and sewerage activities respectively. Their results indicate that as density increases, costs/inputs are reduced, but this effect is exhausted at a sufficiently high level of density. Overall, this indicates that those operating in very rural and very urban areas operate under the most unfavourable operating environments. The density of a company's water operations is calculated as the total water population divided by the length of mains. Similarly, sewerage density is calculated by the total sewerage population divided by the length of sewers.

The proportion of water abstracted from rivers is included to take into account the differences in costs from abstracting from boreholes, rivers and reservoirs. Abstraction from river sources has a relatively low abstraction costs but a relatively high treatment costs. However abstraction from boreholes requires higher power costs for abstraction and requires less treatment due to the high purity of underground water. Erbetta and Cave (2007), Saal et al (2011) and Cherchye et al (2013) find a higher proportion of distribution inputs from boreholes is favourable. On the other hand, Saal et al (2007) report a higher proportion of distribution input from boreholes increases input requirements. Bottasso and Conti (2003, 2009) and Cherchye et al (2013) incorporate the proportion of distribution input abstracted from rivers. Bottasso and Conti (2003) and Cherchye et al (2013) report that a higher proportion of abstraction from rivers is unfavourable, on the other hand Bottasso and Conti (2009) find that the proportion of abstraction from rivers is favourable, although imprecisely estimated.

The final environmental variable is the proportion of trade effluent, calculated as the volume of trade effluent divided by the volume of waste water returned. This represents the proportion of industrial effluent in the total waste water. The higher the proportion of trade effluent one expects to incur higher costs, especially with regards to treatment and energy costs due to the higher intensity of industrial trade effluent. Saal et al (2007) report that a higher proportion of trade effluent requires higher input requirements. Erbetta and Cave (2007) find an insignificant relationship between labour, other input slacks and trade effluent but report a negative and significant relationship for capital, therefore indicating that a higher proportion of trade effluent reduces slacks. This indicates a better performance.

4.5 Summary

This chapter has firstly outlined several specification issues and secondly outlined the data to be used within the following three chapters for the measurement of efficiency within the English and Welsh water and sewerage industry. The majority of the data was obtained from the June return and supplementary data was obtained from the companies' limited accounts and the DWI annual reports. The data does differ slightly between each chapter due to differing data requirements for each technique, which are outlined in their respective chapters. The following chapter will examine the rate of convergence of the efficiency scores within the English and Welsh water and sewerage.

5. Convergence in Efficiency

5.1 Introduction

The regulatory price controls are undertaken every five years and operate through the use of $RPI + Q - X$. The X factor is decomposed into two elements: a frontier shift and catch-up factor. The frontier shift is the amount by which Ofwat assumes that all companies can make a minimum efficiency saving, the level of technical change. The second component is the catch-up factor based on Ofwat's estimates of inefficiency through yardstick competition. The catch up factor is designed to encourage the inefficient firms to catch-up with the most efficient company by imposing stricter efficiency targets for the inefficient companies.

Relative efficiency is determined through a series of COLS regressions at the functional level for water and sewerage activities. The predicted and actual expenditure are summed to the activity level and the distance to the frontier company is determined. Capital expenditure up until the 2004 price review was examined through a combination of COLS functional models and standard costs. The efficiency challenge is based on closing a proportion of the efficiency gap to the frontier, which is reported in table 2.5. The catch-up factor in 1999 was 60% for opex, 50% for capital maintenance and 75% for capital enhancement for the econometric models. The catch-up factors for the standard costs are reported in table 2.6. The efficiency challenge is larger for those firms who are inefficient to encourage firms to catch up with the frontier. Capital expenditure in the 2009 price review was assessed through menu regulation using the CIS which is designed to encourage truth-telling within firm's submissions of predicted costs. Companies submit their plan and Ofwat establishes an independent baseline which they believe an efficient company would spend. Companies are incentivised to submit realistic and challenging plans as they can obtain favourable incentive rates and additional income. If firms achieve beyond expectation they receive incentives based upon the level of outperformance. On the other hand any overspend is penalised by a proportion of that overspend.

Overall, if regulation has been an effective tool this should result in productivity improvement through a shift in frontier and the catch-up factor should have encouraged convergence of efficiency scores towards best practice. A large amount of literature has emerged to examine whether privatisation and regulation of the industry has led to productivity improvements. Overall, the results indicate that privatisation has not led to an improvement in productivity whilst regulation has been effective in encouraging productivity improvement. This chapter examines whether the larger efficiency targets for the least efficient companies has resulted in the least efficient firms catching up with the most efficient firms, resulting in convergence. Cubbin (2005) state that over time we should expect the range of inefficiencies to narrow by the extent in which the incentives of regulation are effective. Bottasso and Conti (2003) and Saal et al (2007) examine the evolution of efficiency scores over time and report that the standard deviation of efficiency scores has decreased.

This chapter extends upon the literature to estimate the presence of beta- and sigma-convergence (β- and σ-convergence) in the efficiency scores. The companies' efficiency scores are measured using DEA whilst accounting for changes in quality through a quality-adjusted measure of output. β -convergence determines whether the least efficiency firms are improving at a faster than the most efficient firms, whereas σ -convergence determines whether the dispersion of efficiency scores has fallen over time. Finally, a Partial Adjustment Model (PAM) is estimated to determine whether firms are converging towards the frontier or whether inefficiencies are persistent. If the catch-up efficiency challenge has been effective, one would expect to see convergence amongst the firms within the industry towards the frontier.

The rest of this chapter is organised as follows. Section 2 will review productivity literature within the English and Welsh water and sewerage industry. Section 3 outlines the methodologies applied for the measurement of efficiency and convergence. Section 4 will describe the data and variables used within estimation. Finally, the empirical results and discussions are reported in section 5.

5.2 Literature review

The privatisation and regulation of the English and Welsh water and sewerage industry has spawned a large amount of literature examining whether policy change has been effective in improving productivity and efficiency. Productivity and efficiency have been analysed through the measurement of cost functions and frontier models, applying both parametric and non-parametric approaches. This section will provide a brief overview of the literature examining whether privatisation and different price reviews have been effective in encouraging Total Factor Productivity (TFP) growth, technical change, scale change and efficiency change.

One of the first paper to examine the impact of privatisation and regulation was Saal and Parker (2000). Saal and Parker (2000) examine the impact of privatisation and regulation in the English and Welsh water and sewerage industry for the WaSCs over the period 1985–1999 through the estimation of a multi-output translog cost function. To allow for the effect of substantial improvements in quality driven by the DWI and EA, Saal and Parker (2000) model a quality adjusted cost function alongside a quality-unadjusted model. Quality is incorporated through a quality-adjusted measure of output. These results highlight the importance of taking into consideration substantial changes in quality standard within the industry through the impact on the interaction term between water and sewerage activities and finding an improvement in the quality of one output may reduce the cost of producing the other. The multi-output specification allows for the joint production of water and sewerage activities. Saal and Parker (2000) find there are intrinsic links between water and sewerage activities although no presence of economies of scope. The impact of privatisation and regulation upon productivity is examined. This paper reports an increase in productivity post-privatisation however reports that the 1994 price review was the main contributor rather than privatisation. The parameter estimates for the 1994 review suggest that after the regulatory review costs rose more slowly by 0.5% per annum.

Ashton (2000a) employs a variable translog cost function to analyse firm-specific cost efficiency for the period 1987–1997. The paper finds a moderate level of dispersion in the efficiency scores which is described as an indicator of the differences in the operating environments and performance within the sector. Ashton (2000b) examines TFP and technical change through the estimation of a translog cost function for the period 1989–1997. Their results display a TFP productivity growth of -0.046% which has worsened over the period, indicating a decline in productivity. TFP is decomposed into economies of scale, technical change and changes in the level of outputs. The measure of technical change over the period reports a negligible effect of -0.048%. Overall, the paper concludes that the initial level of investment post-privatisation has not influenced the productivity of the industry over the period and that privatisation has not improved the level of technical change or productivity growth.

Bottasso and Conti (2003) examine operating cost efficiency for the WaSCs and WoCs for water activities and estimate an SFA cost function to examine the evolution of efficiency over the period 1995 and 2001. The model incorporates environmental variables within the error term to allow for the environmental variables to directly impact efficiency. Their results find that operating cost inefficiency has decreased over the sample period and that inefficiency differentials amongst firms has narrowed. Bottasso and Conti (2003) apply a Kruskal-Wallis test which rejects the null hypothesis that the median scores for each year are equal and find a cost reduction of 5% between 1995 and 2001. The Kruskal-Wallis test is a non-parametric technique for testing whether the samples originate from the same distribution. They report that that standard deviation has reduced from 0.08 in 1995 to 0.03 in 2001, implying convergence in the efficiency scores.

Saal et al (2007) examine productivity growth and efficiency for the English and Welsh water and sewerage industry through the estimation of an SFA input distance function for water and sewerage activities for WaSCs. The parameter estimate for the time trend indicates that the sample average firm has experienced technical change of 1.8% in the mid-year sample. The second order term although insignificant indicates that technical change has decreased over the period from 2.2% in 1986 to 1.39% in 2000. Saal et al (2007) examine the evolution of the efficiency scores over time and they highlight a fall in the mean and median efficiency score post-privatisation which might be the result of the particularly lax price review in 1989. Efficiency began to improve after 1993 which they highlight may have been due to the signalling by Ofwat of its intensions to tighten the regulation of the industry to raise efficiency. Over the period 1990 to 2000 the efficiency level for the least efficient firms has improved. Saal et al (2007) conclude that the price cap had a positive impact on those firms which were inefficient at privatisation. However, the paper reports that the average efficiency score at the end of the period was moderately lower than at privatisation. The paper highlights that efficiency scores for the average firm are relatively high and therefore substantial gains in efficiency may be hard to achieve.

An index number approach is examined by Saal and Parker (2001) who report that privatisation led to an increase in labour productivity, in which majority was after the 1994 price review. Overall TFP did not increase significantly; this suggests that the exclusion of capital expenditure may leave to overestimated productivity gains. Overall the paper concludes that neither privatisation nor the 1994 price review improved productivity.

Saal et al (2007) measure TFP for the WaSCs for water and sewerage activities estimating an input distance function using SFA. Saal et al (2007) note that there are substantial differences in the average TFP estimates compared to Saal and Parker (2001) due to the increased complexity of the methodology in Saal et al (2007). Both papers find that productivity growth was not statistically different after privatisation and that productivity growth rates were lower in the 1995–2000 period than they had been before privatisation.

Saal et al (2007) decompose TFP growth rates into technical change, efficiency change and scale change. Technical change is always positive and relatively stable. The average annual increase in technical change increased from 1.61% before privatisation to 2.19% for the entire post privatisation period. The scale effect has a negative impact on TFP over the whole period, which is consistent with the results that WaSCs are characterised by diseconomies of scale. Saal et al (2007) highlight that efficiency change trends are the main determinant of the trends in the average WaSC TFP growth. The average annual efficiency change decreased from 0.4% in the pre-privatised period to -0.16% for the post-privatised period, however the difference is not statistically significant. The average efficiency score in 2000 was 98.4% of the level of privatisation in 1990. Saal et al (2007) conclude that it would appear that if privatisation and/or the imposition of $RPI + K$ regulation has had an impact on WaSC performance, it has been primarily through encouraging faster technical change rather than encouraging firms on average to move closer to the industry efficient frontier. However it has been effective in encouraging the least efficient firms to catch-up to the frontier.

The previous studies discussed have examined the impact of privatisation and regulation upon efficiency and productivity. Erbetta and Cave (2007), Bottasso and Conti (2009) and Saal and Reid (2004) examine the influence of the 1994 and 1999 price review. Erbetta and Cave (2007) measure cost efficiency and examine the impact of regulation using a two-stage DEA model. Within the first stage, the technical and allocative efficiency scores are calculated. Graphing the average efficiency score over time finds an improvement in both the technical and allocative efficiency score. The graph indicates that technical efficiency has improved by around 5% since the 1999 price review and that the 1994 price reviews has had no effect on allocative efficiency. The second stage employs an SFA model to examine the impact of environmental variables on the input slacks. Separate regressions are analysed for allocative[12] and technical input slacks for each input. A time trend is included within the second stage regression to examine whether the utilisation of inputs has increased or decreased over the period. The time trend for technical efficiency input slacks indicates a more intensive use of inputs over time. The results also indicate positive technical change for the allocative component. Overall, the two time trends represent a more intensive use of inputs offset by an improvement in the combination used.

Erbetta and Cave (2007) examine the impact of the different regulatory price reviews on the excess input slacks by including regulatory dummies alongside the environmental variables within the second stage regression. The regulatory dummies indicate efficiency improvement as a result of the 1994 and 1999 price review for technical efficiency, although the effect of 1994 price review is not significant. The regulatory dummies are not significant for allocative efficiency.

Saal and Reid (2004) estimate opex productivity by employing a quasi-fixed variable capital translog cost function for water and sewerage activities for the WaSCs. Whilst their results did not consider pre-privatisation, their results coincide with those of Saal and Parker (2000) to highlight that the 1994 price review improved productivity. However, the paper finds that the 1999 price review did not bring any improvements in productivity. The paper reports

[12] The allocative efficiency slack is measured as $TE.x - x^{*}$. TE is the technical efficiency and x^{*} is the optimal input derived from the DEA linear programme. Allocative efficiency can take a positive or negative value, the absolute value is taken as a measure of the level of distortion without distinguishing between input over- or under-utilisation

substantial productivity gains have been achieved through substitution of opex with capital investment. The elasticity of opex with respect to labour costs has declined considerably, which is consistent with the substitution away from labour towards other inputs that has occurred within the industry.

Bottasso and Conti (2009) examine economies of scale and technical change for the WoCs using a variable cost function for the period 1995–2005. Their results on total factor productivity indicate that the rate of technical change is much higher after the 1999 price review, rising to 1.6% a year in comparison to the rate after the 1994 price review which was close to zero. Bottasso and Conti (2009) formally test whether variable costs were significantly different as a result of the 1999 price review through the inclusion of an interaction term between the time trend and a dummy variable for the five years after the 1999 price review. Their estimates reveal that the 1999 price review reduced costs, although the effect was insignificant. Their results indicate that the improvement was partly due to labour saving technological progress.

Saal and Parker (2005) examine productivity for water activities for WaSCs and WoCs through the application of an input distance function. Saal and Parker (2005) find an average positive productivity growth for water activities between 1994 and 2000 for WaSCs and WoCs. They note that the main contribution of the positive productivity growth is attributed to technical change.

Portela et al (2011) examine productivity for the period 1993–2007 through a meta-Malmquist frontier for operating expenditure. The paper reports productivity improvements between 1993 and 2005 of which most were due to technical change instead of catching up to the frontier. Portela et al (2011) report that over time companies' efficiency scores are moving closer to the meta-frontier. The meta-frontier envelops all companies within all time periods and the meta-efficiencies are therefore measured against a stationary frontier. The meta-frontier efficiency improved until 2002, remained relatively stable between 2002 and 2005 and began to decline between 2005/06 and 2006/07. Portela et al (2011) highlight that their results coincide with Ofwat's statement that the improvement in relative efficiency since the 1999 price review is striking. Towards the end of the period the year-specific frontiers are moving closer to the meta-frontier, therefore indicating that the main driver of productivity is technological change. The data is divided into four groups representing the different regulatory periods and the results show an improvement in productivity over the regulatory periods with the exception of the 2004 price review. The paper does highlight that productivity as a result 2004 price review may have been underestimated due to the rise in electricity prices and the implementation of leakage control being charged to opex rather than capex.

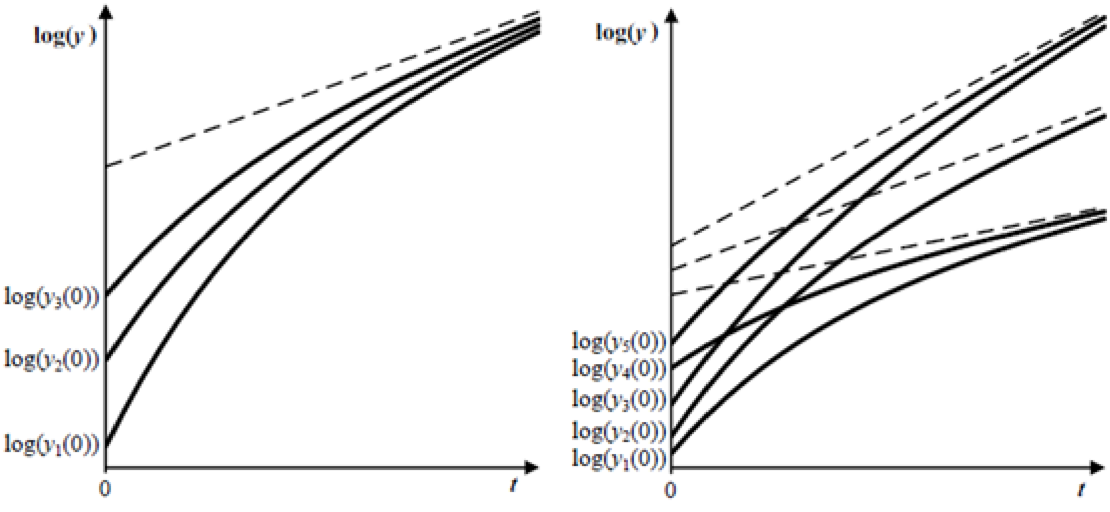
Overall, the studies examining the implication of privatisation and the 1994 price review report that privatisation did not statistically improve productivity. The results of the effect of the 1994 price review are mixed. Saal and Parker (2000) report an improvement in TFP due to the price review, whereas Saal and Parker (2001) and Saal et al (2007) report an insignificant impact on TFP. Bottasso and Conti (2009), Erbetta and Cave (2007) and Saal and Reid (2004) all examine the impact of the 1994 and 1999 price review. The results from Bottasso and Conti (2009) and Erbetta and Cave (2007) find no statistically different impact of the 1994 price review and report a significant improvement as a result of the 1999 price review. On the other hand, Saal and Reid (2004) report that the 1994 price review has a positive impact on TFP, whereas the 1999 price review did not have any effect of productivity. With regards to the dispersion of efficiency scores. Bottasso and Conti (2003) and Saal et al (2007) report a fall in the dispersion of the efficiency scores with the least efficient firms improving and therefore implying convergence. This chapter will extend upon this analysis to examine and quantify whether there has been any convergence in the efficiency scores from 1997 to 2011 through the estimation of β- and σ-convergence.

5.3 Literature Review on Convergence

This section outlines the procedure for identifying whether firms' efficiency scores have converged within the English and Welsh water and sewerage industry. The issue of convergence originates within the growth literature. Barro (1991) and Barro and Sala-i-Marin (1991, 1992) introduce the concepts of β- and σ-convergence.

β-convergence examines whether firms with a low initial level of efficiency improve faster than those with a higher initial level of efficiency. In other words, those firms with a lower initial level of efficiency are catching up with those of an initial higher efficiency. Therefore, over time, firms' efficiency scores should converge. This is measured by regressing the growth rate of efficiency on the initial level of efficiency. The concept of β-convergence has been applied to the analysis of efficiency by Fung (2006) and Casu and Girardone (2010).

Figure 5.1: Conditional Vs Unconditional β-convergence



A B

Source: Gluschenko (2012)

Convergence can be measured either through absolute or conditional β-convergence. Absolute convergence implies that firms are converging towards the same steady state efficiency level as depicted by the dashed line in figure 5.1a. If absolute β-convergence is present then firms converge to the same steady state where the least efficient firms are improving faster than the relatively efficient firms. Barro (2012) suggests the use of unconditional or absolute convergence in the situation where firms are reasonably homogeneous in terms of long-run or steady-state characteristics.

Conditional convergence depicted in figure 5.1b suggests that each firm or group of firms with a set of characteristics possess its own steady-state productivity level. This is measured by regressing the growth rate of the efficiency score on the initial level of efficiency and a set of characteristics. The fixed effects model is an extreme case of conditional convergence in which each firm has its own steady state which it is converging towards. Gluschenko (2012) highlight that under conditional convergence the interpretation differs. β-convergence no longer implies the speed of convergence to the industry steady state level of efficiency but rather the speed of convergence to the firm's own steady state efficiency. This is determined not by considering the distance between the most inefficient firm to the industry steady state but by the distance of each firm's efficiency level to its own steady state. If conditional β-convergence is present Barro and Sala-i-Martin (2004) state that efficiency grows faster the further a given firm is from its individual steady state efficiency score. Gluschenko (2012) states that conditional convergence is uninformative to how growth rates of different firms relate and therefore should be interpreted with caution.

There are several limitations of the application of β-convergence. Quah (1996) highlights that if β-convergence is present this implies that firms with a low efficiency score are growing faster than those with a high efficiency score. The situation may arise whereby those firms with lower initial efficiency may overtake the higher efficiency firms, therefore implying a lack of convergence. Quah (1996) also reports that β-convergence provides no information about the evolution of the dispersion of the efficiency scores. Sala-i-Martin (1996) states that a negative coefficient for β-convergence could be the result of measurement error and random shocks instead of real convergence; this is known as regression towards the mean (Friedman, 1992; Quah, 1993).

To overcome the shortcomings of β-convergence, σ-convergence is also examined. σ-convergence considers whether the dispersion (the standard deviation) of efficiency scores amongst firms diminishes over time. This is measured by regressing the growth rate of the standard deviation by the initial standard deviation. Sala-i-Martin (1996) states that β-convergence is a sufficient but not necessary condition for the existence of σ- convergence. Fung (2006) states that if β-convergence is to measure real convergence then it has to coincide with σ-convergence.

Evidence of β-convergence implies that the least efficient firms are improving at a faster rate than the most efficient firms. This might be due to the least efficient firms catching up with the most efficient firms or could be due to the most efficient firm's efficiency scores declining, or lagging behind. The regulatory tools utilised by Ofwat are designed to encourage catch-up to the frontier. Alongside the measurement of β-convergence and σ- convergence, a PAM is estimated to measure the adjustment of the efficiency scores to the best practice frontier. This identifies whether firms are converging towards the frontier or whether inefficiencies are persistent.

5.4 Methodology

5.4.1 Evaluating Efficiency

Technical efficiency is calculated through the BCC DEA linear programming problem in equation (5.1). Y is a R by N matrix, where R is the number of outputs and N the number of DMU's. X is a J by N matrix of inputs. θ_n is the efficiency score for *n*-th unit which takes the value $0 \leq \theta_n \leq 1$. As a result of the discussion in the methodology chapter, efficiency is examined under input-orientation and VRS. Input orientation is widely considered within the literature (Thanassoulis, 2000a,b; Cubbin and Tzanidakis, 1998 and Erbetta and Cave, 2007) as the demand level faced by suppliers is exogenous due to their legal obligation to supply water and sewerage services. VRS does not assume that firms are operating at their optimal size and therefore excludes any scale inefficiencies. Efficiency is examined under VRS as managers have very little control over their size, apart from mergers and acquisitions which are subject to review by the competition commission. A separate frontier is calculated for each time period to allow for technical regress and progress.

$$\theta_n = \min_{\theta,\lambda} \theta$$

$$s.t. \quad Y\lambda \geq y_n \qquad (5.1)$$

$$X\lambda \leq \theta x_n$$

$$\lambda \geq 0$$

$$i'\lambda = 1$$

5.4.2 Convergence

β- convergence is inferred by running the following panel data regression model:

$$\Delta\theta_{n,t} = \alpha + \beta(\ln \theta_{n,t-1}) + \varepsilon_{n,t} \qquad (5.2)$$

$\theta_{n,t}$= Efficiency of firm n at time t

$\varepsilon_{n,t}$ is a random effect and $\varepsilon_{n,t} \sim iid\ N(0, \sigma_\varepsilon^2)$

$\Delta\theta_{n,t} = (\ln \theta_{n,t}) - (\ln \theta_{n,t-1})$

Here $n = 1, . . , N$ and $t = 1, \dots, T$ and α and β are the parameters to be estimated. The dependent variable in the equation is the growth rate of efficiency. If the coefficient β between the growth rate of efficiency over time and the initial level of efficiency is negative, this indicates that presence of β-convergence. The larger the absolute value of the coefficient implies a greater tendency of convergence. Arbia and Piras (2005) calculate the half-life speed of convergence as $\tau = -\ln(2)/\ln(1 + \beta)$, this calculates the time span which is necessary for the current disparities to be halved. Therefore, a larger absolute value of β implies a faster rate of convergence and a faster half-life.

The steady state mean efficiency in which firms are converging towards is calculated as:

$$\theta = \exp\left(\frac{\alpha}{-\beta}\right) \qquad (5.3)$$

The estimation of σ-convergence is given in equation 5.4. This evaluates whether the dispersion of efficiency falls over time following Parikh and Shibata (2004), Weill (2009) and Casu and Girardone (2010).

$$\Delta E_{n,t} = \mu + \varphi \ln(E_{n,t-1}) + \xi_{n,t} \qquad (5.4)$$

Where:

$E_{n,t} = \ln(\theta_{n,t}) - \ln(\overline{\theta_t})$, where $\overline{\theta_t}$ is the mean efficiency score at time t,

$\Delta E_{n,t} = E_{n,t} - E_{n,t-1}$

$\xi_{n,t}$ is a random error and $\xi_{n,t} \sim iid\ N(0, \sigma_\xi^2)$.

A negative value for the parameter φ implies unconditional σ-convergence. The exponential of the intercept μ indicates the average dispersion from the mean.

To analyse the adjustment of efficiency scores to the frontier, the standard Partial Adjustment Model (PAM) is employed following Casu and Girardone (2010). The convergence towards the best practice; an efficiency score of 1 is evaluated by the following adjustment mechanism

$$\ln \theta_{n,t} - \ln \theta_{n,t-1} = \gamma(\ln \theta_{max} - \ln \theta_{n,t-1}) + \varepsilon_{it} \qquad (5.5)$$

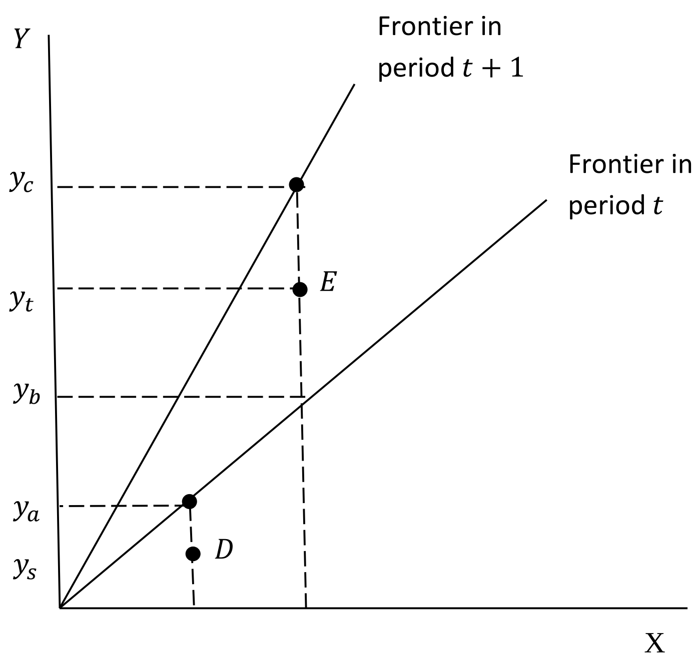
Here, $\theta_{n,t}$ and $\theta_{n,t-1}$ are defined as before and θ_{max} is the maximum attainable efficiency score i.e. unity. ε_{it} is the error term. γ is the adjustment parameter and measures the speed of adjustment towards the maximum attainable efficiency score. A positive value of γ signifies convergence towards the frontier whereas a negative value implies lack of convergence, or persistent inefficiency.

5.5 Interpretation of Convergence

β-convergence can be related to the Malmquist index of productivity discussed within section 3.5.5.8. Figure 5.1 shows a one input and one output production function for two time periods. Point D is the point of production in period t and point E is the point of production in period

$t+1$. The efficiency score in period t is calculated by the ratio y_s/y_a and y_t/y_c for period $t+1$. The Malmquist index decomposes productivity into technical change and efficiency change. Efficiency change is measured as $\frac{y_t/y_c}{y_s/y_a}$. The efficiency change is reported for each company and each time period and the geometric mean reports the average efficiency for each period. Portela et al (2011) examine the geometric mean of the efficiency change for WaSCs and WoCs for each period. Their results find a negligible influence on the average efficiency over the period. β-convergence is an extension on the efficiency change to examine whether the least efficient companies are improving at a faster rate than the most efficient firms. The growth rate of efficiency for each firm is calculated as $\log\left(\frac{y_t/y_c}{y_s/y_a}\right)$. This chapter will measure efficiency change alongside β-convergence to examine whether the average rate of efficiency change differs in comparison to the rate of efficiency growth for the least efficiency company compared the most the most efficient.

Figure 5.2: β-convergence and efficiency change



If both β- and σ-convergence are present this can be interpreted as an indicator of general improvement within the industry. The least efficient firms are improving at a faster rate, and therefore are catching up with the most efficient firms. σ-convergence indicates that the dispersion of efficiency scores across the firms is getting smaller over time. If the presence of β- and σ-convergence coincides with a positive coefficient for the PAM this implies that firms are converging towards the frontier. A negative value indicates that although the dispersion of firms is reduced, the level of inefficiency is persistent. Casu and Girardone (2010) state that the situation can be interpreted as those firms with lower initial efficiency level are catching up with the average and those above the mean are regressing towards the average.

Sala-i- Martin (1996) state that β-convergence is a necessary but not sufficient condition for σ-convergence and not vice-versa. Therefore there is a possibility to experience β-convergence without the presence of σ-convergence. The presence of β-convergence and lack of σ-convergence implies that those firms with lower initial efficiency levels experience a higher growth rate, although the dispersion of efficiency scores has not fallen. This scenario can arise if the least efficient firms overtake the more efficient firms.

5.6 Estimation

The unconditional β-convergence in equation 5.2 is estimated by applying pooled OLS with robust standard errors to correct for the presence of heteroskedasticity. Unconditional convergence implies that the firms are converging towards the same steady state efficiency score. Islam (1995) highlights that there might be underlying firm heterogeneities. In the presence of firm heterogeneities, the β-convergence can be denoted in equation 5.6

$$\Delta\theta_{n,t} = \beta\left(\ln \theta_{n,t-1}\right) + \eta_n + v_{nt} \qquad (5.6)$$

Where η_n is a company-specific effect to represent time-invariant firm heterogeneities. Equation 5.6 can be re-written as an AR(1) equation[13]:

$$\ln \theta_{n,t} = b(\ln \theta_{n,t-1}) + \eta_n + v_{nt} \quad (5.7)$$

Where $b = \beta + 1$.

In the presence of an AR(1) model, OLS gives an estimate of b that is biased upwards in the presence of company-specific effects, therefore giving a downwards estimates of the rate of convergence β (Bond et al, 2001). To incorporate company-specific effects, equation 5.6 can be estimated by using Fixed Effects (FE) or Random Effects (RE). FE controls for time-invariant effect so that the estimated coefficients are not biased because of omitted time-invariant characteristics. RE assumes that individual country effects are uncorrelated with the exogenous variables, an assumption that is sometimes violated. If the individual country effects are correlated with the exogenous variables then the estimates would be biased and inconsistent and FE would be preferred.

To determine the correct estimation procedure a series of tests are conducted. To determine whether FE or RE is preferable, a Hausman test is applied. The Hausman test examines the null hypothesis that the individual effects are uncorrelated with the other regressors in the model. If rejected the random effects model produced bias estimators, and therefore a FE model is preferred. To test for FE against the OLS model an F-test is applied which tests whether the fixed effects are zero. To test for RE against OLS, a Breusch Pagan Lagrangian multiplier test is applied which tests whether the cross-section variance components are zero.

[13] The β-convergence equation is re-written as an AR(1) to estimate β-convergence using dynamic panel data models of Arellano and Bond (1991).

The application of FE or RE addresses the issue of the omitted variable bias within the OLS model. The inclusion of fixed effects measures conditional β-convergence. Conditional β-convergence estimates the rate of convergence after accounting for differences in the steady states across countries (Islam, 1995). Islam (1995) highlights that the rate of β-convergence is faster under conditional β-convergence than unconditional β-convergence due to omitted variable bias. It is important to reiterate that Gluschenko (2012) highlights that the interpretation differs for conditional convergence.

The fixed effects estimator controls for the time-invariant characteristics, but introduces an issue with regards to endogeneity. By construction, the lagged dependent variable in equation 5.7 is correlated with the individual specific effect, i.e $E(\mu_n|\theta_{n,t-1}) \neq 0$. Nickell (1981) shows that the coefficient on the lagged dependent variable (b) is biased downwards in the presence of fixed effects, therefore biasing the estimate of β-convergence upwards. The bias tends to zero as the time span gets large, however the bias can be large in short panels. Bond et al (2001) state that a consistent estimate of b is expected to lie in between the OLS levels and FE estimate.

To correct for the presence of fixed effects within dynamic panel data models, Arellano and Bond (1991) introduce the first difference GMM estimator (GMM-DIFF) for panel data with large N and finite T. The first differencing in equation 5.8 removes the fixed effects; however the first differences error term $v_{n,t} - v_{n,t-1}$ is correlated with the lagged dependent variable $\theta_{n,t-1}$ through the term $v_{n,t-1}$.

$$\Delta \ln(\theta_{n,t}) = b\Delta \ln(\theta_{n,t-1}) + \Delta v_{n,t} \qquad (5.8)$$

Therefore, the GMM-DIFF instruments the first difference with the lagged levels. Bond et al (2001) highlight that the GMM-DIFF can be problematic as the lagged levels might be weak instruments for the first differences. To address the issue of weak instruments the System-GMM (SYS-GMM) introduced by Blundell and Bond (1998) is applied. The SYS-GMM instruments the lagged dependent variable by the first difference.

5.7 Data

The specification of the English and Welsh water and sewerage industry was extensively discussed in the data chapter. The rate of convergence is analysed for the WaSCs for the period 1996/97–2010/11. Efficiency is examined for each period under a separate frontier. A fundamental stage within DEA analysis is the choice of inputs and outputs. One of the limitations of DEA is the issue of dimensionality; this being that the convergence rate becomes slower as the number of inputs and outputs is increased. The issue of dimensionality reduces the discriminatory power of the model. Therefore a trade-off is considered between specifying the most accurate model whilst avoiding the problem of dimensionality. In the presence of panel data, this problem can be overcome by pooling the data under a common frontier; treating each time period as if it were a separate DMU. However, this implicitly assumes that there is no technological progress and/or regress. The evaluation of convergence requires separate technology for each time period[14], therefore the minimum number of DMUs within the analysis is ten WaSCs.

A one input and two output model is analysed with the input as the costs. As the inputs within the models are costs, efficiency can be interpreted as cost efficiency. This specification therefore implies that firms face the same input prices. Thanassoulis (2000a,b) models operating expenditure using DEA and states that companies face similar staff and material

[14] A separate frontier is required; otherwise the speed of convergence will be biased as firms will not be examined against their true technology available within the period. In the presence of technological change firms at the beginning of the period will be relatively inefficient, as the technology is not available. Improvements in the efficiency score under a meta-frontier can be as a result of technical change or efficiency. This chapter aims to examine convergence in the efficiency score.

prices, therefore once differences in the environment have been taken into account the remaining differences in costs will relate to operating efficiency. Portela et al (2011) use one input, opex, to measure productivity within the English and Welsh water and sewerage industry, and therefore state that their measure of efficiency relates to cost efficiency. Portela et al (2011) apply a single deflation index based on RPI, and highlight that their results may be biased if RPI does not reflect price rises for each component of opex, for example, the fast rise in electricity prices. Following Erbetta and Cave (2007) all costs apart from power costs are deflated using RPI to 2009 prices. Power is deflated by an energy price index for the industrial sector derived from the Department for Trade and Industry (DTI).

Two separate models are considered for water and sewerage activities; total and variable costs. Variable costs are calculated as total operating expenditure net of third party services, exceptional items, doubtful debts, service charge and local authority rates[15]. Total costs are calculated as variable costs and capital costs.

Outputs

Two outputs are incorporated within the model: firstly the water delivered (Y_1) and secondly, the equivalent population served (Y_2). Bottasso and Conti (2003) highlight the estimation gains for including not only the physical quantity of outputs but also the number of properties served for water and sewerage[16]. However as the estimation of β-convergence requires a separate

[15] These costs are deemed as non-controllable costs by Ofwat and are not incorporated within their assessment of efficiency. Exceptional items are by definition atypical. Third party services relate to costs incurred for output produced by other companies. Local authority rates and doubtful debts are considered as non-controllable. High levels of doubtful debts are due to the legal and regulatory decision of prohibiting the shutting off of water and sewerage activities when bills are not paid. Service charges are charges by the Environment Agency for water abstraction.

[16] The physical measure of output is applied instead of the number of properties billed, as sewerage quality is measured as the proportion of total load receiving sewerage treatment (BOD/Year). This is directly related to the measure of equivalent population served which is calculated based on the assumption that one person is equivalent to a load of 60g of BOD. This approach allows for a quality-adjusted measure of the physical output.

DEA model for each time period, a two output model is applied to avoid the problem of dimensionality.

Saal and Parker (2000) highlight the importance of the including changes in quality standard by the regulators within the measurement of efficiency. Large capital programmes have been undertaken since privatisation to improve quality standards within the industry. Following Saal and Parker (2000) a quality-adjusted measure of output is applied to account for changes in quality. The quality-adjusted measure of output implicitly makes the assumption that quality improvements require additional inputs. Water quality is regulated by the DWI. A quality index is defined as the ratio of the average percentage of each WaSCs water supply zones that are compliant with key water quality indicators defined by the DWI[17] (Q_w). Sewerage quality is accounted for by calculating the proportion of the total load receiving secondary treatment (Q_s). Adopting the approach of Saal and Parker (2000), by applying a quality-adjusted measure of output, the water delivered is multiplied by the water quality index $(Y_1 = Water\ Delieved \times Q_w)$ and sewerage is multiplied by the sewerage quality index $(Y_2 = Equivalent\ Population \times Q_S)$.

Ofwat considers the impact of other exogenous variables which may impact upon costs within their assessment of efficiency, such as the source of distribution input and the company's operating density[18]. The importance and impact of the inclusion of these additional variables is highlighted by Erbetta and Cave (2007), Saal et al (2007) and within the chapter 6 of this thesis. Additional operating characteristics have not been accounted for within the analysis of

---

[17] The average of several key indicators are considered; taste, odour, nitrate, aluminium, iron, lead and pesticides

[18] Portela et al (2011) extend the quality adjusted output approach to adjust for density. This approach is not applied within this study. Firstly, as output is adjusted for quality and secondly, the approach violates the translation invariance property of DEA and therefore biases the efficiency scores.

convergence due to the problem of dimensionality. Table 5.1 shows a snapshot of the data used in this chapter.

Table 5.1: Convergence Sample Descriptive Statistics

Variable		Mean	Std.Dev	Min	Max
Outputs					
Water Delivered	Ml/Day	1,014.6	551.9	284.2	2,179.4
Equivalent Population	(,000)	6,179.8	3,704.1	1,118.4	14,271.9
Inputs					
Total Costs	£m	648.9	303.9	220.6	1,346.7
Variable Costs	£m	241.7	124.8	76.9	577.2
Operating Characteristics					
Water Quality		0.967	0.026	0.836	0.995
Sewerage Quality		0.904	0.152	0.302	1

Note. 150 Observations. Costs and Input prices are expressed in real terms in 2009 prices

5.8 Results

5.8.1 Unconditional Convergence

This section outlines the results for the estimation of efficiency within the English and Welsh water and sewerage industry for total and variable costs. The results for the estimation of β- and σ-convergence have been estimated to examine whether companies' efficiency scores have converged over the period 1996/97–2010/11. Firstly, unconditional β-convergence is examined; the convergence to a common steady state within the industry. To account for regional heterogeneities, conditional β-convergence is examined through the estimation of fixed effects, random effects, GMM-DIFF and a GMM-SYS model. Finally, σ-convergence and a PAM are estimated to determine whether the dispersion of the efficiency scores has decreased and if the efficiency scores are persistent.

The yearly efficiency scores for total and variable costs are reported in table 5.2. The estimates reveal that an average inefficiency of 13% for total costs and 10% for variable costs respectively. An average of three companies for variable costs and four companies for total costs make up the frontier and are efficient. The minimum efficiency score over the period is 59.6% for total costs and 57.6% for variable costs. The majority of companies are more efficient for variable costs than the total costs models with the exception of WaSC2 which becomes less efficient.

Table 5.2- β-convergence efficiency scores

Total Cost Efficiency

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC1	0.751	0.737	0.729	0.778	0.689	0.779	0.792	0.759	0.753	0.776	0.796	0.766	0.761	0.738	0.754
WaSC2	0.718	0.684	0.636	0.691	0.678	0.691	0.67	0.707	0.719	0.738	0.733	0.709	0.707	0.795	0.788
WaSC3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC4	0.919	0.907	0.936	0.919	0.861	0.953	0.94	0.932	0.888	0.972	0.938	0.876	0.881	0.901	0.91
WaSC5	0.937	0.91	0.919	0.932	0.985	1	1	1	0.962	0.932	0.911	0.895	0.937	0.91	0.947
WaSC6	0.647	0.651	0.596	0.645	0.659	0.656	0.711	0.676	0.705	0.727	0.689	0.694	0.71	0.734	0.699
WaSC7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC8	0.740	0.787	0.777	0.74	0.845	0.865	0.857	0.846	0.813	0.828	0.839	0.790	0.747	0.722	0.69
WaSC9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC10	0.859	0.869	0.842	0.9	0.858	0.961	0.954	0.914	0.949	0.965	0.962	0.93	0.916	0.924	0.895
Average	0.857	0.855	0.843	0.861	0.858	0.890	0.892	0.884	0.879	0.894	0.887	0.866	0.866	0.873	0.868
Minimum	0.647	0.651	0.596	0.645	0.659	0.656	0.67	0.676	0.705	0.727	0.689	0.694	0.707	0.722	0.69
Std Deviation	0.133	0.133	0.153	0.135	0.14	0.135	0.127	0.128	0.121	0.114	0.116	0.12	0.123	0.115	0.125

Variable Cost Efficiency

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC1	0.799	0.819	0.778	0.808	0.673	0.798	0.839	0.81	0.777	0.8	0.842	0.873	0.837	0.824	0.812
WaSC2	0.682	0.639	0.576	0.629	0.597	0.629	0.644	0.658	0.673	0.746	0.755	0.742	0.725	0.771	0.741
WaSC3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC4	1	1	1	0.957	0.904	0.984	0.969	1	0.894	1	1	0.952	0.925	0.949	0.889
WaSC5	0.925	0.948	0.926	0.935	0.939	0.982	0.96	0.94	0.963	0.903	0.864	0.834	0.922	0.853	0.836
WaSC6	0.656	0.732	0.679	0.718	0.789	0.74	0.727	0.692	0.721	0.785	0.729	0.753	0.851	0.848	0.709
WaSC7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC8	1	1	0.963	0.893	1	1	0.986	0.973	0.877	0.968	0.997	0.936	0.923	0.898	0.856
WaSC9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WaSC10	0.849	0.905	0.87	0.963	0.86	0.995	1	1	1	1	1	1	1	1	0.93
Average	0.891	0.904	0.879	0.890	0.876	0.913	0.913	0.907	0.890	0.920	0.919	0.909	0.918	0.914	0.877
Minimum	0.656	0.639	0.576	0.629	0.597	0.629	0.644	0.658	0.673	0.746	0.729	0.742	0.725	0.771	0.709
Std Deviation	0.137	0.131	0.153	0.13	0.147	0.138	0.131	0.136	0.126	0.104	0.111	0.103	0.091	0.087	0.106

The evolution of efficiency scores and the standard deviation for total costs are depicted in figure 5.3.The mean efficiency score over the period has increased from 85.7% in 1997 to 86.8% in 2011. The minimum efficiency score has improved throughout the period from 64.7% to 69%. The standard deviation of the efficiency scores, depicted in figure 5.4 has fallen over the period from 0.133 and 0.125. The basic description of total cost efficiency would therefore indicate that there is limited convergence within the industry.

Figure 5.3: WaSC Total Costs Efficiency Score

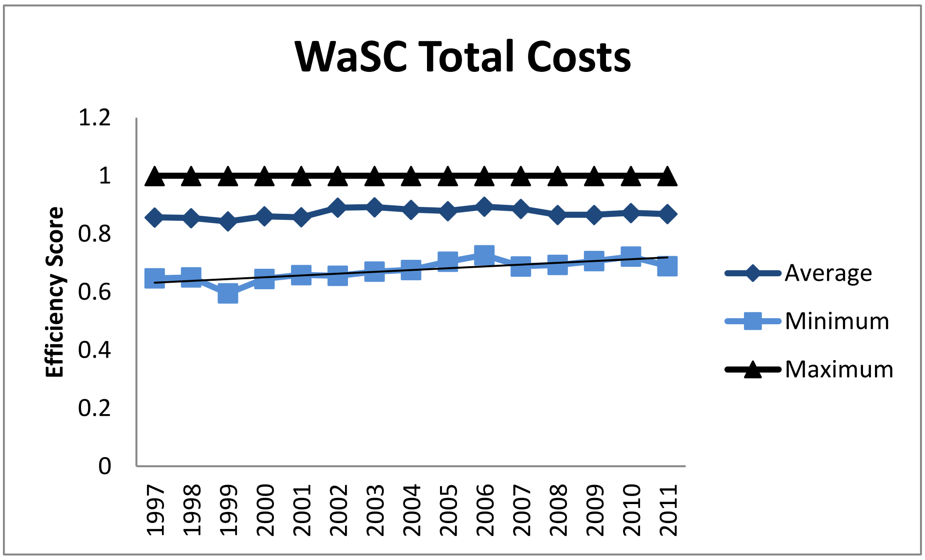
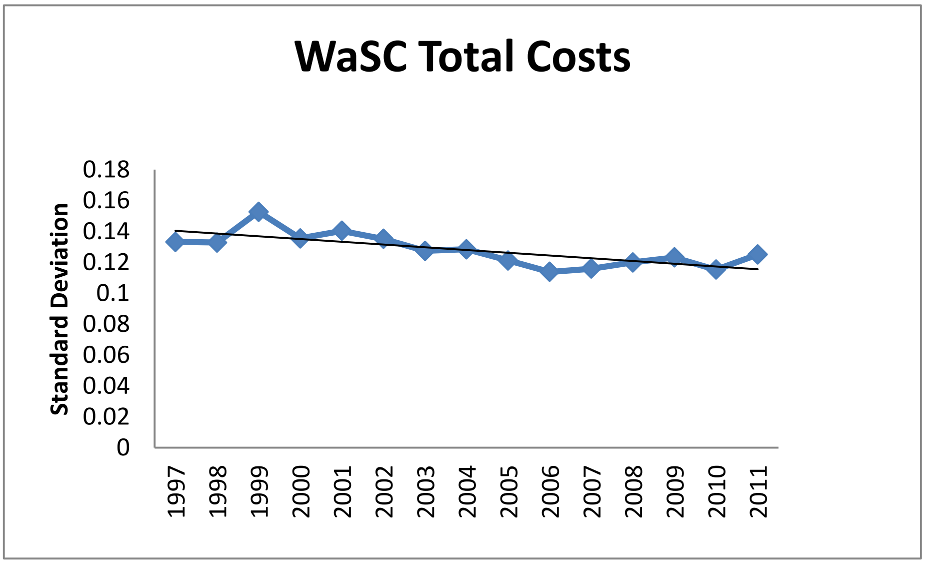


Figure 5.4: WaSC Total Costs Efficiency Score Standard Deviation



The evolution of the efficiency for variable cost efficiency is shown in figure 5.5. The mean efficiency score has decreased from 89.1% to 87.6% over the period. Although the average efficiency score appears to have fallen, there was a large drop in the efficiency score in 2011 from 77.1% in 2010 to 70.9% in 2011. The minimum efficiency score has increased from 65.6% to 70.9%. Overall the standard deviation of the efficiency scores has reduced over time from 0.14 to 0.11 which is depicted in figure 5.6.

Figure 5.5: WaSC Variable Costs Efficiency Score

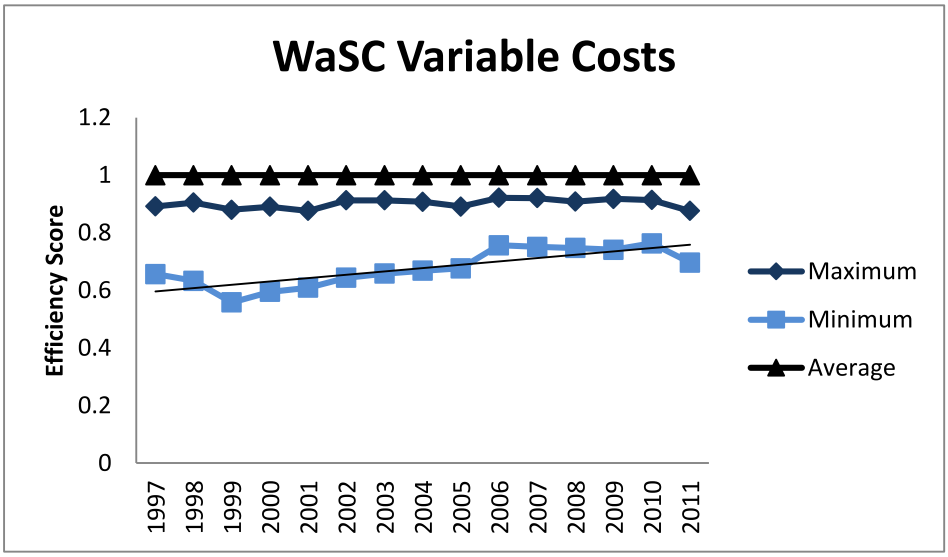
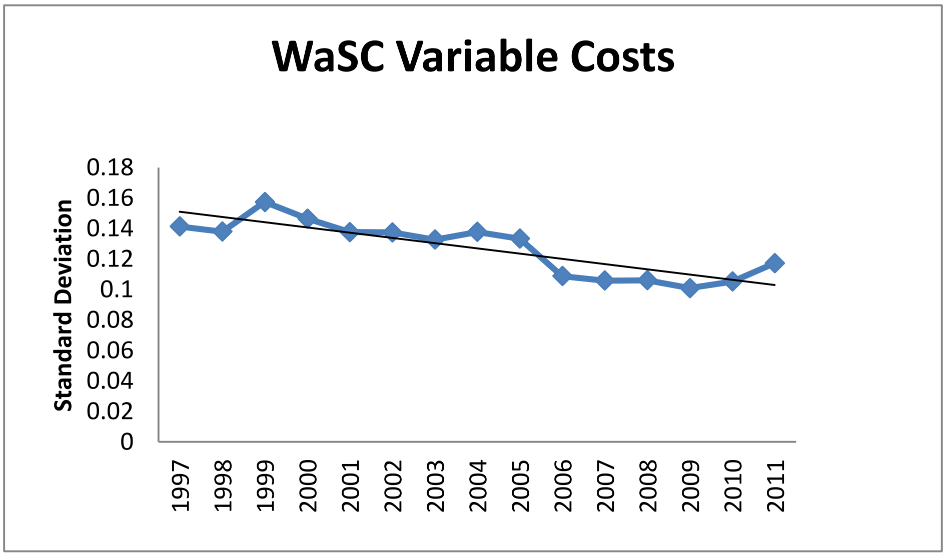


Figure 5.6: WaSC Variable Costs Efficiency Score Standard Deviation



The results for the estimation of unconditional convergence are shown in table 5.3. Examining convergence for total costs, the coefficient for β-convergence implies a rate of convergence of 4.2% a year which is significant at the 10% level, implying a half-life of 16.25 years with a steady state efficiency score of 88%. The coefficient for variable costs implies a rate of convergence of 8.8% a year which is significant at the 5% level, implying a half-life of 7.49 years. The rate of convergence is faster for variable costs than total costs. This might be due to the long adjustment time for the capital stock; firms may not be able to instantaneously adjust the capital stock to the optimal level. On the other hand this result may also be due to the presence of a capex bias. Stone and Webster (2004) and Bottasso and Conti (2009) report a presence of overcapitalisation which may be due to the Averch-Johnson effect. Overall the results indicate that the least efficient firms are catching up with the most efficient firms. On the other hand unconditional β-convergence ignores the presence of company-specific factors. The exclusion of these factors biases the efficiency scores downwards. We would expect company-specific factors to have a larger influence of total costs due to the different capital requirements in each operating area, this may explain the slower rate of convergence for total costs.

Ofwat aims to encourage convergence and technical change within the regulatory process through setting efficiency targets based upon the rate of technical change and a catch-up factor to the frontier. Firms are therefore expected to improve by a given percent which represents technological improvements and each firm is given a specific catch-up target based on closing a proportion of the efficiency gap between the company and the frontier. Opex efficiency targets for 1999 and 2004 were set based on closing 60% of the gap within the five years of the price review. This therefore implies catching-up to the frontier of 12% each year if all companies achieve the continuing efficiency target. The variable cost convergence model reports a rate of β-convergence of 8% a year. The results indicate that Ofwat has been successful in encouraging convergence within the variable costs and at similar rate implied by Ofwat within their price review.

Table 5.3: Unconditional β-convergence, σ-convergence and PAM

Unconditional Convergence

	Total Costs	Variable Costs
Beta-Convergence		
Intercept	-0.00513*	-0.01060***
	(0.00291)	(0.00395)
$\ln \theta_{t-1}$	-0.04174*	-0.08835**
	(0.02499)	(0.035035)
Steady State Efficiency	0.88	0.88
Half- life	16.25	7.49
R-Squared		
Number of Obs	140	140
Sigma-Convergence		
Intercept	-0.00022	-0.00038
	(0.00414)	(0.00389)
$\ln E_{t-1}$	-0.01381	-0.07709**
	(0.02670)	(0.02995)
R-Squared	0.0212	0.0514
Number of Obs	140	140
PAM		
Intercept	-0.00513*	-0.01060***
	(0.00291)	(0.00395)
$\ln \theta_{max} - \ln \theta_{t-1}$	0.04175*	0.08836**
	(0.02499)	(0.03504)
R-squared	0.0243	0.0556
Number of Obs	140	140

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.
Robust Standard errors

The results from β-convergence can be compared to those reported for the efficiency change[19]. Efficiency change is measured as θ_{t+1}/θ_t where a value greater than 1 implies that a firm's efficiency has improved while a value less than 1 implies that the efficiency score is decreasing. The geometric mean efficiency change for total and variable costs are reported in table 5.4. The efficiency change fluctuates throughout the period however the average of the whole period examined indicates a small and insignificant improvement in efficiency change. These results coincide with Portela et al (2011) whom report a negligible change in the average efficiency change of the WaSCs and WoCs over the period 1993/94–2006/07. The overall efficiency change indicates that there has been limited change in the average efficiency change however β-convergence indicates that the least efficient firms are growing at a faster rate than the most efficient firms. Therefore the price reviews have had a larger impact on those companies who were inefficient.

.

[19] Fare et al (1994) introduced a Malmquist index under VRS but Ray and Desli (1997) questioned the validity of the model to incorporate technical change alongside scale change. To avoid biases introduced with the VRS Malmquist index the efficiency change component has only been examined.

Table 5.4: Geometric Mean Efficiency Change

WaSC Geometric Mean Efficiency Change

	Total Costs	Variable Costs
1997/98	1.008	1.029
1998/99	0.981	0.956
1999/00	0.997	1.025
2000/01	1.006	0.942
2001/02	1.074	1.079
2002/03	1.002	1.012
2003/04	0.992	0.982
2004/05	0.983	0.967
2005/06	1.042	1.040
2006/07	0.996	0.985
2007/08	0.951	1.018
2008/09	0.989	1.035
2009/10	1.005	1.001
2010/11	1.004	0.948
Mean	1.002	1.001

Quah (1996) and Sala-i-Martin (1996) highlight several limitations of β-convergence and therefore suggest the estimation of σ-convergence. σ-convergence examines whether the dispersion of efficiency scores reduce over time. Alongside σ-convergence, a PAM model is estimated to determine whether companies' inefficiencies are persistent or whether they are converging towards the mean. Table 5.3 shows the estimation results for σ-convergence and the PAM. The results for σ-convergence indicate a negative coefficient, which is statistically significant for variable costs, however insignificant for total costs. The results indicate convergence for variable costs but there is a lack of σ-convergence for total costs. The coefficient for the PAM is positive and statistically significant at the 5% level for both variable costs and the 10% level for total costs. This result therefore concludes that companies are not only converging, but are also converging towards the frontier.

The results for total costs indicate a rate of β-convergence which is significant at the 10% level and insignificant σ-convergence. As previously discussed the lack of convergence may be as result of the slow adjustment time of capital and the presence of a capex bias. The lack of σ-convergence might be as a result of WaSC 6 water overtaking WaSC 2 and WaSC 8 within the last three years of the period. The least efficient company has improved at a faster rate, implying β -convergence. However, the least efficient company has overtaken other companies, implying a lack of σ-convergence.

5.8.2 Conditional Convergence

Unconditional β-convergence assumes that all firms converge to the same steady state efficiency level. Conditional β-convergence incorporates time specific heterogeneities through the incorporation of time invariant company specific factors. Instead of implying that all companies' convergence to the same steady state, the model implies that companies have differing steady state convergence efficiency scores depending on their characteristics. The FE model assumes that each company has its own steady state efficiency score.

Conditional β-convergence has been estimated using RE, FE, GMM-DIFF and GMM-SYS. The FE model and RE effects model are both estimated and reported in table 5.5. The results indicate that the rates of convergence are substantially different under the RE and FE models. The RE model implies a rate of convergence of 8.8% for total costs and 7.7% for variable costs. The estimates under RE are of a similar magnitude to those under OLS. The Breusch-Pagan test fails to reject the null hypothesis that the variance amongst entities is zero, therefore OLS is preferred.

The rate of convergence under FE is 46% a year for total costs and 54% for variable costs which are significant at the 1% level. This implies a half-life of 1.1 years for total costs and 0.89 years for variable costs. This rate of convergence is significantly higher than that reported for the pooled OLS. A Hausman test is applied to test whether the country effects are correlated with the exogenous variables. The result for the Hausman test rejects the null hypothesis and therefore indicates that RE produces biased and inconsistent estimates. An F-test is applied to examine the preference of OLS and FE and the null hypothesis is rejected which indicates that fixed effects are present and that the FE estimator is preferred.

Table 5.5: FE and RE β-convergence

	Total Costs		Variable Costs	
Beta-Convergence	FE	RE	FE	RE
Intercept	-0.06756***	-0.00513	-0.06129***	-0.01060*
	(0.01216)	(0.00470)	(0.00964)	(0.00563)
$\ln \theta_{t-1}$	-0.46381***	-0.04174**	-0.54270***	-0.08834***
	(0.07952)	(0.02250)	(0.07858)	(0.03099)
Steady State Efficiency				
Half- life	1.11	16.25	0.89	7.49
R-Squared				
Number of Observations	140	140	140	140
F-Test	3.34***		4.43***	
Hausman Test	30.62***		39.58***	
B-P Test	0		0	

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.

The steady state efficiency level for each WaSC is displayed in table 5.6. The presence of heterogeneities between the companies causes the convergence point to differ between the firms. The steady state efficiency scores for total costs range from 84% to 100%, indicating a

16 percentage points difference in the efficiency score convergence level. The variable cost steady state efficiency level ranges from 81% to 100%. The difference within the operating environment influences a firm's obtainable efficiency score by 19 percentage point. The results indicate that due to the company specific time invariant heterogeneities the point of convergence differs substantially between firms.

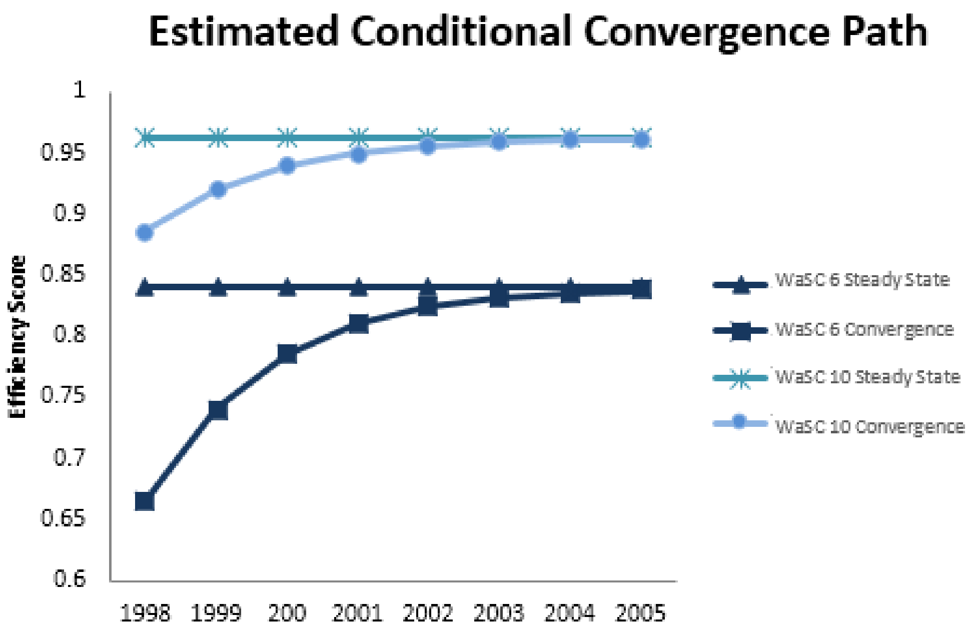
Table 5.6: FE Conditional β-convergence steady state efficiency score

	Total Cost Steady State Efficiency Score	Variable Cost Steady State Efficiency Score
WaSC 1	0.879	0.890
WaSC 2	0.856	0.812
WaSC 3	1	1
WaSC 4	0.959	0.973
WaSC 5	0.975	0.949
WaSC 6	0.839	0.855
WaSC 7	1	1
WaSC 8	0.896	0.966
WaSC 9	1	1
WaSC 10	0.962	0.984

Gluschenko (2012) highlights that the interpretation of the rate of convergence differs when applying fixed effects as the rate of convergence is no longer measured by the distance of the most inefficient firm to the steady state efficiency score. Conditional convergence observes the distance from the firm's own steady state efficiency score. Figure 5.7 plots the predicted convergence path for WaSC 6 and WaSC 10 estimated by FE. The graph depicts the estimated steady state efficiency score alongside the expected path of convergence for total costs given the rate of convergence for the total cost model. WaSC 6 efficiency score in 1997 is further away from its steady state efficiency score than WaSC 10, therefore the efficiency score grows at a faster rate.

This chapter measures conditional convergence through the application of FE. An alternative is to incorporate a set of environmental variables within the measurement of conditional β-convergence. Instead of applying the environmental variables within the measurement of convergence, environmental variables are incorporated to influence the efficiency scores in chapter 6. Environmental variables are incorporated within the measurement of efficiency as it is important to obtain a measure of efficiency once incorporating the differences into the operating environment.

Figure 5.7: Estimated FE Conditional Convergence Path



Nickell (1981) states that the efficiency scores under FE are biased upwards due to the presence of endogeneity between the lagged efficiency score and the fixed effects term. Bond et al (2001) state that a consistent estimate of b in equation 5.7 is expected to lie in between the OLS levels and FE estimate. To address the bias, a GMM-DIFF and GMM-SYS model is applied. The results from GMM-DIFF are displayed in table 5.7 whereby equation 5.7 is estimated therefore the dependent variable is θ_t, the rate of convergence is calculated as $\lambda = \beta - 1$.

A crucial assumption of the validity of GMM is that the instruments are exogenous. The Hansen test has the null hypothesis that the instruments as a group are exogenous. The null hypothesis for the Hansen test is not rejected, therefore implying that the instruments are exogenous. The Arellano-Bond test for autocorrelation has a null hypothesis of no autocorrelation and is applied to the differenced residual. There is first-order but no second-order autocorrelation, therefore the model specification is correct. It is important that second-order autocorrelation is not present as it will detect autocorrelation in levels.

The GMM-DIFF takes first differences to eliminate the fixed effects and instruments the first difference with the lagged levels. GMM-DIFF implies a rate of convergence of 45% for total costs and 59% for variable costs. The rate of convergence is significantly different to the rate implied by OLS but is within a similar range of that implied by the FE model. Blundell and Bond (1998) highlight that the lagged levels may be weak instruments and therefore propose the use of the SYS-GMM. Roodman (2009) shows that using too many estimates can produce biased results in GMM estimation. Roodman (2009) highlights that the minimum standard is to have the number of instruments lower than the number of individuals. GMM-SYS results are displayed within Appendix 2 as the number of instruments is greater than the number of groups of companies. The GMM-SYS model reports significant convergence at a rate of 11% for total costs and 12% for variable cost. However these results are incorporated as robustness check and do suffer from a lack of instruments.

Table 5.7: GMM-DIFF estimate

	Total Costs GMM-DIFF	Variable Costs GMM-DIFF
Intercept		
$\ln \theta_{t-1}$	0.5455493*** (0.1033028)	0.4086844* (0.1983891)
Convergence Speed	-0.4544507	-0.5913156
Number of Observations	140	140
Sargan Test	26.11**	34.79***
AR(1)	-2.42**	-1.97**
AR(2)	6.77	0.22
Half life	1.14	0.77

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.
Estimations are perfomed by xtabond2 in STATA by Roodman (2009)

5.8.3 Impact of Regulation

Up to now we have examined if the firms efficiency scores have converged over the period examined. The influence of the price reviews upon the rate of convergence can be examined through the inclusion of regulatory dummies. The impact of the 2004 price review on the rate of convergence is examined. It would be desirable to examine the influence of the 1999 price review, however due to the differencing in the GMM-DIFF this reduces our panel and therefore we are not able to distinguish the influence of the 1999 price review. To examine the influence of the 2004 price review we incorporate an interactive dummy variable ($D04$) which takes a value of 1 after the 2004 review. Equation 5.8 incorporates the influence of the 2004 price review within unconditional β-convergence.

$$\Delta\theta_{n,t} = \alpha + \beta(\ln \theta_{n,t-1}) + \delta(\ln \theta_{n,t-1} \times \mathrm{D04}) + \varepsilon_{n,t} \qquad (5.8)$$

A negative coefficient for 2004 dummy for OLS and FE would indicate that the 2004 price review has increased the rate of convergence.

Table 5.8: Convergence 2004 Regulatory Dummy

Convergence 2004 Regulatory Dummy

	Total Costs		Variable Costs	
Beta-Convergence	OLS	FE	OLS	FE
Intercept	-0.05257	-0.07454***	-0.01094***	-0.07573***
	(0.03460)	(0.01318)	(0.00391)	(0.01001)
$\ln \theta_{t-1}$	0.02821	-0.49053***	-0.07982*	-0.59456***
	(0.04115)	(0.08172)	(0.04661)	(0.07661)
$\ln \theta_{t-1} \times D04$	0.02821	-0.04372	-0.02583	-0.17289***
	(0.04115)	-0.03247	-0.06218	(0.04831)
R-Squared	0.0297	0.2197	0.0574	0.3363
Number of Observations	140	140	140	140
F-Test		3.58***		6.22***

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.

The results for variable and total costs for OLS and FE are reported in table 5.8 and for GMM-DIFF in table 5.9. Note that the dependent variable in GMM-DIFF is the log of the efficiency score in period t as in equation 5.9 instead of the growth rate of efficiency. Examining the results for total costs, the interactive dummy is insignificant for both unconditional and condition β-convergence, therefore indicating that the 2004 price review had no influence in changing the rate of convergence for total costs. The influence of the 2004 price review for variable costs is insignificant for unconditional β-convergence. The FE model indicates that the 2004 price review is significant at increasing the rate of convergence. However accounting for endogeneity applying the GMM-DIFF the influence is not significant. Overall the results indicate that the 2004 price review had no influence increasing the rate of β-convergence for total costs. The results for variable costs indicate that the 2004 price review had no influence on unconditional β-convergence and a limited influence for conditional β-convergence, although the results are not robust.

Figure 5.9: GMM-DIFF 2004 Regulatory Dummy

GMM-DIFF 2004 Regulatory Dummy

Beta-Convergence	Total Costs	Variable Costs
$\ln \theta_{t-1}$	0.47248* (0.25015)	0.33432* (0.16706)
$\ln \theta_{t-1} \times D04$	-0.02798 (0.10591)	-0.07983 (0.07397)
Number of Observations	130	130

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.

5.9 Conclusion

The regulatory framework encourages technical change and convergence within the industry through an industry efficiency challenge and a company specific catch-up factor. Convergence is encouraged through the measurement of relative efficiency and efficiency targets are based on closing a proportion of the gap to the frontier. This chapter aims to examine whether the regulatory policy has been effective in encouraging convergence in the efficiency scores towards the frontier.

The WaSC efficiency scores are estimated through the application of DEA for total and variable costs. To account for improvements in quality over the period a quality-adjusted measure of output is incorporated. This chapter draws upon the growth literature to estimate β- and σ-convergence to determine whether the least efficient firms are growing faster than the more efficient firms. It also examines whether the dispersion of efficiency scores has reduced. The results highlight the presence of unconditional β-convergence within the industry for both total and variable costs, with companies converging to the same steady state with the least efficient firms growing at a faster rate than the more efficient firms. The rate of convergence for total costs is only significant at the 10% level and the rate of convergence is slower than that of variable costs. The estimate of σ-convergence for variable costs implies a significant reduction in the dispersion of efficiency scores. The results for total costs imply a lack of σ-convergence. The lack of σ-convergence may be as a result of a capex bias, the slow adjustment rate of capital or not accounting for the differences in the companies' operating environments. Chapter 7 will incorporate capital as an intertemporal factor of production to account for capital long life and to examine the presence of a capex bias. The PAM indicates that the efficiency scores are converging towards the frontier.

The chapter aimed to examine whether Ofwat has been effective in encouraging convergence. The chapter reports convergence over the period, however we are unable to confirm whether convergence would have occurred in the absence of privatisation and the introduction of regulation. The chapter does find that the 2004 price review had no influence on the rate of convergence for total costs and limited evidence of a faster rate of convergence for variable costs. Although the chapter cannot identify whether convergence is a result of the introduction of regulation it does find that there is convergence within the industry with the rate of unconditional β-convergence for variable costs that is of a similar magnitude to that implied within the price review.

The chapter compares convergence to the average efficiency change within the industry. Efficiency change implies that there is a negligible amount of efficiency change over the period. These results coincide with those of Saal et al (2007) and Portela et al (2011) who examine efficiency change within the industry through a parametric assessment of water and sewerage activities and non-parametric assessment of water activities for WaSCs and WoCs respectively. This research therefore expands on the literature to find that although there is no change amongst the average firm efficiency change, the least efficient firms are growing at a faster rate than the most efficient companies.

Conditional β-convergence is examined to incorporate companies' heterogeneities through the estimation of fixed effects. The fixed effects model allows for companies to have their own steady state efficiency. The steady state convergence scores differ significantly between companies which highlights the significant differences within their operating environments. To address the issue of endogeneity within the fixed effects model, a GMM-DIFF estimator is applied. The rate of convergence when incorporating company specific heterogeneities is faster than estimated by unconditional convergence. The discovery of significant differences within companies' operating environments poses the following research question of how to incorporate these differences within the measurement of efficiency by DEA.

The following chapter investigates the influence of environmental variables upon firm's efficiency scores and the importance of controlling for differing operating environments within the measurement of efficiency using DEA. Environmental variables are assessed within the measurement of DEA to allow for the environmental variables to influence the frontier and therefore influence the efficiency score.

6. Incorporating the environment: three-stage DEA approach

6.1 Introduction

DEA is an effective tool utilised by firms and regulators to determine the level of managerial inefficiency. The basic DEA model measures the efficiency of a DMU based upon the effectiveness of managers to transform inputs into outputs. However, Avkiran and Rowlands (2008) highlight that the performance of an operational unit depends as much on managerial efficiency as on the operational environment and measurement noise. The methodology of DEA makes the implicit assumption that firms are homogeneous; assuming that firms undertake the same activities in the same operating environments (Golany and Roll, 1989). WaSCs operate under different operating characteristics which are outside of managerial control, known as non-discretionary or environmental variables. Non-discretionary variables can influence the production function, increasing or decreasing the maximum attainable output level if the environmental variables are favourable or unfavourable respectively. Failure to account for the differences in the operating environments may create biased results which can lead to unreliable economic decisions from management and the regulators.

Littlechild (1988) highlight that there are substantial differences between the ten RWAs (now WaSCs) with respect to size, condition of assets and environmental standards. Ofwat acknowledges differences within companies' operating environments by incorporating non-discretionary variables alongside key cost drivers. The operating characteristics of the firms vary substantially across the industry in terms of density, source of abstraction and topography. Quality within the industry is regulated by the DWI, NRW and the EA for water and sewerage activities, who have imposed increasingly strict compliance standards. Since privatisation the industry has invested £98 billion of capital investment to improve quality standards (Ofwat, 2012). Saal and Parker (2000) highlight the importance of incorporating quality standards for understanding cost differentials amongst firms and over time within the English and Welsh water and sewerage industry.

There are several common approaches in which environmental variables can be incorporated in DEA: quality-adjusted measure of output, one-stage approach and a two-stage approach. The one-stage model makes the distinction between controllable variables and non-controllable variables. The one-stage model suffers from the issue of dimensionality as the number of environmental variables increase. The two-stage approach estimates the efficiency of the controllable inputs and applies a second stage regression to identify the influence of non-discretionary variables on the efficiency scores or inputs slacks. This approach identifies whether the environmental variables are favourable or unfavourable, increasing or decreasing the efficiency score. Within the estimation of DEA for the English and Welsh water and sewerage industry, Maziotis et al (2013) and Erbetta and Cave (2007) incorporate quality variables through a quality-adjusted measure of output. Erbetta and Cave (2007) also examine the influence of additional non-discretionary variables through a second stage SFA regression.

The two-stage approach does not calculate the efficiency score once controlling for environmental variables[20]. This chapter extends upon the current literature for the measurement of efficiency within the English and Welsh water and sewerage industry by applying a three-stage approach introduced by Fried et al (1999, 2002) extended by Tone and Tsutsui (2009) and Cordero-Ferrera et al (2010). The methodology enables the measurement of efficiency whilst controlling for the differences in firms' operating environments. The three-stage

[20] The error term of the second stage regression can be interpreted as the level of inefficiency after controlling for environmental variables. However the error term has a two-sided distribution and a true measure of efficiency should have a one-sided distribution.

approach introduced by Fried et al (1999, 2002) extends upon the two-stage model by adjusting inputs, depending on whether firms operate within favourable or unfavourable environments. The final stage re-runs the DEA problem to obtain environmentally adjusted efficiency scores. Cordero-Ferrera et al (2010) apply a bootstrap truncated second stage regression proposed by Simar and Wilson (2007) to correct for the presence of serial correlation amongst the DEA efficiency scores. Tone and Tsutsui (2009) highlight that the adjustment procedure of Fried et al (1999, 2002) violates the translation invariance property, therefore biases the efficiency scores. Their paper introduces an alternative procedure for adjusting the input variables. As well as highlighting the importance of incorporating environmental variables, the chapter also examines the impact of the price reviews. It also examines the question of whether different price reviews have been effective in improving efficiency or reducing input slacks. The impact of privatisation and the 1994 price review has been examined by Saal and Parker (2000), Saal and Parker (2001) and Saal et al (2007). Saal and Reid (2004), Erbetta and Cave (2007) and Bottasso and Conti (2009) examine the impact of the 1994 and 1999 price reviews. This chapter extends upon current research to examine the impact of both the 1999 and 2004 price reviews upon efficiency.

The chapter will firstly review the relevant literature for the incorporation of environmental variables within efficiency estimates for the English and Welsh water and sewerage industry and the empirical literature of the impact of privatisation and regulation. The different methodologies for incorporating environment variables are outlined and the proposed methodology. The data and variables used within the model are outlined followed by the results.

6.2 Incorporating non-discretionary variables

The efficiency of a DMU measures a manager’s ability to transform inputs into outputs. The ability to do so depends on managerial efficiency and other factors outside of managerial control. The theoretical importance of controlling for non-discretionary variables within the measurement of efficiency is outlined within the methodology chapter. The ignorance of the non-discretionary variables can bias the measurement of efficiency depending upon whether firms operate within a favourable or unfavourable environment (Ray, 1988). As DEA assumes homogeneity it is important to ensure that non-discretionary variables which influence the production function are incorporated within the analysis. The following section outlines several approaches to incorporate non-discretionary variables in the measurement of efficiency through DEA.

Environmental variables have been included within DEA for the measurement of efficiency through several different methodologies. Environmental variables are defined as those factors outside of the control of management such as population density. Several authors[21] incorporate the proportion of leakage and the measurement of water and sewerage quality as non-discretionary variables. Although leakage and quality can be influenced by management, these are incorporated as non-discretionary variables due to the leakage targets imposed by Ofwat and quality standards imposed by the DWI and EA. The following section will outline the methodologies employed for the incorporation of non-discretionary variables for the measurement of efficiency within the English and Welsh water and sewerage industry and international water utilities.

[21] Saal and Parker (2000,2001), Saal and Reid (2004), Saal and Parker (2005), Saal et al (2007), Erbetta and Cave (2007), Bottasso and Conti (2009) and Saal et al (2011).

6.2.1 Quality-Adjusted Output

Saal and Parker (2000) was the first study to account for changes in quality in the English and Welsh water and sewerage through a quality-adjusted measure of output for the estimation of a translog cost function to measure productivity within the English and Welsh water and sewerage industry. The quality-adjusted measure of output (Y_Q) is calculated by multiplying the level of output (Y) by a quality index (Q) which takes a value between 0 and 1 as in equation (6.1)

$$Y_Q = Y \times Q \qquad (6.1)$$

The application of a quality-adjusted measure of output implicitly makes the assumption that an improvement in the measure of quality requires additional inputs. An increase in quality increases the quality-adjusted measure of output, therefore holding efficiency constant an increase in output requires additional inputs. Using a quality-adjusted measure of output requires the interpretation that firms desire to produce the output at a given quality. This is a reasonable assumption within the English and Welsh water and sewerage industry due to minimum quality standards. Saal and Parker (2000) highlight the importance of adjusting for changes in quality standards through the impact on the interaction terms between water and sewerage activities, finding an improvement in the quality of one output may reduce the cost of producing the other. The quality-adjusted measure of output is applied within the English and Welsh water and sewerage industry by Saal and Parker (2001) and Saal et al (2007) within the parametric approach. Maziotis et al (2013) and Erbetta and Cave (2007) apply this within the non-parametric approach. Portela et al (2011) extend the approach to account for the differences in density by including an adjusted measure of billed properties alongside the output variable for billed properties. The measure of billed properties is multiplied by an adjustment factor. The adjustment factor is based upon the residual from the regression of the number of billed properties on the length of main[22]. To ensure the adjustment factor is positive a constant is added[23].

The incorporation of environmental variables through an adjusted measure of output has the advantage of allowing for quality to be incorporated directly within the frontier whilst not requiring additional constraints within the DEA problem. Additional constraints within the DEA problem can lead to dimensionality problems and therefore reduce the discriminatory power of the model. On the other hand, the methodology has the disadvantage of assuming the directional impact of the environmental variables a priori. The methodology only allows for the inclusion of several environmental variables and those which impact the output variables under input-orientation. The incorporation of non-quality variables can lead to difficulties in the interpretation.

6.2.2 One-Stage Approach

Within the parametric methodology environmental variables can be incorporated directly within the frontier to determine the impact of environmental variables upon efficiency. Environmental variables can be incorporated on the right hand side of the cost function (input distance function) to influence costs (inputs). Under output-orientation, environmental variables can be incorporated on the right hand side to influence the level of outputs. This methodology allows for the influence and significance of the environmental variables upon

[22] Adjusted billed properties= billed properties*(standardized regression residual+ constant). Standardized residual is the raw residual divided by the estimated standard error of the residual. A positive residual indicates that the company has more billed properties than expected, therefore operate within a highly dense operating area.
[23] This approach violates the translation invariance property however Portela et al (2011) find that the comparison of the results with and without the adjusted output for density is negligible except for one highly urbanised company.

costs (inputs) to be determined. This approach is applied by Bottasso and Conti (2009), Saal and Reid (2004), Saal and Parker (2005) and Saal et al (2007) for the estimation of a translog cost function and Saal et al (2011) who estimate a quadratic cost function. The translog cost function is shown in equation (6.2) where TC represents total costs, W the input prices, Y the output variables and Z is a vector of environmental variables which are included as non-interactive variables within the cost function.

$$\ln TC = \delta + \sum_{j} \alpha_j \ln W_j + \sum_{r} \chi_r \ln Y_r + \frac{1}{2}\sum_{j}\sum_{v} \gamma_{\mathrm{jv}} \ln W_j \ln W_v + \sum_{j}\sum_{r} \kappa_{jr} \ln W_j \ln Y_r + \frac{1}{2}\sum_{r}\sum_{z} \xi_{r,z} \ln Y_r \ln Y_z + \sum_{j} \psi_j \ln Z_j \quad (6.2)$$

Saal et al (2007) highlight the advantage of incorporating environmental variables as fully interactive variables to allow for the environmental variables to influence the measurement of both the cost function and the measurement of productivity. However, the incorporation of the environmental variables as fully interactive variables significantly increases the number of parameters to be estimated, and therefore decreases the degrees of freedom.

Environmental variables can be incorporated directly within the DEA frontier as either input or output variables. The measurement of efficiency under input (output) orientation examines the radial contraction (expansion) of all inputs (outputs). The assumption of radial contraction or expansion for non-discretionary variables makes little sense as by definition managers do not have full control over these variables (Fried et al, 1999). To overcome the problem of radial contraction (expansion) of inputs (outputs), Banker and Morey (1986) hold the level of the environmental variables fixed and non-discretionary variables are incorporated as uncontrollable variables. Uncontrollable variables enter the DEA model to influence the frontier, however, are held constant for the calculation of radial efficiency.

The inclusion of an additional input enables the production of additional outputs holding efficiency constant. Including an environment variable (z) as an input in equation (6.3) implies more outputs can be produced; therefore the environmental variable is assumed to be favourable.

$$\begin{aligned} & min_{\theta,\lambda}\, \theta \\ st \quad & -y_n + Y\lambda \geq 0 \qquad (6.3) \\ & \theta x_n - X\lambda \geq 0 \\ & z_n \geq Z\lambda \\ & \lambda \geq 0 \end{aligned}$$

On the other hand, environmental variables can be incorporated as an output variable. The inclusion of additional outputs requires additional inputs holding efficiency constant. If the environmental variables are included as output variables in equation (6.4), additional inputs are required, therefore the environmental variable is assumed to be unfavourable.

$$min_{\theta,\lambda}\ \theta$$

$$st \quad -y_n + Y\lambda \geq 0 \qquad (6.4)$$

$$\theta x_n - X\lambda \geq 0$$

$$z_n \leq Z\lambda$$

$$\lambda \geq 0$$

If one is unsure of the directional impact of the environmental variables, then the environmental variables can be included in an equality form in equation (6.5). The inclusion of the environment variables using an equality form ensures that firms are only compared with a (theoretical) frontier that has the same environment. This approach is advantageous as it does not require the directional impact of the environmental variables to be assumed a priori. However the approach substantially reduces the reference set therefore biases the efficiency scores upwards (Coelli et al, 2005).

$$min_{\theta,\lambda}\ \theta$$

$$st \quad -y_n + Y\lambda \geq 0 \qquad (6.5)$$

$$\theta x_n - X\lambda \geq 0$$

$$z_n = Z\lambda$$

$$\lambda \geq 0$$

Picazo-Tadeo et al (2008) apply the one-stage DEA model to analyse the importance of including quality within the Spanish water utilities. Quality is incorporated within the frontier as a non-controllable variable. The model makes two assumptions; firstly that a lack of quality is a bad output and secondly that there is a trade-off between quantity and quality. The efficiency scores are compared for the conventional DEA and for the quality-adjusted DEA for Spanish water utilities. Their results indicate that there is a trade-off between quality and quantity. The maximum attainable quantity firms could potentially reach is reduced when accounting for quality. The paper compares the efficiency scores before and after adjusting for quality and performs several tests; t-test, Wilcoxon test and the Spearman's rank test. The results reveal that the technical efficiency scores and their distribution are statistically different when controlling for quality, however their ranks do not statistically change.

The all-in-one method is advantageous as it allows for the environmental variables to impact directly upon the frontier. If several environmental variables are considered within the analysis the additional constraints can reduce the dimensionality of the model which reduces its discriminatory power. Another drawback of incorporating quality directly within the frontier is that it makes the implicit assumption of the directional impact of the environmental variables. This approach is not suitable if the researcher wants to determine whether environmental variables are favourable or unfavourable, such as the impact of regulatory regimes.

The drawbacks of the one-stage approach within the non-parametric framework has led to the evaluation of the significance and directional impact of the environmental variables upon the firm's efficiency scores through the application of a second stage regression.

6.2.3 Two-Stage Approach

To overcome the problems of the one-stage model, Ray (1988) introduces a two-stage model to determine the directional impact of the environmental variables by regressing the efficiency scores on a set of environmental variables. The second stage regression aims to explain differences in the efficiency scores by features of the operating environment (Fried et al, 1999). The second stage regression is shown in equation 6.6 where θ is a vector of efficiency scores for N firms. The Farrell efficiency score takes a value between 0 and 1, where a value of 1 implies the firm is technically efficient. Z is a vector of environmental variables and β is a vector of coefficients to be estimated and ε is the error term. The estimated β coefficients indicate the directional impact of the environmental variables upon the efficiency score. A positive coefficient indicates that the environmental variables impact positively on the efficiency score and therefore operates under a favourable environment. On the other hand, a negative coefficient implies that the environmental variable reduces the efficiency score, and therefore the operating environment is unfavourable.

$$\theta_n = z_n\beta + \varepsilon_n \quad (6.6)$$

The approach is advantageous as it can incorporate categorical and continuous variables whilst allowing several variables to be incorporated simultaneously. However, the use of a second stage regression implies the separability condition $y_t = h(z_t) \times f(x_t)$ for the production function (Ray, 1988). The separability condition implies that the non-discretionary variables act like Hicks neutral technical change. Therefore, the environmental variables impact on the maximum attainable outputs, but do not affect the optimal choices between the discretionary inputs. For the second stage regression of the environmental variables on the efficiency scores, the separability condition implies that the environment variables cannot affect the support of the input and output variables within the first stage DEA problem. The environmental variables affect the distribution of inefficiencies, affecting their mean and variance.

The second stage regression has been estimated by OLS by Ray (1991) and Sexton et al (1994) and Stanton (2002). The dependent variable, the Farrell efficiency score, is bounded between 0 and 1, therefore several authors (Anwandter and Ozuna, 2002; Tupper and Resende, 2004 and Garcia-Sanchez, 2006) use a Tobit regression. The Tobit regression is distributed according to the normal distribution but its likelihood function reflects the fact that the dependent variable is bounded between 0 and 1.

The second stage regression allows the directional impact and the significance of the environmental variable upon efficiency to be realised. Anwandter and Ozuna (2002) apply a second stage Tobit regression to examine the impact of environmental variables within the Mexican Water utilities. Alongside the environmental variables Anwandter and Ozuna (2002) include a series of dummy variables to determine whether public sector reforms in the Mexican water utilities have improved efficiency. Regulatory dummies within the second stage regression are insignificant, therefore indicating that the public sector reforms had no impact upon the efficiency scores.

The error term of the second stage regression can be interpreted as the efficiency score once controlling for environmental variables. However, this is a two-sided distribution and a true measure of efficiency score has a one-sided distribution. Tupper and Resende (2004) examine the impact of environmental variables for the Brazilian water and sewerage industry. The paper applies an adjustment procedure to ensure that the efficiency scores are bounded between 0 and 1. The residuals are adjusted by the adjustment process in equation (6.7), where ε_n is the residual from the Tobit regression and θ_n^{ADJ} is the adjusted efficiency score for firm n.

$$\theta_n^{ADJ} = \varepsilon_n + \left(1 - \max_{\mathrm{n=1,\ldots,N}} \varepsilon_n\right) \quad (6.7)$$

The adjusted efficiency score can be interpreted as the efficiency score once controlling for the environmental variables. However, this methodology suffers from the problem associated with COLS, whereby the parallel shift makes the assumption that the structure of the production technology is the same for the industry average and the frontier.

Fried et al (1999) extend on the two-stage approach by regressing total radial and non-radial input slacks instead of the efficiency scores on the environmental variables. The total radial and non-radial input slacks represent the excess level of inputs, where the slacks are zero if the firm is efficient and positive for the excess utilisation of inputs. The interpretation of the second stage coefficients varies whereby a positive coefficient indicates that the environmental variables impacts positively on the input slacks, which indicates that the environmental variable is unfavourable. On the other hand a negative coefficients indicates that the environmental variables are favourable, which reduces the input slacks. Fried et al (1999) estimate separate regressions for each input slack, therefore allowing the impact of the environmental variables to differ between inputs.

One of the downfalls of DEA is that the model does not allow for the incorporation of noise. To allow for the decomposition of the DEA efficiency score into managerial efficiency, environmental effects and statistical noise, Fried et al (2002) employ a second stage SFA model for each j input which takes the form.

$$s_{nj} = f^j\left(z_n; \beta^j\right) + u_{nj} + v_{nj} \quad (6.8)$$

Where s_{nj}denotes the slacks for input $j, j = 1,,.,J$ and firm $n, n = 1,,.,N$ and $f^j\left(z_n; \beta^j\right)$ is the deterministic feasible slack frontier, z is the vector of environmental variables and β is the vector of coefficients to be estimated. SFA allows for the error term to be decomposed into two components $v_{nj} \sim N(0, \sigma_{vn}^2)$ which is a two sided normal distributed error reflecting statistical noise and a one-sided error term $u_{nj} \geq 0$ reflecting managerial efficiency. The researcher must make distributional assumptions with regards to the error term representing inefficiency; common distributions are half-normal, truncated normal and gamma distribution.

Erbetta and Cave (2007) measure cost efficiency for the WaSCs in the English and Welsh water and sewerage industry using DEA. The paper examines the impact of environmental variables through a quality-adjusted measure of output (Saal and Parker, 2000) and a second stage SFA regression based upon the methodology of Fried et al (2002). The first stage is to calculate the efficiency score and obtain the technical efficiency slacks and the absolute value of the allocative inefficiency slacks[24]. The absolute value of the allocative inefficiency slack therefore does not represent an over- or under-utilisation of inputs but rather a mere distortion. The second stage regresses the slacks for each input relating to technical and allocative efficiency separately using a SFA model in equation (6.8). Separate regressions are applied for each input to allow for the impact of the environmental variables to differ for each input. The coefficient on the environmental variables indicates the directional impact of the environmental variables on the excessive input slacks. A positive coefficient indicates a positive relationship between the environmental variables and excess input slacks, indicating that the environmental variable is unfavourable. Erbetta and Cave (2007) SFA results report that the one-sided managerial inefficiency term dominates noise. Therefore, once accounting for differences in the operational characteristics and the presence of noise, firms still utilise inputs differently. The technical over-utilisation of inputs and allocative distortion is a function of the technical efficiency error term u_{nj}. Erbetta and Cave (2007) model technical inefficiency and the absolute value of allocative inefficiency. As the absolute value of allocative efficiency is taken,

[24] Allocative efficiency represents the over or under-utilization of inputs measured as $TE_n x_n - x_n^*$, where TE_n is the technical efficiency score and x_n^* is the optimal inputs under cost efficiency. This measure is negative, positive or zero for underutilisation, overutilization and optimal use respectively. The absolute value of the allocative efficiency represents the mere distortion.

this approach does not allow for the overall measurement of cost efficiency score once accounting for the differences in the operating environment.

Although Tobit, OLS and SFA regressions are common within the literature Simar and Wilson (2007) state that the inference obtained from the second stage regression is invalid due to complicated, unknown serial correlation among the efficiency estimates. The actual efficiency score θ_n in equation 6.6 is unknown and is therefore estimated through DEA using the observed data pairs (x_n, y_n) to estimate θ_n for all $n = 1, \ldots, N$ yielding a set of estimates $\{\hat{\theta}_n\}_{n=1}^{N}$.

The actual technical efficiency in the second stage regression is replaced by the estimated $\hat{\theta}_n$.

$$\hat{\theta}_n = z_n\hat{\beta} + \xi_n \geq 1 \quad (6.9)$$

Simar and Wilson (2007) note that $\hat{\theta}_n$ are serially correlated in a complicated and unknown way. The estimation of technical efficiency depends on all the observations and consequently so must the error term ξ_n in the second stage regression. Moreover x_n and y_n are correlated with z_n, the environmental variables; otherwise there would be no motivation for the second stage regression. Hence this implies that the error term ξ_n is correlated with z_n. To overcome the problem of serial correlation Simar and Wilson (2007) denote two algorithms; algorithm #1 allows the correct standard errors and algorithm #2 additionally corrects for the bias of the DEA estimates. As the inverse of the Farrell efficiency scores are bounded at 1. Algorithm #1 applies a bootstrapped truncated regression with left truncation at 1 to obtain correct standard errors.

Renzetti and Dupont (2009) measure the technical efficiency of water municipals in Canada and evaluate the impact of environmental variables. The paper computes technical efficiency through a three-stage DEA approach proposed by Fried et al (1999). The first stage evaluates the efficiency scores applying DEA whilst the second stage evaluates the influence of the environmental variables through a second stage regression. Their work compares the use of a second stage Tobit regression and the bootstrapped truncated regression algorithm #1 proposed by Simar and Wilson (2007) to correct for the presence of serial correlation. Their results indicate that the coefficients for the bootstrapped truncated regression and tobit regression are similar, however the bootstrapped truncated regression has a higher number of coefficients which are significant.

The contribution of Simar and Wilson (2007) allows for the estimation of a second stage regression with correct inference. However, the second stage regression only allows for the directional impact and significance of the environmental variables to be analysed. To allow for the incorporation of the impact environmental variables estimated in the second stage regressions, Renzetti and Dupont (2009) and Cordero-Ferrera et al (2010) apply a 3rd stage introduced by Fried et al (1999,2002) to adjust the input data relative to whether firms operate within a favourable or unfavourable environment.

6.2.4 Three-Stage Approach

The three stage approach introduced by Fried et al (1999, 2002) allows for the impact of the environmental variables to be reflected within the DEA scores. The efficiency scores and input slacks are estimated within the first stage by DEA. The impact of the environmental variables is examined through a second stage regression for each input slack in equation (6.10). s_{nj} is the total input slacks for input $j = 1,..,J$ and firm $n = 1,..,N$ and Z is a vector of environmental variables and v is a iid error term.

$$s_{nj} = f^j(z_n; \beta^j) + v_{nj} \quad (6.10)$$

The third stage adjusts the firm's inputs x_{nj} by the amount in which they operate under a favourable or unfavourable environment based upon the predicted slack $z_n\hat{\beta}^j$ from the second stage regression following Fried et al (1999) in equation 6.11. One can adjust the inputs downwards for those firms which operate under unfavourable conditions. The amount by which they have been disadvantaged is revealed within the second stage regression. If a firm is significantly disadvantaged it is possible that the inputs may be adjusted so far downwards that they may become negative. Therefore, an alternative approach is to adjust the input variables upwards for those firms operating under relatively favourable environments. Inputs are adjusted upwards relative to the firms operating under the most unfavourable environment. $max_n\{z_n\hat{\beta}^j\}$ is the most unfavourable operating environment as they have the highest predicted slacks. The firms with a relatively unfavourable operating environment have their inputs adjusted by a relatively small amount, whilst those with favourable operating environments are adjusted upwards by a relatively large amount. Adjusting the inputs upwards provides a performance target that managers can reach regardless of their operating environment (Fried et al, 1999).

$$x_{ni}^A = x_{ni} + \left[max_n\{z_n\hat{\beta}^j\} - z_n\hat{\beta}^j\right] \quad (6.11)$$

The DEA model is re-run with the adjusted input variables allowing for the impact of the environmental variables to influence the shape of the frontier and to obtain an environmental adjusted DEA score.

Fried et al (1999) apply a second stage OLS regression whilst Fried et al (2002) apply a second SFA model to account for noise. However, neither second stage regression, OLS or SFA model address the issue of serial correlation highlighted by Simar and Wilson (2007). Cordero-

Ferrera et al (2010) combine the two approaches of Simar and Wilson (2007) and Fried et al (1999) to allow for both the correct inference and environmental adjusted DEA scores.

To allow for the correct inference for the second-stage regression in the three-stage approach, Cordero-Ferrera et al (2010) apply the second stage bootstrap truncated regression introduced by Simar and Wilson (2007). However, the dependent variable is the radial and non-radial input slacks instead of the efficiency scores. The inverse of the Farrell efficiency score is bounded at 1, whereas the input slacks are bounded at 0. Therefore Cordero-Ferrera et al (2010) apply a bootstrap truncated regression with left truncation at 0 instead of left truncation at 1 which is applied for the efficiency scores. Cordero-Ferrera et al (2010) then follow the adjustment procedure of Fried et al (1999) to obtain environmentally adjusted DEA scores.

Tone and Tsutsui (2009) demonstrate that the adjustment procedure of Fried et al (1999, 2002) violates the translation variance property of DEA due to the addition of the constant which relates to the company operating under the most unfavourable conditions. The translation invariance property is that the efficiency score is biased if a constant is added to each DMU for an input (output) under an input (output) orientation model. Tone and Tsutsui (2009) demonstrate that the addition of the constant skews the distribution and recommend an alternative adjustment process which is discussed in the methodology. This chapter applies the methodology of Fried et al (1999) extended by Cordero-Ferrera et al (2010) and Tone and Tsutsui (2009) to obtain efficiency scores once controlling for differences in the operating environment. A detail outline of the approach is provided within the methodology.

6.3 Privatisation and Regulation

The second component of the chapter is to examine the impact of regulation upon efficiency. The English and Welsh water and sewerage industry was privatised in 1989 and subject to price-cap regulation. The price caps are determined every five years through Price Reviews (PR), the first price review was undertaken by the government in 1989 and subsequent reviews (from 1994 onwards) were undertaken by Ofwat. Two key aims of privatisation and regulation are to improve economic efficiency and to encourage investment to improve quality. A body of literature has emerged to determine whether privatisation and regulation has been an effective tool for encouraging efficiency improvements. The literature for the measurement of productivity and efficiency was outlined in chapter 5. The following section will provide a brief review of the literature examining the influence of the difference price reviews.

Overall, the empirical literature concludes that privatisation has had little or no impact on efficiency and productivity and that the industry has experienced improvements in productivity due to the imposition of price-cap regulation. The studies examining privatisation and the 1994 price review find mixed results. Saal and Parker (2000) report that privatisation had no impact on productivity whereas the 1994 price review reports significant improvements. On the other hand Saal and Parker (2001) and Saal et al (2007) report no statistically significant difference due to the 1994 price review. However, when examining the 1994 and 1999 price review the literature Erbetta and Cave, 2007 and Bottasso and Conti, 2009 conclude that the 1999 price review brought around significant improvements in productivity and efficiency in comparison to the 1994 review, with the exception of Saal and Reid (2004). Portela et al (2011) is the only study known to examine the 2004 price review and state that that review may have impacted negatively upon the efficiency scores.

This chapter acknowledges the differences in the operating conditions when measuring efficiency and also contributes to the existing literature about the impact of different regulatory regimes by determining whether the 1999 and 2004 price review had a significant impact on reducing input slacks.

6.4 Methodology

Incorporating non-discretionary variables within the measurement efficiency using DEA is based upon the methodology of Fried et al (1999), Simar and Wilson (2007), Cordero-Ferrera et al (2010) and Tone and Tsutsui (2009). Firstly, the methodology of measuring efficiency through the application of DEA is outlined. Secondly, a second-stage bootstrapped truncated regression is applied to evaluate the impact of the environmental variables upon the efficiency scores. Finally, the adjustment procedure of Tone and Tsutsui (2009) is examined to adjust the input slacks by the amount in which firms operate under a favourable environment.

The efficiency scores are measured using DEA under VRS and input-orientation as output is demand derived and managers have limited control over their operating size as described within the methodology chapter. The BCC DEA linear programming problem to measure technical efficiency is denoted in equation 6.12.

Here N $(n = 1, \dots, N)$ denotes the number of DMUs, x_j the j-th input $(j = 1, \dots, J)$and y_r the r-th output $(r = 1, .., R)$, λ is the vector of weights and θ is the radial contraction of inputs to the convex frontier; the efficiency scores.

$$min_{\theta,\lambda}\ \theta$$

$$st \quad -y_n + Y\lambda \geq 0 \qquad (6.12)$$

$$\theta x_n - X\lambda \geq 0$$

$$\lambda \geq 0$$

$$I\lambda' = 1$$

The efficiency is examined over the period 1996/97–2010/11 in which the data is pooled and is examined under a meta-frontier, described in section 3.6.5.9. The meta-frontier assumes the construction of a single frontier, therefore implying no technological change over the period. A meta-frontier is examined as the number of DMUs within the sample is limited, therefore to avoid the dimensionality problem the data is pooled over the period. Although the meta-frontier does not assume technical change, the impact of technical change and regulatory regimes is examined within the second stage regression. An alternative approach is the application of a moving average window analysis, however distinguishing the impact of regulatory regimes is difficult as the frontier is changing between the periods which will include the influence of regulation.

The DEA linear programme measures the Farrell efficiency score θ which takes a value between 0 and 1. A value of 1 indicates that the firm is efficient and therefore is on the frontier, while a value less than 1 indicates that the firm is inefficient. The amount in which management should reduce inputs by to be efficient is calculated as $(1-\theta)x$, the radial slacks. As outlined in the methodology chapter this ignores the presence of any non-radial slacks. Fried et al (1999, 2002), Cordero-Ferrera et al (2010) and Tone and Tsutusi (2009) use total radial and non-radial excess inputs slacks within the second stage regression which are measured as

$$s_{nj} = x_{nj} - X_n\lambda \geq 0, n = 1, \dots, N, j = 1, \dots, J \quad (6.13)$$

Cordero-Ferrera et al (2010) highlight the importance of controlling for non-radial slacks due to the additional information obtained which can be extremely useful when accounting for the potential sources of production inefficiencies.

To account for the impact of environmental variables on the input slacks, a second stage regression is employed. To enable the adjustment of input variables for differences in the operating characteristics, input slacks are regressed upon the environmental variables instead of the efficiency scores. Separate equations are examined for each input to allow for the impact of the environmental variables to differ for each input. The dependent variable in equation 6.14 is the total input slacks s_{nj} for $n = 1, .., N$ DMUs and $j = 1, \dots, J$ inputs. The independent variable is a vector of environmental variables z_n.

$$s_{nj} = f^j(z_n; \beta^j) + \xi_{nj} \geq 0 \qquad n = 1, \dots, N, \qquad j = 1, \dots, J \quad (6.14)$$

As discussed previously, Simar and Wilson (2007) propose the use of a bootstrapped truncated regression to account for serial correlation in the efficiency scores. As the input slacks are a function of the estimated efficiency scores, the issue of serial correlation is still present within the second stage regression. Therefore, a bootstrap truncated regression is applied to obtain correct standard errors.

Simar and Wilson (2007) apply a truncated bootstrap regression for the environmental variables on the firm's efficiency score. The inverse of the Farrell efficiency is greater than or equal to 1, therefore by construction is truncated with a lower limit of 1. As the efficiency scores are truncated at 1 Simar and Wilson (2007) apply a left truncated regression at $(1 - Z\beta)$. To correct for the presence of serial correlation whilst using slacks Cordero-Ferrera et al (2010) note that the input slacks are greater than or equal to 0, therefore are left truncated at 0. Instead of the truncation point at $(1 - Z\beta)$ for the regression of the efficiency scores, the point of truncation is $-Z\beta$ for the input slacks. Following Simar and Wilson (2007) and Cordero-Ferrera et al (2010), the steps for each of the j truncated regression are:

1. The computation of s_{nj} for each input variable j and for all n decision making units using the original data.

   Run steps 2-4 for each j input equations

2. Use the method of maximum likelihood to obtain an estimation $\hat{\beta}$ of β as well as an estimate of $\hat{\sigma}_\varepsilon$ of σ_ε from $s_n = f(z_n, \beta) + \varepsilon_n$, considering it is a truncated regression at zero.
3. The computation of L (e.g. $L = 2000$) bootstrap estimates for $\hat{\beta}^*$ and $\hat{\sigma}_\varepsilon^*$ in the following way :
   3.1. For each $n = 1, \ldots, N$ draw from ε_n from the $N(0, \hat{\sigma}_\varepsilon^2)$ distribution with truncation at $-z_n\hat{\beta}_n$.
   3.2. Compute $s_n^* = f(Z_n, \beta_n) + \varepsilon_n$again for each $n = 1, \ldots, N$.
   3.3. Use the maximum likelihood method to estimate the truncated regression of s_n^* on z_n, yielding a bootstrap estimates $\left(\hat{\beta}^*, \hat{\sigma}_u^*\right)$.
4. Use the bootstrap values and original estimates $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$ to construct estimated confidence intervals for each element of β and σ_ε.

Simar and Wilson (2007) also propose a second algorithm which allows for the adjustment of the bootstrap bias as well as the estimation of the second stage regression with correct inference. Cordero-Ferrera et al (2010) apply Simar and Wilson (2007) algorithm #1 as it performs better in small samples than algorithm #2. Simar and Wilson (2007) highlight that when the number of inputs and outputs equal 1, 2 or 3 a sample size of 400 is required in order for the Root-Mean Square Error (RMSE) of algorithm #2 to be lower than algorithm #1. The RMSE measures the difference between the predicted model and the observed values. Algorithm #1 is applied to avoid additional biases due to the small sample size.

Separate regressions are considered for each J input to allow for the impact of the environmental variables to differ for each input. Fried et al (1999) apply a Seemingly Unrelated Regression (SUR) to allow for the error terms to be correlated as the slacks are derived from the first stage DEA. Fried et al (2002) use SFA within their estimation but highlight that the errors are probably not iid. They acknowledge the estimation gains from using SUR but note this is not yet possible with the composite error term. To my knowledge, SUR is not applicable for a bootstrap truncated regression.

The coefficients within the second stage regression allow for the determination of the magnitude and directional impact of the environmental variables on the input slacks. The total input slacks take a value greater than or equal to 0, and an efficient firm has 0 input slacks. A positive coefficient on the environmental variables indicates that the environmental variable is unfavourable as it leads to additional input slacks. On the other hand a negative coefficient indicates that the environmental variable reduces input slacks, and therefore operates in a favourable environment. Given the environmental operating characteristics of the firms the predicted slacks are obtained, $z_n\hat{\beta}^j$.

The third stage is to adjust the input variables by the amount in which are operating under a relatively favourable or unfavourable environment to ensure firms are compared when controlling for environmental differences. Fried et al (1999) adjust the inputs by adding a constant representing the difference between the predicted slack of the firm and the most unfavourable operating environment in equation 6.15. Tone and Tsutusi (2009) highlight that Fried et al (1999) adjustment process violates the translation invariance property as the addition of the constant skews the distribution. Tone and Tsutsui (2009) suggest the following input re-adjustment:

$$x_{nj}^{AA} = \frac{x_{jmax} - x_{jmin}}{x_{jmax}^{A} - x_{jmin}^{A}}\left(x_{nj}^{A} - x_{jmin}^{A}\right) + x_{jmin} \quad (6.15)$$

Where $x_{jmax} = max_n\{x_{nj}\}$, $x_{jmin} = min_n\{x_{nj}\}$, $x_{jmax}^{A} = max_n\{x_{nj}{}^{A}\}$ and $x_{jmin}^{A} = min_n\{x_{nj}{}^{A}\}$

Where $x_{nj}^{A} = x_{nj} - z_n\hat{\beta}^{j} - \hat{v}_{nj}$

Tone and Tsutsui (2009) highlight that adjustment process has the following properties:

1. x_{nj}^{AA} increases in x_{nj}^{A}. The re-adjusted data has the same ranking with the adjusted data. x_{nj}^{AA} is a linear transformation of x_{nj}^{A} with a positive coefficient, the coefficient and constant term are constant with the respective input item.
2. At x_{jmax}^{A}, x_{jmax}^{AA}attains the maximum value $x_{jmax}^{AA} = x_{jmax}$
3. At x_{jmin}^{A}, x_{jmin}^{AA}attains the minimum value $x_{jmax}^{AA} = x_{jmax}$

They highlight that the re-adjusted dataset $\{x_{nj}^{AA}\}$ remains within the range $[x_{jmin}, x_{jmax}]$ and that the maximum and minimum values are the same between $\{x_{nj}^{AA}\}$ and $\{x_{nj}\}$. They state that these properties are desirable as the inputs remain within the same range, which they find impacts upon the translation invariance property for the DEA scores. The adjustment process of Tone and Tsutsui (2009) adjusts the inputs both upwards and downwards, and the firm which has the most favourable environment has its inputs adjusted upwards by the largest amount.

The final stage re-runs the original DEA with the adjusted input values. This approach therefore allows for the identification of whether environmental variables impact upon the input slacks but also allows for adjusted-DEA scores to be calculated.

6.5 Data

This section will outline the data used within the paper and the definitions of the variables considered. The data is available for the period 1996/97–2010/11 for the ten WaSCs.

6.5.1 Definition of Variables

The functions undertaken by WaSCs are the abstraction, treatment and distribution of water followed by the collection, treatment and disposal of sewerage. Previously highlighted within chapter 2, the measurement of efficiency is analysed at the activity level and under a joint specification for water and sewerage activities. Efficiency modelled at the activity level allows for cost interactions between water and sewerage activities.

A four output model is analysed with two outputs measuring the physical volume of water and sewerage, the amount of potable water delivered (Y_1) and the equivalent population served (Y_2). The equivalent population served is a proxy for the physical output considered by Ofwat within their efficiency analysis. Garcia and Thomas (2001) and Stone and Webster (2004) highlight the estimation gains for including not only the physical quantity of outputs but also the number of properties served for water and sewerage. Therefore, alongside the physical outputs the number of water properties served (Y_3) and the number of sewerage properties served(Y_4) are also used.

To account for changes in quality requirements, a quality-adjusted measure of the physical inputs is applied. Adopting the approach of Saal and Parker (2000) as applied in chapter 5 the quality-adjusted measure of output is calculated as the water delivered is multiplied by the water quality index ($Y_1 = Water\ Delieved \times Q_w$) and the equivalent population served is multiplied by the sewerage quality index ($Y_2 = Equivalent\ Population \times Q_S$).

Three inputs are considered within the model: labour, capital and other. Labour costs are obtained from the statutory accounts following Saal et al (2011) rather than the June Return following Erbetta and Cave (2007) as the figures from the June Return only relate to direct staff costs and therefore ignore any indirect costs such as head office functions. The staff costs from the statutory accounts are not a perfect measure as they may include other group activities.

Capital costs are calculated as the sum of depreciation (including IRC[25]) and the opportunity cost of capital. The opportunity cost of capital is calculated as the Weighted Average Cost of Capital (WACC), applying assumptions made by Ofwat at each price review multiplied by the company's RCV. This measure is used following Saal et al (2011) and Maziotis et al (2013) rather than the MEA value as in other studies[26] as the RCV reflects the actual amount companies have spent rather than the value of an equivalent asset today. MEA has previously been used as it is reported separately for other water and sewerage activities whereas the RCV is reported for the WaSC.

Total costs are calculated as the sum of capital costs and operating expenditure net of third party services, exceptional items, doubtful debts, service charge and local authority rates[27]. Other costs are therefore calculated as total operating costs less labour costs and capital costs. Following Erbetta and Cave (2007) all costs apart from power costs are deflated using RPI to

[25] IRC is the infrastructure renewal charge, is a charge on infrastructure assets which acts as a depreciation.
[26] Saal and Parker (2000), Stone and Webster (2004), Saal and Reid (2004), Saal and Parker (2005), Saal et al (2007) and Bottasso and Conti (2009).
[27] These costs are deemed as non-controllable costs by Ofwat and are not incorporated within their assessment of efficiency. Exceptional items are by definition atypical. Third party services relate to costs incurred for output produced by other companies. Local authority rates and doubtful debts are considered as non-controllable. High levels of doubtful debts are due to the legal and regulatory decision of prohibiting the shutting off water and sewerage activities when bills are not paid. Service charges are charges by the Environment Agency for water abstraction.

2009 prices and power is deflated by an energy price index for the industrial sector derived from the Department for Trade and Industry (DTI).

Efficiency under input-orientation can either be measured under technical efficiency or cost efficiency. Technical efficiency examines the proportional reduction of all inputs to the frontier. Cost efficiency examines the optimal level of inputs given input prices and the frontier. Cost efficiency incorporates allocative efficiency and inputs can either be at their optimal level, under-utilised or over-utilised. The ratio of the actual input and the optimal input derived from cost efficiency can take a value of greater than, less than or equal to 1. Under cost efficiency, the dependent variable within the second stage regression represents the optimal, under- or over-utilisation of inputs. The interpretation of the directional impact of the environmental variable will depend on whether the input is over- or underutilised, which varies between firms. The interpretation of the impact of environmental variables on these input slacks is not as intuitive as those relating solely to technical efficiency, which represents the over-utilisation of inputs[28,29]. The use of costs instead of a quantity measure of inputs within the measurement of technical efficiency allows for the interpretation of the efficiency scores as cost efficiency. This approach has the drawback that prices are not assumed to be exogenous as they are in the application of a cost function DEA model. Thanassoulis (2000a, b) and Portela et al (2011) examine technical efficiency applying a cost measure for input. Thanassoulis (2000a) states that companies face similar staff and material prices. Therefore, after accounting for the level of output and the different operating environment, the remaining cost differentials will reflect

[28] An application of the adjustment of cost efficiency slacks is applied within chapter 7 for the measurement of dynamic efficiency. The measurement of dynamic efficiency requires the measurement of cost efficiency to highlight the impact on allocative inefficiency when ignoring intertemporal links of capital within the production process.

[29] Erbetta and Cave (2007) apply a cost function measure of DEA in which they examine the impact of environmental variables upon both the technical and allocative efficiency. The paper takes the absolute value of the allocative inefficiency and over- or under-utilisation of inputs to determine the influence of environmental variables. The paper does not allow for the final stage adjustment of the DEA scores for the DMUs in the three-stage approach.

managerial inefficiency. Overall, the methodology of using costs within the assessment of technical efficiency within DEA allows for the appropriate incorporation of environmental variables. It also determines the impact of the different regulatory periods whilst the results are comparable with Ofwat's assessment of relative efficiency.

6.5.2 Quality Variables

Environmental variables are incorporated to control for differences in the operating environment faced by the WaSCs. This allows for those companies that operate under a relatively unfavourable (high cost) environment to be analysed on a level playing field. The environmental variables were discussed extensively in section 4.4.8. The first environmental variable is the proportion of distribution input abstracted from rivers to take into account the differences in the resources and treatment (Z1). The second environmental variable is the density of a company's water operations which is calculated as the total water population divided by the length of mains (Z2). Similarly, sewerage density is calculated by the total sewerage population divided by the length of sewers (Z3). The proportion of trade effluent (Z4) is calculated as the volume of trade effluent divided by the volume of waste water returned. The proportion of leakage relative to distribution input (Z5) is incorporated. Alongside the environment variables, a time trend is included to account for technological progress or regress. A negative coefficient is expected due to technological progress over the period examined. Regulatory dummies are included to examine whether regulation has impacted on the economic environment. Two regulatory dummies are included for the price review in 1999 and 2004: Reg99 and PR04 respectively. Reg99 and Reg04 take the value of 1 for the five years after the price review in 1999 and 2004 respectively. The coefficient for the second stage regression will be negative if the regulatory price review has been effective in improving efficiency. Table 6.1 presents a snapshot of the data used in this chapter.

Table 6.1: Three-Stage DEA Sample Descriptive Statistics

Variable		Mean	Std.Dev	Min	Max
Outputs					
Water Delivered	Ml/Day	1,014.6	551.9	284.2	2,179.4
Equivalent Population	(,000)	6,179.8	3,704.1	1,118.4	14,271.9
Water Properties	(,000)	1,854.1	985.7	492.9	3,538.8
Sewerage Properties	(,000)	2,198.0	1,331.0	586.7	5,426.5
Inputs					
Labour Costs	£m	104.5	53.2	34.7	217.0
Capital Costs	£m	407.3	188.1	137.6	813.9
Other Costs	£m	137.2	77.7	15.3	370.0
Operating Characteristics					
Water Density		150.89	45.11	100.65	283.41
Sewerage Density		172.00	18.68	131.80	225.86
% of DI from Rivers		0.39	0.20	0.00	0.78
Proportion of Trade Effluent		0.07	0.04	0.03	0.18
Proportion of Leakage		0.23	0.05	0.15	0.38
Water Quality		0.97	0.03	0.84	1
Sewerage Quality		0.90	0.15	0.30	1

Note. 150 Observations. Costs are expressed in real terms in 2009 prices

6.6 Results

This section will outline the results for the incorporation of non-discretionary variables within DEA for the analysis of efficiency within the English and Welsh water and sewerage industry. The BCC DEA model under input-orientation is applied within the first stage. This DEA model calculates efficiency by examining the level of radial contraction of inputs given outputs and the feasible technology, assuming that WaSCs operate within a homogeneous operating environment. However, WaSCs operate under different environmental characteristics and the impact of the environmental variables on the excess input slacks is analysed through a second-stage bootstrapped truncated regression. Based upon the firm's environmental operating characteristics, its predicted slacks are estimated. Finally, the inputs are adjusted relatively to the amount in which they operate under a favourable environment and the DEA programme is repeated. The adjustment mechanism for inputs is proposed by Tone and Tsutsui (2009) who expand on the work of Fried et al (1999, 2002) to overcome the problem of the violation of the translation invariance property. The DEA scores under the first and third stage are compared to determine whether adjusting for environmental variables impacts on the both the efficiency scores and ranks of WaSCs.

6.7.1 1st Stage Efficiency Scores

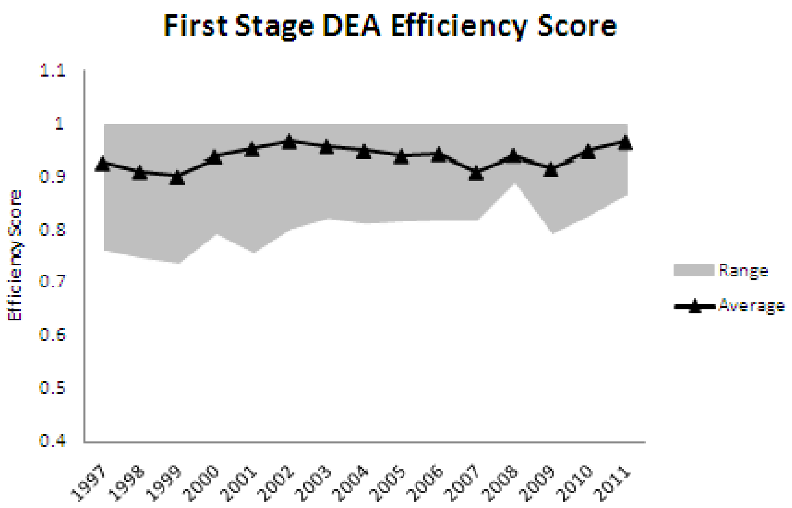
The results from the first stage DEA model are reported in table 6.2; whilst figure 6.1 depicts the range of efficiency scores alongside the average. The efficiency scores range from 1 to 0.74 with a mean efficiency score over the whole period of 0.94, which therefore indicates a mean inefficiency of 6%. The efficiency scores differ from those calculated for the measurement of beta-convergence in chapter 5. Firstly, the number of firms which are deemed as efficient under β-convergence is higher; this is because a separate frontier is analysed within each time period. Secondly, the model suffers from a dimensionality problem, reducing the discriminatory power of the model. The efficiency scores also differ due to the inclusion of labour and capital as a separate input within the measurement of three-stage DEA.

Table 6.2 First Stage DEA Efficiency Scores

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC 1	0.801	0.747	0.738	0.793	0.757	0.802182	0.822	0.813	0.817	0.819	0.819	0.913	0.802	0.827	0.867
WaSC 2	0.762	0.790	0.797	0.885	0.884	1	0.994	0.951	0.904	0.885	0.875	0.911	0.793	0.857	1
WaSC 3	0.994	1	1	1	1	1	0.970	0.980	0.984	1	0.974	1	0.941	1	1
WaSC 4	1	0.973	0.952	0.973	0.955	1	0.999	1	0.976372	1	0.920	0.978	0.934	1	1
WaSC 5	0.992	0.967	0.964	0.992	0.995	0.994	0.978	0.957	0.951	0.930	0.898	0.890	0.931	0.928	0.956
WaSC 6	0.905	0.824	0.831	0.863	0.950	0.884	0.857	0.837	0.882	0.962	0.916	0.936	0.953	1	0.954
WaSC 7	1	1	0.982125	1	1	1	1	1	1	1	0.958	0.945	0.982	1	1
WaSC 8	0.948	0.930	0.913	0.945	1	1	0.973	0.969	0.891	0.900	0.863	0.908	0.904	0.926	0.964
WaSC 9	0.994	1	0.977995	1	1	1	0.996	0.999	1	0.992	0.948	1	1	1	1
WaSC 10	0.871	0.875	0.862	0.944	0.997	1	0.998	0.995	1	0.955	0.929	0.938	0.921	0.961	0.925
Average	0.927	0.911	0.902	0.939	0.954	0.968	0.959	0.950	0.941	0.944	0.910	0.942	0.916	0.950	0.967
Standard Deviation	0.089	0.095	0.090	0.071	0.079	0.069	0.064	0.069	0.064	0.061	0.047	0.039	0.069	0.064	0.044
Minimum	0.762	0.747	0.738	0.793	0.757	0.802	0.822	0.813	0.817	0.819	0.819	0.890	0.793	0.827	0.867

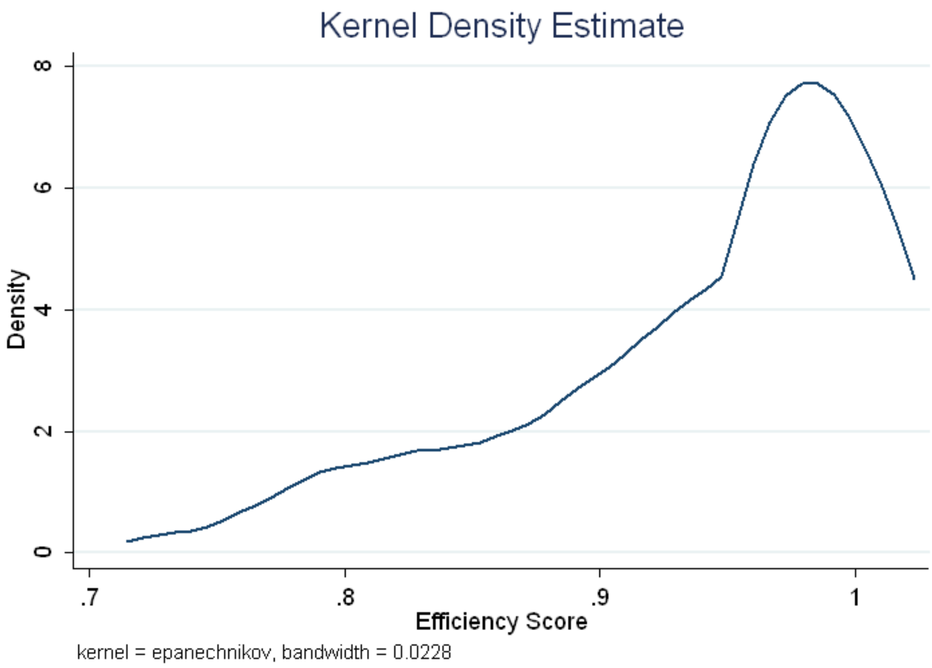
The minimum efficiency score has increased from 0.76 in 1997 to 0.87 in 2011. The average and minimum efficiency scores have generally improved over the period 1997-2011. The dispersion of efficiency scores has fallen over the period; the standard deviation has fallen from 0.09 to 0.044. These results coincide with those from chapter 5 and Saal et al (2007) which show that regulation has been effective in encouraging the least efficient WaSCs to catch up to the frontier. Figure 6.1 depicts that the average efficiency score improved after 1999 however started to deteriorate in 2003 and reached the lowest average in 2007. Portela et al (2011) find that companies are moving closer towards to the meta-frontier until 2002 when they become relatively stable and start to decline in 2006 and 2007.

Figure 6.1: Evolution of First Stage DEA Efficiency Score



The distribution of the efficiency scores over the whole period are displayed in figure 6.2. This shows the typical distribution expected when a small proportion of the firms are inefficient, and the density increases as the efficiency score increases.

Figure 6.2 First Stage DEA Scores Kernel Density Estimate



6.7.2 2nd Stage Bootstrap Regression

The DEA model incorporates input and outputs to determine the level of efficiency and excess input slacks. To account for the influence of non-discretionary variables upon the production possibility set, a second stage regression is applied[30]. The bootstrap truncated regression with left truncation at $-Z\beta$ in equation 6.14 determines the impact of the environmental variables on the excess levels of inputs. The results from the second stage bootstrapped truncated

[30] Company-specific heterogeneities are controlled for by the environmental variables. Fixed effects are not incorporated within the second stage regression as the time-invariant component could capture persistent inefficiencies.

regression for each input with 2,000 iterations following that amount used by Simar and Wilson (2007)[31] are displayed in table 6.3.

Water and sewerage density have been incorporated alongside a squared term; the estimates reveal a negative first order term and a positive second order term. This indicates that as density increases input slacks are reduced until this is exhausted at a sufficiently high level of density and firms require additional resources. Overall, those firms operating in a very rural and very urban environment require higher inputs, and therefore are considered to be operating within a relatively unfavourable environment[32]. The first and second order coefficients for sewerage density for all inputs are significant at least at the 10% level. This indicates that slacks decrease as density increases up to a certain threshold and then slacks begin to increase. The coefficients for water density are insignificant. Previous literature reports that high water density is characterised by a favourable operating environment (Bottasso and Conti, 2003; Erbetta and Cave, 2007). Saal and Parker (2005), Bottasso and Conti (2009)[33] and Saal et al (2011) allow for the non-linear relationship between water density and inputs/costs in which both report that very rural and dense operating characteristics are unfavourable. The results for the impact of sewerage density are less consistent in the literature. Erbetta and Cave (2007) for labour slacks and Tupper and Resende (2004) report a positive impact on slacks therefore indicating that urban areas are unfavourable, although the impact is insignificant. On the other hand, Saal and Reid (2004) report a negative impact on slacks whilst Saal et al (2011) find a negative first order term and positive second order term.

[31] Simar and Wilson (2007) state that more accurate estimates can be achieved with the larger amount of replications. Simar and Wilson (2007, pp. 44) apply 2,000 replications. 5,000 replications were examined but produced little or no change in the standard errors.
[32] For the capital stock the optimal density is 250.5 and 172.9 for water and sewerage respectively. Taking the average density for the firms over the period the optimal water density is only reached for WaSC 7. For sewerage density, firms within the dataset are exhibited on all components of the curve, both the decreasing and increasing component.
[33] Bottasso and Conti (2009) find that the increase in costs as density increases, the congestion effect becomes significant at levels of density not experienced within their sample for WoCs.

The coefficient for the proportion of DI from rivers is positive for all inputs and significant for labour and capital inputs. This therefore indicates that a higher proportion of DI from rivers is unfavourable as it leads to additional slacks. This implies that the additional treatment costs associated with river abstraction out weights the lower abstraction costs. This result is consistent with those reported from Bottasso and Conti (2003) and Cherchye et al (2013)[34]. An alternative to incorporating the DI from rivers is to include the proportion of DI from boreholes. Erbetta and Cave (2007), Saal et al (2007) Saal et al (2011) find that a higher proportion of abstraction from boreholes reduces input requirements[35].

The proportion of total waste water which is trade effluent has a positive and significant coefficient for all input slacks. The results conclude that treating trade effluent requires additional inputs, therefore reflecting the relative intensity of treating trade effluent. These results coincide with those of Saal et al (2007) who report that a higher proportion of trade effluent requires higher input requirements. Erbetta and Cave (2007) find an insignificant relationship between labour and other input slacks and trade effluent however report a negative and significant relationship for capital. This indicates a higher proportion of trade effluent reduces slacks which indicates better performance.

The influence of leakage is negative and significant for labour and capital inputs. This result indicates that higher leakage reduces input slacks and therefore is favourable. The expected influence of the impact of leakage is ambiguous as Erbetta and Cave (2007) argue a higher proportion of leakage indicates a deteriorating asset condition, and therefore is more costly to operate. However, Cherchye et al (2013) argue that higher leakage may be as a result of a lack

[34] Alongside the proportion of DI from rivers, Cherchye et al (2014) incorporate the proportion of DI from boreholes in which they find a higher proportion of borehole is favourable. The proportion of DI from boreholes was also considered alongside rivers however the influence was not significant.
[35] The proportion of DI from boreholes was included however was insignificant and therefore the proportion of DI from rivers was incorporated.

of capital expenditure on fixing leaks. Erbetta and Cave (2007), Saal et al (2007) and Saal et al (2011) report that a higher proportion of leakage increases costs. On the other hand Cherchye et al (2013) report mixed results. These results report that a high proportion of leakage is a result of a lack of capital maintenance expenditure, therefore indicating the environment is favourable.

Finally, the impact of the time trend and regulatory dummies on input slacks is examined. The time trend is negative for labour slacks and other slacks, which indicates a reduction in labour slacks over the period, although it is insignificant. The impact of the time trend for capital inputs is positive. The results therefore indicate a more intensive use of capital inputs over the period. The positive coefficient may be due to the presence of a capex bias due to the Averch-Johnson effect and the nature of the industry which is focused on capex solutions to meet future demand. Erbetta and Cave (2007) report a significant and positive influence of time for technical efficiency, suggesting a more intensive input requirement over the period. Erbetta and Cave (2007) report a negative time trend for the absolute value of the allocative inefficiency, highlighting an improvement in the optimal combination of inputs.

To examine whether the different regulatory periods has influenced the input slacks, two dummy variables have been incorporated for the 1999 and 2004 price review. The regulatory dummies relating to the 1999 and 2004 price review are negative for all input slacks. The main impact of the price review was to decrease input slacks, however the results are only significant for labour and capital inputs for the 1999 price review. The influence of the 1999 price for other inputs was insignificant, therefore to determine whether the overall effect of the 1999 price review was significant, the environmental variables are regressed upon total slacks[36].

[36] All inputs are incorporated in monetary values and therefore can be summed. This approach is not applicable if inputs are not in the same unit of measurement. The regression of total input slacks allows for the incorporation of radial and non-radial slacks. A second-stage bootstrapped truncated regression was also

Column 4 in table 6.3 reports the bootstrapped truncated regression for total slacks which reveals a negative and significant influence of the 1999 price review. Therefore, the overall influence of the 1999 price review resulted in a significant reduction of slacks and therefore improved efficiency within the industry. The coefficient for the 2004 price review is insignificant for all inputs and the total input slacks, indicating the review had no influence on improving the efficiency score.

Erbetta and Cave (2007) report that the 1999 price review was statistically significant in reducing slacks, however the 1994 review was statistically insignificant. They state that this might be due to change in the regulatory policy in 1999. Portela et al (2011) examine productivity and report increasing productivity in the period 1993–2005. Portela et al (2011) find an improvement in the meta-frontier until 2002 which coincides with Ofwat's statement that the improvement in the efficiency scores as a result of the 1999 price review is striking. The paper also reports that productivity fell after 2005 and continued to fall until 2007. This period coincides with the 2004 price review, although they do highlight that the result may be due to the price review alongside other factors such as an increase in electricity and fuel prices.

The 1999 price review was a particularly tough review which was the only review to date to impose a negative K factor whereas the 2004 review imposed a K factor greater than that imposed at privatisation. Overall, this chapter contributes to the literature by reporting that the 1999 price review brought significant efficiency improvements, whereas the impact of the 2004 price review was statistically insignificant. The results found within this chapter coincide with the industry experience of the impact of the price review with Sir Ian Byatt, the former director

estimated for the radial efficiency score, and the results hold that the influence of the 1999 price was negative and significant.

general stating that the 1999 price review was tough whereas the 2004 review was generous (Utility Week, 2014).

Table 6.3: Second Stage Bootstrapped Truncated Regression

	Dependent Variable			
Independent Variable	Labour Slack	Other Slack	Capital Slack	Total Slacks
Constant	462.089***	1322.556*	1637.001***	2281.979***
	(147.086)	(687.598)	(597.111)	(708.942)
Water Density	-1.557	-6.673	-2.201	-4.868
	(1.292)	(4.485)	(3.292)	(4.662)
(Water Density)^2	0.004	0.018	0.004	0.012
	(0.004)	(0.015)	(0.011)	(0.016)
Sewerage Density	-3.661**	-10.855*	-16.998**	-21.972***
	(1.432)	(6.580)	(7.299)	(8.118)
(Sewerage Density)^2	0.010***	0.032*	0.049**	0.063***
	(0.004)	(0.019	(0.021)	(0.023)
Leakage	-288.947***	-304.465	-785.674**	-842.138**
	(103.506)	(305.456)	(317.993)	(401.879)
Proportion of DI from Rivers	25.680*	93.344	185.083**	-842.137**
	(14.771)	(70.858)	(73.813)	(401.879)
Proportion of Trade Effluent	433.457***	1270.763**	1829.951***	2130.455***
	(138.918)	(593.109)	(546.859)	(642.287)
t	-0.051	-1.888	11.634***	9.460***
	(0.645)	(3.256)	(3.615)	(3.378)
Reg99	-16.211**	-23.735	-67.421**	-79.161**
	(6.689)	(27.039)	(26.839)	(31.597)
Reg04	-1.663	-0.942	-5.859	-11.159
	-4.932	(22.214)	(18.125)	(21.535)

Note: Estimates marked within (***) are significant a the 1% level; estimates marked within (**) are significant at the 5% level; estimates marked with (*) are significant at the 10% level.

6.7.3 3rd Stage Adjusted DEA Scores

The 3rd stage adjusts the DEA scores by the amount in which WaSCs operates under a favourable or unfavourable environment. Applying the adjustment procedure of Tone and Tsutsui (2009) in equation 6.15 table 6.4 reports several descriptive statics for the adjusted inputs variables. The adjusted inputs using the Tone and Tsutsui (2009) method remain within the same range of the initial resources. The adjustment process varies the mean and standard deviation of the inputs.

Table 6.4: Adjusted Inputs Descriptive Statistics

	Initial Resources				Tone and Tsutusi (2009)			
	Mean	Std.Dev	Min	Max	Mean	Std.Dev	Min	Max
Other	137.21	77.73	15.30	370.02	166.09	82.45	15.30	370.02
Labour	104.48	53.22	34.68	217.05	97.41	45.88	34.68	217.05
Capital	407.30	188.05	137.57	813.88	406.87	171.62	137.57	813.88

The DEA is re-run using the adjusted inputs and the adjusted efficiency scores are displayed in table 6.5 and depicted in figure 6.3. The minimum efficiency score has fallen from 73.8% within the first stage adjustment to 62.9% in the final stage. The average efficiency score has also decreased from 93.9% to 87.0% after the 3rd stage adjustment. This indicates that several firms were deemed relatively efficient as they were operating within a favourable environment, once controlling for this their efficiency score fell. The efficiency scores from the 1st stage regression show a clear trend over time however after accounting for the environmental variables, technological progress and the influence of the price review the trends in the efficiency scores has dampened.

Table 6.5- Third Stage Environmental Adjusted DEA Scores

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC 1	0.773	0.840	0.869	0.843	0.774	0.795	0.804	0.805	0.867	0.870	1	0.846	0.820	0.867	0.837
WaSC 2	0.948	0.944	0.963	0.849	0.888	0.937	0.846	0.814	0.913	0.863	0.871	0.879	0.708	0.776	0.870
WaSC 3	0.806	0.822	0.851	0.829	0.822	0.815	0.791	0.784	0.857	0.807	1	0.851	0.780	0.842	0.851
WaSC 4	1	0.931	1	1	0.957	0.977	0.878	0.891	0.988	0.892	0.866	0.851	0.873	0.987	1
WaSC 5	0.729	0.884	1	0.769	0.748	0.815	0.755	0.757	0.995	0.902	1	0.810	0.835	0.977	1
WaSC 6	0.843	0.798	0.798	0.744	0.844	0.762	0.763	0.757	0.874	0.850	0.900	0.956	0.908	0.969	0.971
WaSC 7	1	0.984	0.972	0.951	0.956	0.919	0.890	0.928	0.922	1	0.878	0.899	0.909	1	1
WaSC 8	1	1	0.881	0.894	0.943	1	0.924	0.892	0.865	0.830	0.850	0.834	0.778	0.826	0.801
WaSC 9	0.629	0.740	0.748	0.650	0.723	0.817	0.712	0.766	0.901	0.823	0.847	0.812	0.810	0.851	0.893
WaSC 10	0.976	0.913	0.916	0.826	0.861	0.943	0.900	0.910	1	0.948	0.881	0.812	0.792	0.855	0.802
Average	0.870	0.886	0.900	0.835	0.852	0.878	0.826	0.830	0.918	0.879	0.909	0.855	0.821	0.895	0.902
Standard Deviation	0.133	0.084	0.086	0.101	0.086	0.086	0.072	0.068	0.057	0.060	0.064	0.046	0.063	0.080	0.083
Minimum	0.629	0.740	0.748	0.650	0.723	0.762	0.712	0.757	0.857	0.807	0.847	0.810	0.708	0.776	0.801

Figure 6.3 3rd Stage DEA Efficiency Score

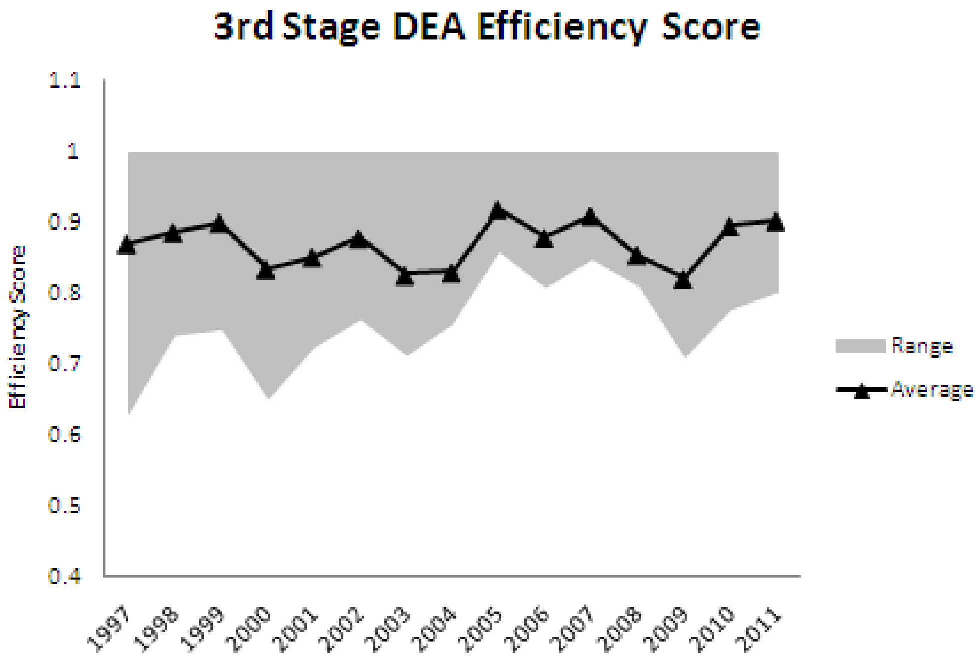


Figure 6.4 shows the kernel density function of the efficiency scores for the 1st and 3rd stage DEA. After controlling for the differences in the environmental operating characteristics, the dispersion of the efficiency scores widens. The number of firms which make up the frontier has decreased from 43 to 22. This indicates again that several firms appear relatively efficient because they operate within favourable operating environments relative to the other firms. To test whether the DEA efficiency scores are significantly different for the 1st and 3rd stages a Wilcoxon-rank sum test is applied alongside a Spearman's rank correlation test and Pearson correlation. The Wilcoxon-rank sum test is a non-parametric test with the null hypothesis that the two samples are from the same population. The Wilcoxon rank-sum test rejects the null hypothesis that the distributions are the same at the 5% level with $z = -6.992$. The result therefore indicates that the efficiency scores are significantly different when controlling for the differences in environmental characteristics.

As well as using the Wilcoxon-Rank Sum test to determine whether there is a relationship between the first and third stage DEA, a Pearson correlation and Spearman rank correlation test is applied. The Pearson coefficient for the pooled sample between the 1st and 3rd stage is 0.149 and is significant at the 10% level. The result therefore indicates that there is a weak correlation between the unadjusted and environmental adjusted DEA scores. This indicates that once controlling for the environment the relative efficiency scores vary significantly. The results indicate that three WaSCs; WaSC 3, WaSC 5 and WaSC 9 have an average reduction in their efficiency score by 15, 9 and 21 percentage points respectively once controlling for differences in their operating environments. This would therefore indicate that these firms operate within a relatively favourable environment and when controlling for this their efficiency scores are reduced.

On the other hand WaSC 1 has experienced an increase in its average efficiency score over the period, highlighting that it operates in an unfavourable environment, and when controlling for environmental differences their efficiency scores improve. The remainder of the firms have experienced a smaller change in their average mean efficiency between 1 and 6 percentage points.

Figure 6.4 1st and 3rd Stage Kernel Density Estimate

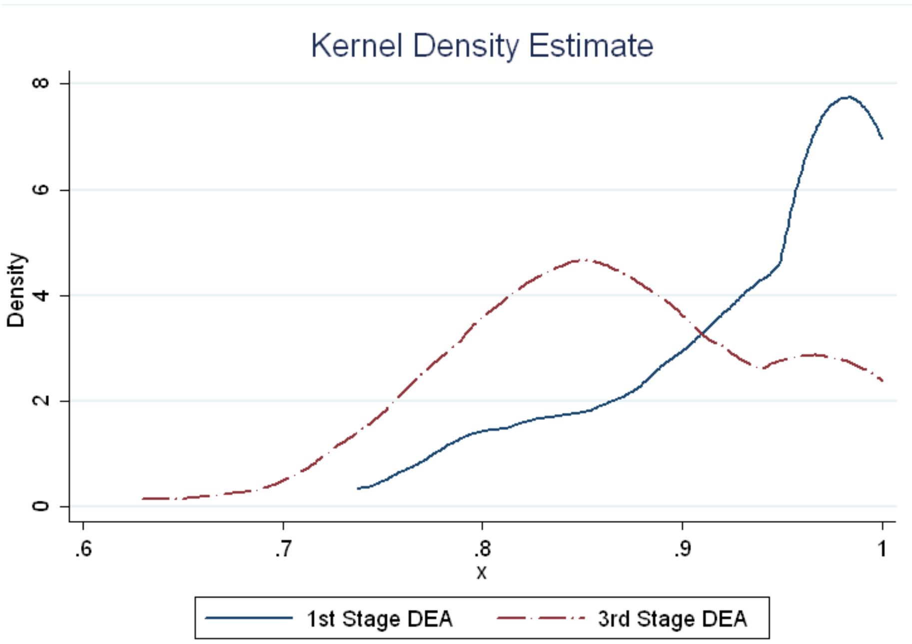


Table 6.6- First and Third Stage DEA Efficiency Score Rank

1st Stage Efficiency Score Rank

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC 1	9	10	10	10	10	10	10	10	10	10	10	7	9	10	10
WaSC 2	10	9	9	8	9	1	5	8	7	9	8	8	10	9	1
WaSC 3	3	1	1	1	1	1	8	5	4	1	1	1	4	1	1
WaSC 4	1	4	5	5	7	1	2	1	5	1	5	3	5	1	1
WaSC 5	5	5	4	4	6	8	6	7	6	7	7	10	6	7	7
WaSC 6	7	8	8	9	8	9	9	9	9	5	6	6	3	1	8
WaSC 7	1	1	2	1	1	1	1	1	1	1	2	4	2	1	1
WaSC 8	6	6	6	6	1	1	7	6	8	8	9	9	8	8	6
WaSC 9	3	1	3	1	1	1	4	3	1	4	3	1	1	1	1
WaSC 10	8	7	7	7	5	1	3	4	1	6	4	5	7	6	9

3rd Stage Adjusted Efficiency Score Rank

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
WaSC 1	8	7	7	5	8	9	6	6	8	5	1	6	5	5	8
WaSC 2	5	3	4	4	4	4	5	5	5	6	7	3	10	10	6
WaSC 3	7	8	8	6	7	7	7	7	10	10	1	4	8	8	7
WaSC 4	1	4	1	1	1	2	4	4	3	4	8	5	3	2	1
WaSC 5	9	6	1	8	9	7	9	9	2	3	1	10	4	3	1
WaSC 6	6	9	9	9	6	10	8	10	7	7	4	1	2	4	4
WaSC 7	1	2	3	2	2	5	3	1	4	1	6	2	1	1	1
WaSC 8	1	1	6	3	3	1	1	3	9	8	9	7	9	9	10
WaSC 9	10	10	10	10	10	6	10	8	6	9	10	8	6	7	5
WaSC 10	4	5	5	7	5	3	2	2	1	2	5	9	7	6	9

The results so far examine whether the incorporation of the environmental variables have influenced the value of the efficiency score. The following section examines whether the ranks of the companies have altered once accounting for the environment. The Spearman's Rank correlation test is the measure of statistical dependencies between the two datasets analysing the differences in their ranks. The Spearman's rank correlation coefficient between the first stage DEA results and Tone and Tsutsui (2009) adjusted efficiency scores have a correlation coefficient for the pooled sample of 0.189. The coefficients are significant at the 1% level. The results therefore conclude that there is a weak correlation between the firm's efficiency scores when controlling for the environmental variables.

Table 6.6 reports the ranks of the efficiency scores under the first stage DEA and the Tone and Tsutsui (2009) adjusted DEA scores. The ranks vary substantially between the 1st and 3rd stage efficiency scores; WaSC 2 was 8th to 10th within the industry with the exception of 2002 and 2011. Once adjusting for the environmental variables the ranks increased to mainly 5th over the period with the exception of 2009 and 2010. On the other hand, WaSC 3 made up the frontier over the period 1998–2002 and once accounting for the environment their efficiency score reduced to 6th to 8th. These results therefore confirm that it is not only the relative efficiency scores that are changing; the influence of adjusting for the environmental variables influences the ranks for the WaSCs.

The results highlight from a methodological viewpoint the importance of incorporating environmental variables within the measurement of efficiency. If environmental variables are not incorporated the efficiency scores will be biased depending on whether firms operate within a favourable or unfavourable environment. The results highlight that within the first stage DEA, several firms make up the frontier because they operate within a favourable environment.

6.7 Conclusion

This chapter employs a three-stage DEA model introduced by Fried et al (1999) and extended by Cordero-Ferrera et al (2010) and Tone and Tsutsui (2009) to account for the differences in firms operating environment when measuring efficiency. DEA makes the implicit assumption of homogeneity; however WaSCs operate in different geographical locations which have different operating characteristics which can influence the production function. Ray (1988) highlights the theoretical importance of controlling for environmental variables as this can influence production function, therefore making firms look efficient or inefficient if they operate within a favourable or unfavourable environment respectively.

Non-discretionary variables in the English and Welsh water and sewerage industry have been incorporated extensively through the parametric approaches. However DEA is advantageous as it does not require any assumptions with regards to its functional form. Erbetta and Cave (2007) produce input-specific efficiency measures applying DEA, this chapter contributes to the literature by producing an overall company measure of efficiency once account for differences in operating characteristics. Within the literature there is a large amount of evidence of the influence of the 1994 and 1999 price review. Portela et al (2011) report limited evidence of deteriorating efficiency as a result of the 2004 price review. The second contribution of this chapter is to examine if the 1999 and 2004 price review has significantly influence the efficiency score.

The impact of the 1999 and 2004 price reviews are examined in which the results indicate that the 1999 price review had a significant impact on improving efficiency, whereas the 2004 price review did not have a significant influence. The 1999 price review is thought to be a particularly tough price review imposing large efficiency challenges on firms whereas the 2004 review is considered as relatively lax.

The second stage regression implies a separability condition; therefore the environmental variables influence the distribution of the input slacks and do not influence the optimal combination of inputs. An extension of this chapter would be to examine conditional DEA introduced by Darario and Simar (2005) which conditions the efficiency scores on a set environmental variable.

The final stage adjusted the input slacks for the different operating environments and repeats the DEA model to obtain environmental adjusted DEA efficiency scores to account for whether firms operate within a relative favourable or unfavourable operating environment. The efficiency scores are compared between the first and third stage. The results highlight a large change in both the efficiency scores and ranks for several firms. The adjusted DEA scores highlight the importance of controlling for environmental variables within the measurement of efficiency using DEA. When firms are examined on a level playing field the distribution and ranking of their efficiency scores vary substantially. The number of firms which make up the frontier decreased from 43 to 22, and therefore several firms were deemed efficient due to operating within a favourable environment. The chapter highlights the importance of controlling for environmental variables; the exclusion of environmental variables can significantly impact the managerial or regulatory decisions.

7. Dynamic Efficiency

7.1 Introduction

The English and Welsh water and sewerage sector is a long-life capital intensive industry, characterised by natural monopolies. The English and Welsh water and sewerage industry is characterised by quasi-fixed inputs such as mains, sewers and treatment works which cannot be adjusted to their optimal level instantaneously. The static DEA model measures efficiency assuming that all inputs can be adjusted to their optimal level instantaneously, firms may face adjustment costs. The static model does not take into account decisions companies make today to influence the production in the future. Firms face a trade-off between increasing output production today or investing in capital to increase production in future periods. To account for the intertemporal nature of capital dynamic DEA is examined.

The aims of the chapter are twofold: firstly to compare the conclusions from static and dynamic DEA highlighting the inefficiencies that arise out of a dynamic framework, and secondly to investigate the presence of a preference for capital expenditure known as the capex bias. Ofwat (2011b) defines the capex bias as the view that companies within the industry have an inappropriate preference for expenditure on capital assets over day-to-day operational expenditure. The 2014 price review to set prices for 2015–2020 is partly designed around eliminating the presence of the perceived capex bias. CEPA (2012) state that the bias is believed to exist for three reasons. Firstly, there are the different financial incentives created by examining capex and opex separately. Secondly, the presence of what is termed as the Averch-Johnson effect; which arises if the allowed rate of return is higher than the true cost of capital (Averch and Johnson, 1962). Opex is recovered within the period; however capex is added to the Regulatory Capital Value (RCV) which earns a return based upon Ofwat's assumptions of the cost of capital. Thirdly, the culture of the sector is one that is focused on capex solutions and infrastructure to meet future demand. Through examining dynamic DEA this chapter aims to examine the presence of the perceived capex bias.

DEA is used to measure efficiency within a dynamic context by examining the presence of quasi-fixed capital. A dynamic perspective of the measurement of efficiency is required as investment today does not only influence today's production, but also future periods. Intertemporal effects are incorporated through the inclusion of capital as an output in the current period production as well as an input from the previous period's production. Firms face an installation cost investing in quasi-fixed inputs; the more resources that are utilised in installing additional quasi-fixed inputs, less are available for the production of output. Firms therefore face a trade-off between increasing output today or producing capital to increase outputs in the subsequent period (Geymueller, 2009). The chapter allows overall efficiency to be decomposed into a dynamic component and a static component. This approach determines the level of efficiency due to variable inputs and the inefficiency due to quasi-fixed inputs.

The chapter takes a three stage approach by including environmental variables[37] within the dynamic framework to ensure firms are compared on a level playing field. Input slacks ratios are obtained from the dynamic DEA which are then regressed upon the environmental variables. The predicted slack ratios are used to adjust the input variables upwards for those firms operating in a relatively favourable environment. The DEA is repeated including the adjusted inputs to obtain efficiency scores when accounting for differences in the operating environments.

The remainder of the chapter is organised as follows: Section 2 describes the regulatory environment; Section 3 briefly reviews the extant literature; Section 4 outlines the methodology

[37] Fried et al (1999, 2002)

to measure dynamic efficiency; Section 5 outlines the sample data and variable definitions; Section 6 presents and discusses the results and Section 6 concludes the paper.

7.2 Literature Review

A common theme running throughout the literature of efficiency within the English and Welsh water and sewerage industry is the treatment of capital. Capital can be modelled either as part of a variable cost function or a long run cost function where the latter assumes that firms have the ability to adjust all inputs in the long run to their optimal level. Within the variable cost function, capital is incorporated as a quasi-fixed input, therefore capital is not considered as a control variable and cost minimisation is only related to variable inputs. Saal and Reid (2004) model capital as a quasi-fixed factor as the technology used within the industry is indivisible and associated with a long service life and therefore it is difficult to vary capital stock in the short-term. In addition, WaSCs face legal obligations to connect customers to the service as well as investment programmes agreed with the DWI and EA to meet increasing quality standards. Stone and Webster (2004) and Bottasso and Conti (2009) estimate a variable cost function for the English and Welsh water and sewerage industry and the coefficient on the quasi-fixed input implies a tendency of overcapitalisation which is a common finding in the literature of public utilities (Caves et al, 1981 and Cowing and Holtman, 1983). Bottasso and Conti (2009) state that overcapitalisation can be interpreted as the Averch-Johnson effect due to the presence of rate of return regulation alongside the capital intensive nature of the industry and the presence of investment to meet future demand. Stone and Webster (2004) and Bottasso and Conti (2009) state that the presence of overcapitalisation could result in a misspecified total cost function where the assumption is that firms can instantaneously vary the level of capital. Saal and Parker (2000) and Erbetta and Cave (2007) consider a total cost function through the use of parametric and non-parametric specifications respectively. Erbetta and Cave

(2007) find an initial under-utilisation of capital and over-utilisation of labour which diminishes over the period. Saal and Parker (2000) report capital-augmenting and labour-saving technological change over the period considered. This result is not surprising as one of the goals of privatisation was to expand capital investment in the industry. Indeed Saal and Parker (2001) examine total factor productivity and find substantial capital growth in the post-privatisation period.

The pioneering work of Farrell (1957) measures efficiency as the distance between an observation and an estimated ideal referred to as an efficient frontier. Cooper et al (2006) defines DEA as a non-parametric technique which uses mathematical linear programming to create a piece wide surface or frontier over the data. Traditionally DEA is studied within a static context, therefore inputs and outputs were considered for a given period. The static model implies that all inputs can be adjusted to their optimal level within the given period and there are no links between time periods. The Malmquist index allows for the evolution of efficiency over time to be measured[38] and is used to explain changes between two consecutive time periods. The Malmquist index allows for the decomposition of the intertemporal efficiency change into a catch-up and innovation (frontier-shift) effect.

The static model is based on the firm's ability to instantaneously adjust the factors of production and ignore the intertemporal linkage of production decisions (Silva and Stefanou, 2003). However, Emrouznejad and Thanassoulis (2005) make the case for the presence of an intertemporal relationship: 1) The existence of a stock of capital whose useful service life and the effects of investment extend over several periods; 2) The presence of lagged outputs which, in addition to contemporaneous effect of the inputs, depends on the inputs used in the previous periods and 3) The production of intermediate outputs.

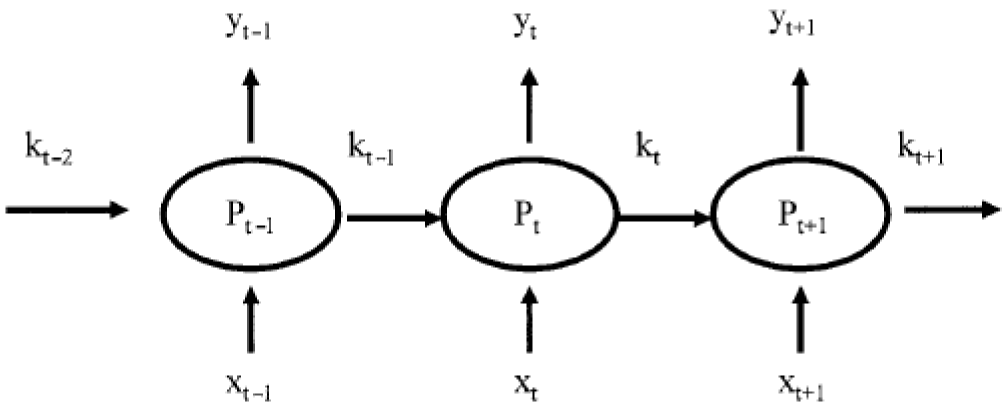
[38] See Fare et al (1994)

Nemoto and Goto (2003) argue that the assumption of static optimisation results in biased measurements of inefficiency if quasi-fixed inputs exist. They show that the allocative inefficiency in particular will be overstated to the extent that quasi-fixed inputs are not instantaneously adjusted to their optimal levels. The weakness of the static theory of production to describe how some inputs gradually adjust has led to the development of dynamic models.

Sengupta (1995) was perhaps the first to introduce dynamics through the adjustment costs of quasi-fixed inputs. Fare and Grosskopf (2000) introduce dynamics through the use of network DEA considering capital as an intermediate outputs. Nemoto and Goto (1999, 2003) propose a model of dynamic DEA using a cost function which is closely related to adjustment-cost theory of investment and therefore provides a nonparametric alternative to the parametric Euler equation. The model augments the conventional DEA by treating quasi-fixed inputs at the end of one period as if they were outputs in the period and essentially inputs in the subsequent one as depicted in figure 7.1. The firm therefore faces a trade-off, whether to myopically increase output or to increase quasi-fixed inputs to increase future production.

Geymuller (2009) extend Nemoto and Goto (2003) by solely considering technical efficiency in the absence of input prices. Tone and Tsutusi (2010) consider a slacks-based approach which considers both radial and non-radial efficiency. Their model allows for the inclusion of desirable, undesirable, discretionary (free) and non-discretionary (fixed) variables. The model however does not allow for the breakdown of overall efficiency to account for the inefficiencies relating to static and dynamic components. Ouellete and Yin (2008) propose a dynamic DEA model which allows for the inclusion of quasi-fixed inputs when investment decisions are outside of the control of the firm and are rather controlled by a higher authority. Silva and Stefanou (2007) develop a dynamic efficiency model which takes into account interdependence of production decisions whilst allowing for temporal efficiency measurements. Capital is treated as a quasi-fixed factor and is managed as an asset where rapid expansion or contraction of the stock of capital is accompanied by adjustment costs.

Figure 7.1: Production Process



Source: Fare and Grosskopf (2000)

The model of Nemoto and Goto (2003) is applied to the English and Welsh water and sewerage industry to incorporate the intertemporal nature of capital. Firstly, the capital within the industry has a long asset life and span over several periods. Secondly, the firms face a trade-off between the performance in the current period and investing in capital to meet future demand and to improve the quality of outputs and productivity.

7.3 Methodology

This section outlines the dynamic DEA model by Nemoto and Goto (2003) which will be applied to the English and Welsh water and sewerage industry. This uses N DMUs $(n = 1, \dots, N)$, J variable inputs $(j = 1, \dots, J)$, S quasi-fixed inputs $(s = 1, \dots, S)$ producing R outputs $(r = 1, \dots, R)$. The DEA model is shown in equation 7.1 where γ^t is a constant discount factor, x_t denotes the variable inputs used in period t, and k_t denotes the quasi-fixed at the end of period t. w_t, v_t are price vectors of variable inputs and quasi-fixed inputs in period t respectively. y_t denotes the outputs produced in period t and i is a $N \times 1$ vector of ones to impose the convexity constraint under Variable Returns to Scale (VRS). VRS specification is examined to obtain a pure measure of managerial inefficiency excluding any scale inefficiencies as WaSCs do not have control over their operating size, unless the regulator permits mergers (Thanassoulis, 2000a). Overall efficiency (OE) examines the cost minimising level of quasi-fixed and variable inputs given prices whilst incorporating the quasi-fixed inputs as an intertemporal factor of production.

$$\hat{C}(\bar{k}_0) = \min_{\{x_t, k_t, \lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \,(w_t' x_t + v_t' k_{t-1})$$

$$s.t. \quad X_t \lambda_t \le x_t\,, \qquad t = 1,2,\dots,T$$

$$K_{t-1} \lambda_t \le k_{t-1}\,, \qquad t = 1,2,\dots,T$$

$$K_t \lambda_t \ge k_t\,, \qquad t = 1,2,\dots,T-1$$

$$Y_t \lambda_t \ge y_t\,, \qquad t = 1,2,\dots,T$$

$$i' \lambda_t = 1\,, \qquad t = 1,2,\dots,T$$

$$k_0 = \bar{k}_0\,, \quad x_t \ge 0\,, \quad k_t \ge 0\,, \quad \lambda_t \ge 0\,, \quad t = 1,2,\dots,T \qquad (7.1)$$

The bar over the variables represents fixed levels of variables. The difference between this DEA model and the static model is the inclusion of the capital stock at time *t* as an output. Nemoto and Goto (2003) highlight that overall efficiency is calculated by:

$$OE = \hat{C}(\bar{k}_0)/C \qquad (7.2)$$

Here C is the discounted sum of actual costs over the period considered. Overall efficiency can be decomposed into dynamic and static efficiency; static efficiency can then be decomposed into technical efficiency and allocative efficiency. Static efficiency is calculated holding the quasi-fixed inputs fixed and examining the optimal level of variable inputs given input prices. The difference between overall efficiency and static efficiency is dynamic efficiency. Static efficiency can be written formally as the linear programming problem in equation 7.3.

$$C_{SE} = \min_{\{x_t,\lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \left(w_t' x_t + v_t' \bar{k}_{t-1}\right)$$

$$\begin{aligned}
s.t.\quad & X_t\lambda_t \le x_t\,, && t = 1,2,\dots,T \\
& K_{t-1}\lambda_t \le \bar{k}_{t-1}\,, && t = 1,2,\dots,T \\
& K_t\lambda_t \ge \bar{k}_t\,, && t = 1,2,\dots,T-1 \\
& Y_t\lambda_t \ge y_t\,, && t = 1,2,\dots,T \\
& i'\lambda_t = 1\,, && t = 1,2,\dots,T \\
& x_t \ge 0\,,\quad \lambda_t \ge 0\,,\quad t = 1,2,\dots,T && \text{(7.3)}
\end{aligned}$$

Static efficiency (SE) and dynamic efficiency (DE) are then calculated by equation 7.4 and 7.5.

$$SE = C_{SE}/C \qquad (7.4)$$

$$DE = \hat{C}\left(\bar{k}_0\right)/C_{SE} \qquad (7.5)$$

Nemoto and Goto (2003) highlight that dynamic efficiency includes forecast errors for input prices and demands for outputs in the future.

Static efficiency can be broken down into allocative and technical efficiency. Technical efficiency is obtained by examining the radial contraction of variable inputs by solving the following linear programming problem:

$$C_{TE} = \min_{\{\theta_t,\lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t \left(\theta_t w_t' \bar{x}_t + v_t' \bar{k}_{t-1}\right)$$

$$s.t. \quad X_t\lambda_t \leq \theta_t \overline{x_t}\ , \qquad t = 1,2,\dots,T$$

$$K_{t-1}\lambda_t \leq \bar{k}_{t-1}\ , \qquad t = 1,2,\dots,T$$

$$K_t\lambda_t \geq \overline{k_t}\ , \qquad t = 1,2,\dots,T-1$$

$$Y_t\lambda_t \geq y_t\ , \qquad t = 1,2,\dots,T$$

$$i'\lambda_t = 1\ , \qquad t = 1,2,\dots,T$$

$$x_t \geq 0\ , \quad \lambda_t \geq 0\ , \quad t = 1,2,\dots,T \qquad (7.6)$$

The radial contraction θ_t is allowed to vary over the periods. Since the quasi-fixed inputs are exogenously given at the actual levels there are no restrictions across the periods. Therefore the LP programme can be reduced to T single period problems that are independent of one another. Technical efficiency (TE) is measured as

$$TE = C_{TE}/C$$

Allocative efficiency (AE) can be calculated as

$$AE = C_{SE}/C_{TE}$$

AE reflects the costs that could be saved if variable inputs were adjusted to the optimal level along the short-run isoquant. The relationship for overall efficiency can be decomposed as

$$OE = TE.AE.DE$$

The measurement of inefficiency in the period t for variable inputs τ_t^x and for quasi-fixed inputs τ_t^k follows as:

$$\tau_t^x = \gamma^t w_t'(x_t - x_t^*)/C \qquad t = 1,2, \dots, T;$$

$$\tau_t^k = \gamma^t v_t'(k_{t-1} - k_{t-1}^*)/C \qquad t = 2,3, \dots, T; \qquad (7.8)$$

Where x_t and k_t are evaluated at the observed values, k_t^* and x_t^* are the optimal values of capital and variable inputs at time t. C is the discounted sum of actual total costs over the planning period. Positive (negative) values of τ_t^x and τ_t^k indicate excess (short) usage of inputs. The equations measure the inefficiencies according to the normalised deviations of observations along from the optimal input usage.

Dynamic DEA is compared to static DEA where all inputs are considered as variable therefore implying that they can be instantaneously adjusted to their optimal level. Static cost efficiency (CE^S) is obtained by the following linear programme:

$$CE^S = \min_{x_t, k_t, \lambda_t} \sum_{t=1}^{T} \gamma^t (w_t' x_t + v_t' k_{t-1})$$

$$s.t. \;\; X_t \lambda_t \leq x_t, \quad t = 1,2,..,T$$

$$K_{t-1} \lambda_t \leq k_{t-1}, \quad t = 1,2,..,T$$

$$Y_t \lambda_t \geq y_t, \quad t = 1,2,..,T$$

$$i' \lambda_t = 1, \quad t = 1,2, \dots, T$$

$$\lambda_t \geq 0, x_t \geq 0, k_t \geq 0, \quad t = 1,2,..,T \qquad (7.9)$$

$$OE^S = CE^S/C \quad (1.10)$$

Cost efficiency can be decomposed into allocative and technical efficiency by equation (7.12). Static technical efficiency (TE^S) is measured through the following linear programme:

$$TE^S = \min_{\theta_t,\lambda_t} \sum_{t=1}^{T} \gamma^t \theta_t (w_t' x_t + v_t' k_{t-1})$$

$$s.t.\ \ X_t\lambda_t \leq \theta_t x_t\,, \quad t = 1,2,..,T$$

$$K_{t-1}\lambda_t \leq \theta_t k_{t-1}, \quad t = 1,2,..,T$$

$$Y_t\lambda_t \geq y_t\,, \quad t = 1,2,..,T$$

$$i'\lambda_t = 1\,, \quad t = 1,2,\ldots,T$$

$$\lambda_t \geq 0\,, \quad t = 1,2,..,T \qquad (7.11)$$

Static allocative efficiency (AE^S) is defined as

$$AE^S = OE^S/TE^S \quad (7.12)$$

As DEA implies homogeneity the second part of this paper incorporates environmental variables within the dynamic DEA framework through the use of the three stage approach based upon Fried et al (1999). The first stage generates the efficiency scores and the input slacks. The second stage accounts for the effect of the environmental impact upon the slacks through a second stage regression. The third stage adjusts inputs variables to create a level playing field before repeating the DEA analysis. Input variables are adjusted upwards for those firms that operate under relatively favourable environments. The firms with a relatively unfavourable operating environment have their inputs adjusted by a relatively small amount, while those with favourable operating environments are adjusted upwards by a relatively large amount. Adjusting the inputs upwards provides a performance target managers can reach regardless of their operating environment (Fried et al, 1999).

Fried et al (1999, 2002) consider the impact of environmental variables on technical efficiency while Blank and Valdmanis (2005) extend their work to consider the impact of environmental variables on cost efficiency through cost efficiency slacks for each firm and time period. The cost slack ratio is the ratio of the optimal input x^*_{jnt} for the j^{th} variable input, n^{th} DMU and t^{th} time period and the actual variable input defined by equation 7.13. The cost slack ratio for quasi-fixed inputs is defined in equation 7.14 for the s^{th} quasi fixed input, n^{th} DMU and t^{th} time period. The cost slack ratios are greater than, less than or equal to 1[39]. A value greater than 1 implies an over utilisation of input, a value less than 1 implies an underutilisation of inputs and a value equal to 1 implies an efficient use of inputs.

$$S_{jnt} = \frac{x_{jnt}}{x^*_{jnt}} \quad (7.13)$$

$$SK_{snt} = \frac{k_{snt}}{k^*_{snt}} \quad (7.14)$$

To determine the impact of environmental variables the input slack ratios are regressed on the environmental variables. The slack ratios are centred on their means and do not have a mass of observations at one point meaning that this approach avoids the censoring problem when using efficiency scores within the second stage regression. To ensure that the predicted slack ratios take positive values only a log transformation of the dependent variable is taken. The data for each time period is pooled and regressed upon M environmental variables

[39] Fried et al (1999, 2002) consider both the radial and non-radial slacks whereas this approach only considers radial slacks.

$Q_{nt} = [Q_{1nt,\dots,}Q_{Mnt}], n = 1, .., \mathrm{N}, t = 1,,,.,T$ separately for each input slack[40]. Simar and Wilson (2007) highlight the presence of serial correlation amongst the efficiency scores generated by DEA which leads to incorrect inference within the second stage regression. To correct for the presence of serial correlation this chapter applies a second stage bootstrapped regression based on Simar and Wilson (2007) as outlined in chapter 6[41].

$$log(S_{jnt}) = f^j(Q_{nt}; \beta^j) + u_{jnt} \quad (j = 1, \dots, J \quad t = 1, \dots, T \quad n = 1, \dots, N) \quad (7.14)$$

$$log(SK_{snt}) = f^s(Q_{nt}; \beta^s) + u_{snt} \quad (s = 1, \dots, S \quad t = 1, \dots, T-1 \quad n = 1, \dots, N)$$

Where $f^j(Q_{nt}; \beta^j)$ and $f^s(Q_{nt}; \beta^s)$ are deterministic feasible slack frontiers, parameter vectors β^j and β^s are to be estimated. The interpretation of the coefficients depends upon whether the resource is under or over-utilised. If the resource is over-utilised and takes a value greater than 1, a negative coefficient will imply moving to the optimal level of resources. However, if the input is under-utilised a negative coefficient will imply that the environmental variable is unfavourable, moving away from the optimal utilisation of inputs.

The predicted slacks are obtained and the inputs are adjusted using the methodology of Blank and Valdmanis (2005) for each WaSC and time period[42]. The inputs are adjusted upwards using equation 7.15 in which the least favourable set of environmental conditions are used as a base[43].

---

[40] Fried et al (2002) highlight that as the slacks are obtained from the first stage DEA model when considering separate equations, the error components are probably not independently and identically distributed (i.i.d). It would be preferable to stack the regressions and estimate via SUR allowing the error terms to be correlated across inputs. Due to the timing difference within the dynamic DEA this has not be considered here.

[41] Simar and Wilson (2007) apply a bootstrap truncated regression as the inverse of the Farrell efficiency scores are bounded at 1. A bootstrap regression is applied as the cost slack ratios can take any value greater than, less than or equal to 1.

[42] Regression for capital slacks is run for the period 1997-2010 as the linear programme does not return an optimal value of the capital stock for the beginning of 1997 and the end of 2011. Using the coefficients obtained, predicted slacks are obtained for the beginning of 1997 and 2011 and capital is adjusted using the same methodology.

[43] The reciprocal of the ratio from Blank and Valdmanis (2005) is used so the maximum slack, considered as the most unfavourable environment is the highest over utilisation of inputs. Inputs are adjusted upwards by the amount in which firms operate under a favourable environment.

$$x_{jnt}^{adj} = x_j^{nt}\left(\frac{max^{nt}(\hat{S}_{jnt})}{\hat{S}_{jnt}}\right) \quad j = 1,..,J;\ t = 1,\ldots,T;\ n = 1,..,N \quad (7.15)$$

$$k_{snt}^{adj} = k_s^{nt}\left(\frac{max^{nt}(\widehat{SK}_{snt})}{\widehat{SK}_{snt}}\right) \quad s = 1,..,S;\ t = 0,\ldots,T;\ n = 1,..,N$$

Where $\hat{S}_{jnt}$ are the predicted slacks of resource j for firm n at time t and $max^{nt}(\hat{S}_{jnt})$ represent the firm with the most unfavourable conditions. Firms operating within a favourable environment have their inputs adjusted upwards by a relatively larger amount whereas those with unfavourable environment have their inputs adjusted upwards by a relatively smaller amount. The final stage re-runs the dynamic DEA programme using the adjusted inputs to control for environmental differences.

7.4 Data

This section describes the data used within the chapter and the definitions of the variables considered. The data is available for the period 1996/97–2010/11 for the ten WaSCs. Within the panel there were three mergers of WaSCs with the smaller WoCs[44].

7.4.1 Definition of Variables

For the application of this methodology for the English and Welsh water and sewerage industry, two physical outputs will be considered for water and sewerage activities: water delivered and the equivalent population served. To take into account changes in quality within the industry the output measures are adjusted by a quality index following Saal and Parker (2000). Water

[44] Mergers occurred in April 2000 between Hartlepool Water and Anglian Water, Northumbrian Water and Essex and Suffolk Water and finally the merger between Yorkshire Water and York Waterworks. As capital from the merged entity is incorporated over time, to avoid the acquisition of capital being treated as investment, the data for the pre-merged companies has been aggregated

input is measured as $(Y_1) = \ Water\ Delivered \times Q_w$where Q_w is a measure of water quality and Sewerage output is measured as $(Y_2) = Equivalent\ Popuation \times Q_s$, where Q_s is a measure of sewerage quality. Two outputs have only been incorporated due to the limited number of DMUs.

Three inputs are considered within the model: labour, capital and other. Staff costs and the number of full time employees are obtained from companies' statutory accounts. The price of labour is calculated as the total labour costs divided by the number for full time equivalent employees.

The value of the capital stock is measured by the MEA value. Capital costs are calculated as the sum of depreciation (including IRC[45]) and the opportunity cost of capital. The opportunity cost of capital is calculated as the WACC applying assumptions made by Ofwat at each price review multiplied by a company's RCV. Capital price is calculated as capital costs divided by the capital stock.

Other costs are therefore calculated as total operating costs less labour costs and capital costs. Other costs comprise of a composite of other goods and therefore, following Saal et al (2011), the price of other goods has been proxied by the price index relating to the price of inputs bought for the distribution and purification of water collected by the ONS. A measure for the physical amount of other inputs is calculated as other costs divided by the price of other costs.

The model is a 2 variable input, 1 quasi-fixed input and 2 output model for 10 DMUs per period[46].

---

[45] IRC is the Infrastructure Renewal Charge. This is a charge on infrastructure assets which acts as a depreciation.

[46] The model suffers from a dimensionality issue, therefore the efficiency scores are biased upwards, although the relative positions stay the same. The chapter seeks to examine the relative differences of the inefficiencies between variable and capital inputs and the comparison of static and dynamic DEA.

7.4.2 Quality Variables

To account for the differences in operating environments several environmental variables are incorporated within the analysis. The first environmental variable is the proportion of distribution input abstracted from rivers to take into account the differences in the resources and treatment (Z1). The density of a company's water operations is calculated as the total water population divided by the length of mains (Z2). Similarly, sewerage density is calculated by the total sewerage population divided by the length of sewers (Z3). The level of leakage is controlled by calculating the proportion of leakage relative to distribution input (Z4). Finally, the proportion of trade effluent (Z5) is calculated as the volume of trade effluent divided by the volume of waste water returned.

Table 7.1: Sample Descriptive Statistics

Variable		Mean	Std.Dev	Min	Max
Outputs					
Water Delivered	Ml/Day	1,014.6	551.9	284.2	2,179.4
Equivalent Population	(,000)	6,179.8	3,704.1	1,118.4	14,271.9
Water Properties	(,000)	1,870.4	979.2	492.9	3,538.8
Sewerage Properties	(,000)	2,198.0	1,331.0	586.7	5,426.5
Input Quantities					
Capital Stock	£m	23,229.9	13,400.8	7,030.2	49,129.9
Employees		2,943.3	1,401.7	1,157.0	5,894.0
Other		184.4	111.4	15.3	491.2
Inputs Prices					
Capital Price	£	0.019	0.004	0.010	0.030
Labour Price	£m	0.035	0.004	0.024	0.045
Other Cost Deflator		0.782	0.140	0.651	1.052
Operating Characteristics					
Water Density		150.893	45.114	100.648	283.411
Sewerage Density		172.001	18.675	131.798	225.861
% of DI from Rivers		0.393	0.203	0.000	0.781
Trade Effluent		0.074	0.035	0.025	0.175
Proportion of Leakage		0.227	0.049	0.147	0.379
Water Quality		0.967	0.026	0.836	0.995
Sewerage Quality		0.904	0.152	0.302	1

Note. 150 Observations. Costs and Input prices are expressed in real terms in 2009 prices

7.5 Results

In the first instance, the results are considered for the implication of modelling dynamic versus static DEA. The results from the dynamic DEA allows for the decomposition of efficiency into dynamic, static, technical and allocative inefficiencies. The optimality conditions and trends for the average efficiency are considered over the period. Secondly, the implication of firms operating environments is examined. The results from the second stage regression allow for the assessment of whether differing operating characteristics are favourable or unfavourable for the firms. Finally, the static and dynamic DEA will be re-estimated and the results are compared against the original DEA to examine the impact on firms' efficiency scores when taking into account their operating environment.

The planning period within the model covers the period 1996/97–2010/11, thus $\bar{k}_0$ corresponds to the initial capital stock at the beginning of 1997 and k_T is the terminal capital stock at the end of 2011. Table 7.2 reports the efficiency scores calculated under variable returns to scale for both the dynamic and static efficiency model. The application of a cost function implies input orientation; firms reduce their inputs given the amount of outputs. Input-orientation is the assumption mostly considered within the literature (Thanassoulis, 2000a,b; Cubbin and Tzanidakis, 1998 and Erbetta and Cave, 2007) as the demand level faced by suppliers is exogenous. Overall efficiency (OE) for dynamic DEA can be decomposed into static efficiency (SE), technical efficiency (TE), allocative efficiency (AE) and dynamic efficiency (DE). The results for the OE reports the efficiency score when firms can adjust both variable and quasi-fixed capital whilst taking into account the intertemporal nature of capital. Static efficiency within the dynamic context considers the efficiency when firms can only consider the reduction of variable inputs and capital is held as fixed. OE scores range from 77.8% to 100%. The results for the SE are higher than those for OE, which indicates that given the level of quasi-fixed inputs firms are between 91.6% and 100% efficient. Dynamic efficiency is calculated as the ratio of OE and SE. Dynamic inefficiency ranges from 0% to 21.9% and these results imply that firms are relatively efficient with regards to the variable inputs and that the quasi-fixed inputs are the major source of overall inefficiency.

The static model allows for the contraction of both variable and quasi-fixed inputs without taking into account the intertemporal nature of quasi-fixed inputs. OE^S indicates that firms are between 73.9% and 100% efficient. These results are all less than or equal to OE but of a similar magnitude with the differences ranging between 0 and 5.4%. The OE^S is decomposed into TE^S and AE^S, and these results show that within the static model the main source of inefficiency is due to the wrong factor mix (AE^S). The relative magnitude of the overall efficiency scores under dynamic and static DEA are of a similar magnitude. The dynamic DEA allows for an extension of the static DEA to incorporate capital over time and to examine the overall and efficiency holding capital fixed. The results indicate that the firms are relatively efficient with regards to variable inputs and the majority of the inefficiencies are due to the over utilisation of capital inputs.

Table 7.2: Dynamic and Static Efficiency Score

	Dynamic					Static		
	OE	SE	TE	AE	DE	OE^S	TE^S	AE^S
WaSC1	0.894	0.932	0.938	0.994	0.959	0.870	0.956	0.910
WaSC2	0.778	0.916	1.000	0.916	0.849	0.739	0.992	0.746
WaSC3	1	1	1	1	1	1	1	1
WaSC4	1	1	1	1	1	1	1	1
WaSC5	1	1	1	1	1	1	1	1
WaSC6	0.859	0.965	0.985	0.979	0.890	0.804	0.959	0.838
WaSC7	1	1	1	1	1	1	1	1
WaSC8	0.779	0.998	1	0.998	0.781	0.761	0.993	0.767
WaSC9	1	1	1	1	1	1	1	1
WaSC10	0.932	0.997	1	0.997	0.935	0.917	0.995	0.922

Figures 2(a-f) demonstrate the deviations over time of dynamic overall inefficiency measured by deviations from the optimality condition τ_t^x and τ_t^k in equation 7.8. Five of the firms are showed below; the remaining five are efficient over the whole period and therefore do not have deviations from optimality.

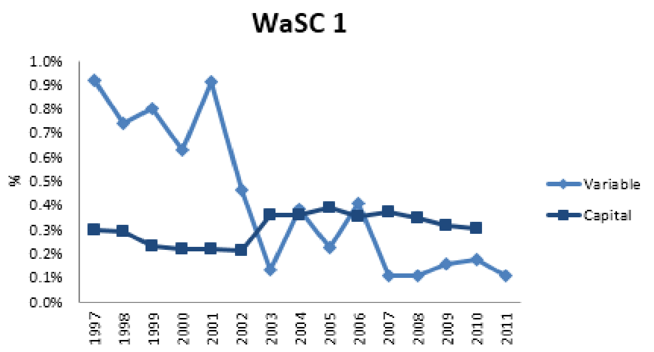


Figure 7.2a

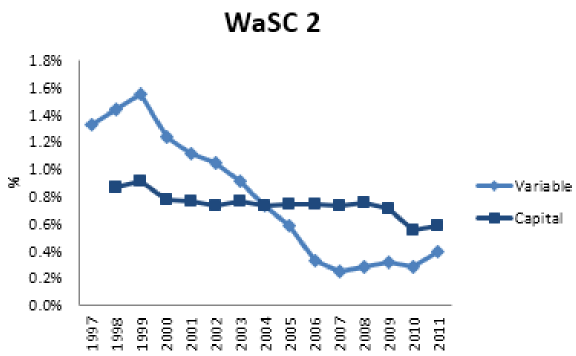


Figure 7.2b

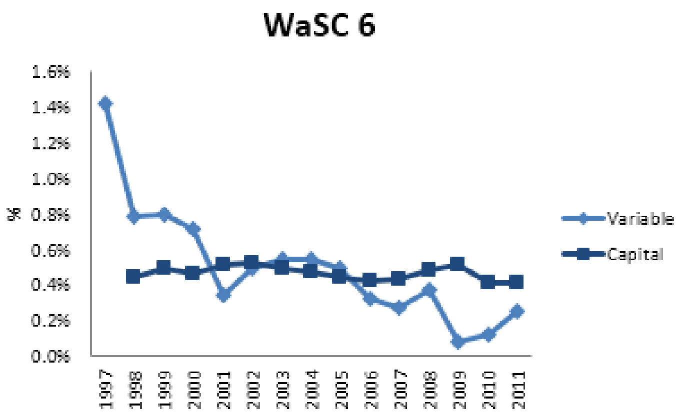


Figure 7.2c

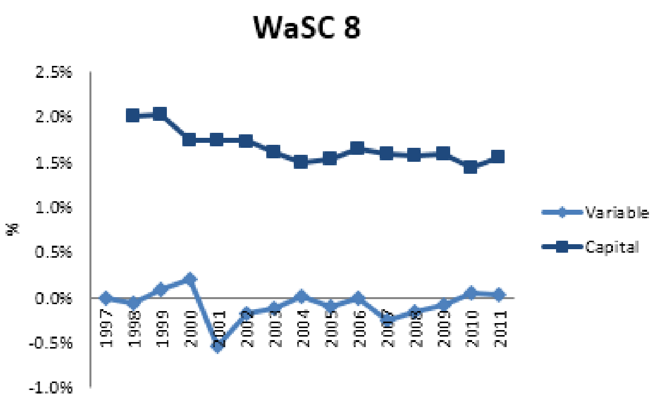


Figure 7.2d

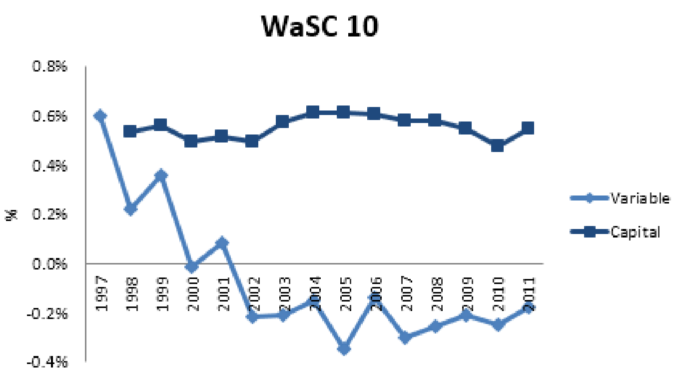


Figure 7.2e

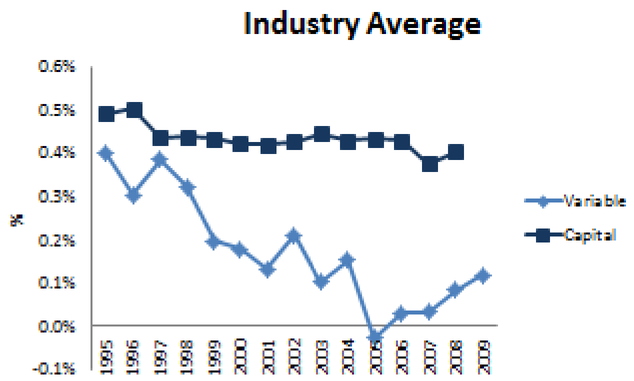


Figure 7.2f

A common finding is the over utilisation of capital inputs. The deviations in variable inputs from the optimal seem more volatile over the period, whereas quasi-fixed inputs are persistently overused. For WaSC 1 and WaSC 2 it can be seen that there has been a considerable improvement in the efficiency of variable inputs over the period considered, whereas quasi-fixed remain persistently overused. Dynamic DEA incorporates the intertemporal nature of capital within the evaluation of efficiency. Figure 7.3 compares the optimal deviation of capital from the actual value τ_t^k for the dynamic and static DEA. The static DEA implies that firms should substantially reduce the level of capital in the first five years of the period in comparison to the dynamic case. The path of capital under dynamic DEA is a lot smoother in comparison to static DEA. This is expected as the static DEA does not take into account the capital needed for the future period and the adjustment cost of capital. WaSC 1, WaSC 2 and WaSC 6 experience large differences in the optimal value of capital determined by the static and dynamic DEA. The difference between the static and dynamic DEA reduces over the period. This is because output increases over the time period, and therefore the additional capital is required to produce output. One limitation within the model is that the final stage DEA is effectively a static DEA model and therefore we expect the results to converge. The results highlight the need to account the intertemporal nature of capital when measuring efficiency as the static DEA model will underestimate the optimal value of capital.

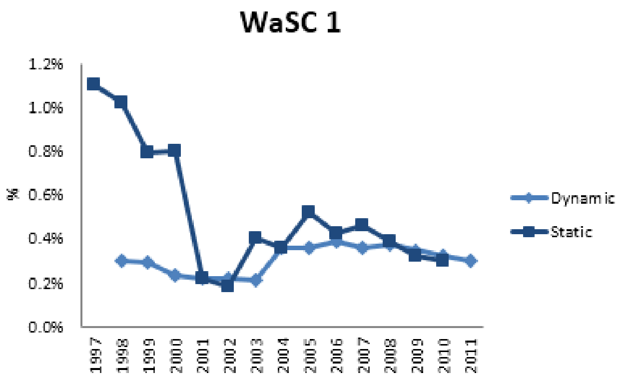


Figure 7.3a

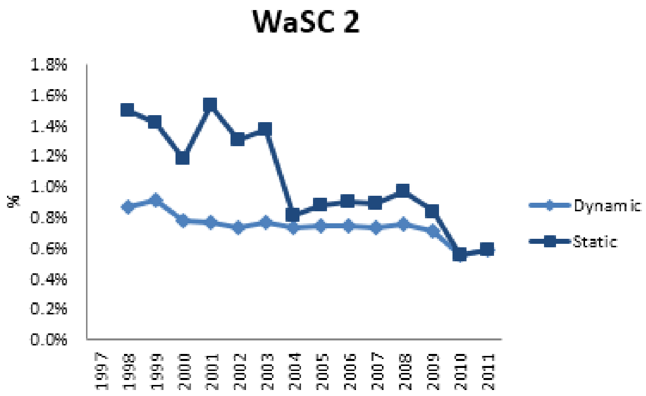


Figure 7.3

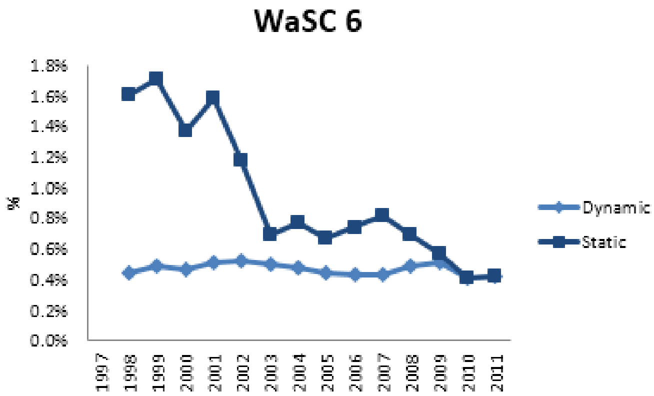


Figure 7.3c

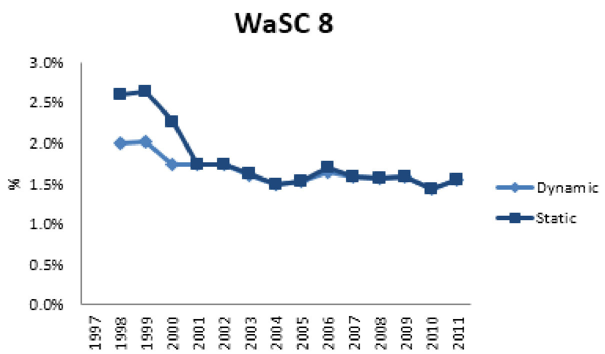


Figure 7.3d

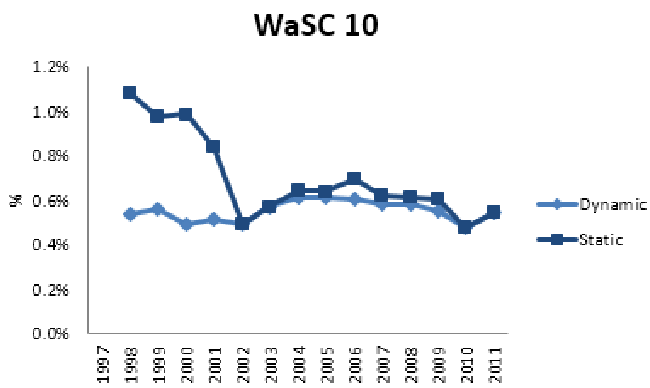


Figure 7.3e

We now turn to the incorporation of environmental variables within the dynamic DEA. Slacks are obtained from the dynamic cost function and the static cost function and are regressed upon a number of environmental variables to take into account the impact of operating conditions on efficiency. Five environmental variables are considered within the analysis, namely water and sewerage density, the proportion of trade effluent, the proportion of leakage and the proportion of DI from boreholes. Table 7.3 shows the coefficients and standard errors of the regressions:

Table 7.3: Slack Regression. Standard errors in parenthesis

	Dependent Variable		
Independent Variable	Labour Slack	Other Slack	Capital Slack
Intercept	-0.6333	-0.6840***	-0.4360***
	(0.1601)	(0.2244)	(0.0565)
Water Density	0.0001	-0.0028***	-0.0002
	(0.0003)	(0.0005)	(0.0001)
Sewerage Density	0.0009	0.0063**	0.0022***
	(0.0008)	(0.0013)	(0.0004)
Leakage	-0.6475	0.6177	-0.0605
	(0.03462)	(0.4204)	(0.1684)
Proportion of DI from Rivers	0.0302	0.0365	0.0032
	(0.0780)	(0.1125)	(0.0388)
Proportion of Trade Effluent	-0.5368	0.4362	2.4011***
	(0.3158)	(0.4793)	(0.2798)

Note: Estimates marked within (***) are significant a the 1% level; estimates marked within (**) are significant at the 5% level; estimates marked with (*) are significant at the 10% level.

All companies over-utilise the capital inputs but the interpretation of the coefficients for labour and other inputs is more complex as some companies over-utilise these inputs whilst others under-utilise them. The coefficient for the proportion of DI obtained from rivers for capital inputs have a positive coefficient. Capital is over-utilised by firms; the positive value would therefore imply that those with a higher proportion of DI from rivers operate under a relatively more unfavourable operating condition. Rivers require more power with regards to treatment than the other sources of abstraction. The coefficient for water density for capital is negative which implies a higher density is favourable; however this result is statistically insignificant. On the other hand, the coefficient for sewerage density for capital is positive; therefore it implies that operating in an urban environment is unfavourable. A squared term for capital was incorporated, however was insignificant, therefore remove for the final specification. WaSCs with a higher proportion of trade effluent operate within an unfavourable operating environment, which is indicated through the positive coefficient for capital. This result is intuitive as trade effluent may require a higher level of treatment, therefore imposing higher costs upon the companies. Leakage has a negative coefficient for capital slacks, which indicates that utilities with higher leakage operate under a favourable environment; this may be due to a lack of capital maintenance expenditure, although the coefficients are insignificant. These results are consistent with those reported in chapter 6, with the exception of density.

The predicted efficiency scores are generated from the regressions and the actual variables are adjusted upwards by the amount in which they operate under a favourable environment relative to the most unfavourable using equation 7.15. Summary statistics are reported in table 7.4 for the inputs before and after adjustments for environmental factors. The results indicate a higher mean for all input variables and the standard deviation between inputs increases when taking into account the differing operating environments.

Table 7.4: Environmental Adjusted Data Description

	Initial Resources				Adjusted Resources			
	Mean	Std.Dev	Min	Max	Mean	Std.Dev	Min	Max
Employees	2943.3	1401.7	1157.0	5894.0	3268.0	1586.7	1241.6	6812.3
Other Inputs	184.4	111.4	15.3	491.2	237.8	173.9	17.8	877.8
Capital	23229.9	13400.3	7030.2	49129.9	27344.0	16130.7	9408.2	83262.7

The DEA efficiency scores are recalculated with the adjusted data and the dynamic and static efficiency results are shown in table 7.5 below.

Table 7.5: Environment Adjusted Dynamic and Static DEA Efficiency Scores

	Dynamic					**Static**		
	OE	**SE**	**TE**	**AE**	**DE**	**OE^S**	**TE^S**	**AE^S**
WaSC1	0.889	0.948	0.958	0.990	0.937	0.875	0.960	0.911
WaSC2	0.859	0.958	0.989	0.969	0.897	0.828	0.984	0.841
WaSC3	1	1	1	1	1	1	1	1
WaSC4	1	1	1	1	1	1	1	1
WaSC5	1	1	1	1	1	1	1	1
WaSC6	0.883	0.977	0.991	0.986	0.904	0.844	0.981	0.860
WaSC7	1	1	1	1	1	1	1	1
WaSC8	0.874	0.998	1	0.998	0.876	0.863	0.997	0.866
WaSC9	1	1	1	1	1	1	1	1
WaSC10	0.976	0.997	0.997	1	0.979	0.969	0.998	0.971

As with the unadjusted results, the DEA scores highlight the finding that most of the inefficiencies are due to dynamic efficiency. All of the efficiency scores increase when taking into account differences in the operating environment. The incorporation of environmental variables decreases the average overall efficiency score by 0.023 points and the maximum increase is 0.095 points.

The largest increase in efficiency scores relates to the dynamic component. This implies that the environmental variables have the largest effect in equalising the playing field for quasi-fixed inputs. This is to be expected as a proportion of the differences in companies capital will be explained through the differences in the exogenous operating conditions.

When accounting for the environmental variables there are no changes in the firms' rankings. However, the dispersion of efficiency is reduced with the range falling from 22% to 14%. There are several potential reasons why the ranks of the firms do not vary when accounting for quality. Firstly, as separate frontiers are examined for each period and capital is incorporated over time, there is less volatility amongst the efficiency scores. Secondly, due to the adjustment process, the dependent variable in the second-stage regression is the under and over-utilisation of inputs and the least efficient environment is assumed to be with the highest predicted over-utilisation of inputs. It would be advantageous to develop an adjustment process to examine the technical efficiency and allocative efficiency separately whilst generating an environmental adjusted DEA efficiency score. Overall, the results indicate that when taking into account intertemporal relationships the main contributor to overall inefficiency is the dynamic efficiency. The optimality conditions highlight the persistent over-use of quasi-fixed inputs. These results are consistent with those of Stone and Webster (2004) and Bottasso and Conti (2009) who also find a presence of over capitalisation within the industry. According to Bottasso and Conti (2009), our findings can be interpreted as the result of an Averch-Johnson effect. This is due to the presence of rate of return regulation, as well as the nature of the industry where infrastructure is built to meet future demand.

When examining the trends of inefficiencies over the period there appears to be a large improvement of the efficiency of variable inputs, whereas the inefficiencies of the capital stock are more persistent. This could be as a result of the differing incentive rates between opex and capex and rate of return regulation, which overall indicate a presence of capex bias. The inclusion of environmental variables allows firms to be considered when taking into account differences in the operating environment. The inclusion of environmental variables increases the majority of efficiency scores, dampening the magnitude of the perceived capex bias.

To account for the cost of the capex bias to the industry, table 7.6 reports the monetary saving per property if the capex bias is eliminated. This is calculated as the difference between total operating costs and the operating costs if capital was at its optimal value. This is derived from the dynamic DEA equation after adjusting for environmental variables and divided by the number of properties shown in equation 7.16.

$$optk_t = \frac{\left(\gamma^t(w_t'x_t + v_t'k_{t-1}^*) - \gamma^t(w_t'x_t + v_t'k_{t-1})\right)}{Properties_t} \qquad (7.16)$$

The result for equation 7.16 reported in table 7.6 indicates that the cost of the over-utilisation of capital per property served ranges from zero to £36. The cost is negative for WaSC 10 for the period 1998–2000 and this indicates that the firm under-utilises capital in those periods when accounting for the operating environment. Table 7.7 shows the percentage of total costs which is attributed to the over-utilisation of capital. The cost of the over-utilisation of capital for WaSC 8 is equivalent to 17% of its total costs. The cost of excess capital for the average WaSC is £8.85 per property which is equivalent to 4.5% of total costs. If Ofwat were able to eliminate the presence of the capex bias this could reduce average total costs by 4.5%, which would have major implications on customers' bills.

Table 7.6- Cost of the over-utilisation of capital £ per property billed

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Average
WaSC1	33.4	22.9	18.6	11.8	12.8	0.5	5.2	5.5	7.1	6.6	13.2	13.2	10.4	9.9	12.2
WaSC2	28.8	23.4	16.0	10.4	13.4	17.2	19.6	17.8	23.7	24.6	30.5	23.9	17.2	18.6	20.3
WaSC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaSC4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaSC5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaSC6	36.2	29.0	22.2	19.6	16.5	15.2	11.9	9.8	9.8	11.3	9.6	10.1	6.5	6.0	15.3
WaSC7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaSC8	33.9	29.2	36.6	38.2	35.0	35.5	33.4	33.7	37.1	34.9	32.0	32.3	32.2	37.1	34.4
WaSC9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaSC10	18.4	-0.5	-0.5	-0.5	0.0	3.9	5.4	5.6	7.0	6.6	11.3	10.0	9.8	11.3	6.3

Table 7.7 - Proportion of the cost of over-capitalisation to total costs

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Average
WaSC1	11.8%	8.3%	8.2%	5.3%	6.1%	0.3%	3.1%	3.4%	4.7%	4.5%	8.7%	8.8%	7.9%	8.0%	6.4%
WaSC2	8.0%	6.6%	5.5%	3.8%	5.2%	6.7%	7.9%	7.6%	10.5%	11.1%	13.4%	11.0%	10.0%	11.4%	8.5%
WaSC3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
WaSC4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
WaSC5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
WaSC6	14.7%	12.5%	11.7%	11.5%	10.0%	9.8%	7.9%	7.1%	7.6%	8.6%	7.0%	7.6%	5.9%	5.6%	9.1%
WaSC7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
WaSC8	13.2%	10.7%	14.9%	16.6%	16.1%	17.2%	17.2%	17.9%	19.9%	19.1%	17.9%	18.6%	20.7%	24.2%	17.4%
WaSC9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
WaSC10	6.6%	-0.2%	-0.3%	-0.2%	0.0%	2.1%	3.0%	3.4%	4.4%	4.2%	7.2%	6.7%	7.5%	9.1%	3.8%

7.6 Conclusion

This chapter has evaluated the effect of dynamic DEA in the English and Welsh water and sewerage industry. A two output model has been constructed for the ten WaSCs within the industry for the period 1996/97–2010/11. Environmental variables have been considered within the analysis in order to account for those firms that operate under relatively favourable or unfavourable environments.

The estimates show that the main estimated inefficiencies are due to the quasi-fixed inputs. The overall inefficiencies within the industry range from 0 to 22%. The optimality conditions show the persistent over-utilisation of quasi-fixed inputs. Our results indicate that the inefficiencies from variable inputs have improved over the period, whereas those from capital stock remain persistent. These results are consistent with the Averch-Johnson effect created by the presence of rate of return regulation. Overall the results infer the presence of a capex bias within the industry.

The incorporation of environmental variables allows for firms to be considered on a level playing field. The effect of adjusting the input slacks for those operating in unfavourable conditions improves the efficiency scores of the majority of firms. The environmental variables impact significantly upon capital differences within the industry, thereby reducing the magnitude of any capex bias. The presence of the capex bias drives the heart of the regulator by not encouraging the optimal use of inputs. The presence of a capex bias leads to intergeneration distortions for consumers bills. A pound spent on opex is fully recovered in the year in which it is spent whereas for a pound spent on capex only a proportion is recovered as capex is added to the regulatory capital value and only the rate of return and an element for depreciation charge is recovered from customers annually for the life of the asset. An inappropriate preference for capex reduces costs to be recovered within the current period and hence lowers consumer costs in the short term at the expense of higher bills for future generations. The elimination of the capex bias could reduce the average WaSCs total costs by 4.5%, equivalent to £8.85 for each property billed. Therefore the presence of the capex has implications for the consumer's bill, influencing the K factor within the price cap. This model suffers from a dimensionality problem, therefore the efficiency scores are biased upwards. Therefore the estimate of the impact of the capex bias is a best case estimate. Within the 2014 price review Ofwat aims to eliminate the capex through by equalising out the incentives rate for opex and capex by earning a rate of return based on a proportion of the total expenditure instead of solely capital expenditure.

8. Overview and Implications

8.1 Overview

This study explores the impact of regulation upon the English and Welsh water and sewerage industry. The motivation arose as a direct result of the nature of the industry which is characterised by a series of regulated regional monopolies in which the majority of consumers cannot choose their supplier. The research set out to answer five research questions.

Research Question 1: Did the 1999 and 2004 price reviews improve efficiency?

Research Question 2: Did the English and Welsh water and sewerage industry exhibit convergence in terms of efficiency performance in the period 1997–2011?

Research Question 3: How can the measure of efficiency using DEA incorporate non-discriminatory factors to allow differences in the local environment?

Research Question 4: How should long-life and indivisible capital stock be treated in the measurement of efficiency?

Research Question 5: Is there a capex bias within the industry?

Question 1 is answered in chapter 6 which finds that the 1999 review significantly improved efficiency whereas the 2004 review had no impact. Question 2 is examined in chapter 5 which reports a significant rate of convergence to a common steady state efficiency level for variable costs at a comparable rate implied by price review. Chapter 6 applies a three-stage DEA approach to allow for the incorporation of environmental variables in DEA to address question 3.When accounting for the differences in the operating environments, the efficiency score and ranks of the WaSCs vary significantly. To answer question 4, chapter 7 applies a dynamic DEA model that highlights the importance of incorporating the intertemporal nature of capital. Finally, question 5 is answered in chapter 7 which reports a persistent over-utilisation of capital over the period, finding the presence of a capex bias.

8.2 Implications

There is a large body of literature within the industry that examines the effectiveness of privatisation and regulation. This thesis contributes to the literature by examining the convergence in efficiency scores, the impact of the 1999 and 2004 price reviews, and the presence of a capex bias. Bottasso and Conti (2003) and Saal et al (2007) report an improvement in the minimum efficiency scores after privatisation. Saal et al (2007) and Portela et al (2011) both examine the average efficiency change within the industry, reporting a negligible impact as a result of regulation. This thesis contributes towards the literature by finding that the rate of growth of the efficiency scores is significantly larger for inefficient firms, implying convergence within the industry. Within the price review, Ofwat sets higher efficiency targets for those firms which are deemed as inefficient. Oxera (2005) highlights that within Ofwat's 2004 assessment of efficiency, a substantial number of companies are in the top efficiency bands compared with ten years earlier where the spread of relative efficiencies was much larger. The 8% a year rate of β -convergence implied by unconditional β-convergence is of a similar magnitude to that implied within the regulatory framework. The results conclude that the regulatory framework has been effective in encouraging convergence within the industry. However, regulation has had only a minor influence on the average efficiency change.

This thesis contributes to the literature by examining whether the 1999 and 2004 price reviews have been significant in improving efficiency. To knowledge, Portela et al (2011) is the only study to examine the 2004 review in which they indicate that productivity has fallen over the period of the 2004 review. However, it does not determine whether this is due to the price review. The results presented here indicate that that 1999 price review significantly reduced the over-utilisation of inputs, whereas the impact of the 2004 price review was insignificant. Utility Week (2014) interviewed Sir Ian Byatt, the former director general of Ofwat, who assessed the price reviews to date and stated "1994, tough; 1999, tough; 2004, generous; 2009, not as tough as it could have been had people known what would happen to interest rates". The findings that the 1999 price review induced significant improvements in efficiency and the 2004 price review had no effect on efficiency coincide with the industry's perceptions of the strengths of the price reviews.

There are a number of advantages associated with the application of DEA to the water and sewerage industry. The principle advantage of DEA is that it does not require assumptions with regards to the functional form. Another is that it can decompose overall inefficiency into its technical and allocative components. This thesis illustrates the use of a three-stage DEA model for the measurement of efficiency within the English and Welsh water and sewerage industry. The three-stage DEA incorporates non-discretionary variables within the measurement of efficiency to account for differences in the WaSCs' operating environments. The 2014 price review applies a totex approach to measuring efficiency through a triangulation approach of random effects and COLS models for the measurement of efficiency to predict the future baseline expenditure. The findings of this study suggest that the three-stage approach could be used in conjunction with the parametric approach or as a robustness check within the price review.

The final contribution is the identification of the perceived capex bias within the industry. The capex bias is the preference of capital expenditure over operating expenditure. The presence of a perceived capex bias is acknowledged by Ofwat (2011b). CEPA (2012) highlight that the capex bias may be a result of differing incentive rates, the Averch-Johnson effect and the culture of the industry which is focused on capital solutions. If the capital inefficiencies were eliminated this would amount to an average reduction of the total costs by 4.5%.

8.3 State of play

The measurement of efficiency is a key component within the price review for determining the level of expenditure an efficient company is required to finance its functions. Ofwat measures comparative efficiency to determine the level of inefficiencies, and therefore the amount in which costs can be reduced for the next five years. Efficiency is examined within this thesis to examine the historical trends and to examine whether regulation has been effective in encouraging efficiency and the correct combination of inputs. Although the techniques employed within this thesis and those of Ofwat differ in terms of the methodology, the level at which efficiency is examined and inputs and outputs, a top level comparison can be made.

Ofwat examined operating efficiency through a series of functional level OLS models. In PR94 efficiencies ranged from band A to band E, where band E is 35% away from the frontier whereas in PR09 the banding ranged from A to C, 30% from the frontier. The level of inefficiencies reported within the variable DEA function in chapter 5 range from 34.5% from the frontier in 1997 to 29.1% in 2011, which are of a similar magnitude of those reported by Ofwat. These results are also consistent when taking into account differences within the operating environment, although the rankings of companies differ when accounting for the operating characteristics. The measurement of β-convergence reports a reduction in the dispersion of the efficiency scores and convergence, with the least efficient firm catching up with the most efficient firm. The catch-up factor within the price reviews is determined by closing 60% of the gap towards the frontier company, equivalent to 12.5% a year. Unconditional β-convergence implies a rate of convergence of 8.8% a year which implies that the regulation has been effective at encouraging convergence at a similar rate. The results of convergence are consistent with those reported by Ofwat who report an improvement in the efficiency scores over the price reviews. Ofwat state that as a result of the 1999 and 2004 price review there was a clustering towards the best performance. As a result of the 1999 price review in 2004 all of the companies are in the top three relative efficiency bands compared with half in 1999 (Ofwat, 2004)

The results are this thesis are consistent with Ofwat's view of a perceived capex bias (Ofwat, 2011). Ofwat's current approach to determining the required revenue is to examine the opex and capex separately. Opex is recovered pound for pound whereas capex is added to the regulatory capital value which earns a rate of return. CEPA (2012) denote that the difference incentives between opex and capex and the Averch-Johnson effect as a result of the rate of return regulation has led to a preference towards capital expenditure within the industry. Ofwat (2011b) highlight the presence of a perceived capex bias, therefore the 2014 price review is partly designed around eliminating the capex bias through a total expenditure (totex) approach. The totex approach examines opex and capex together and companies earn a rate of return on a percentage of the totex, equalising the incentive rates between opex and capex. The application of dynamic DEA reveals an over-utilisation of capital and a reduction in the over-utilisation of other inputs, indicating a capex bias. Although the model is able to distinguish that there is a persistent over-utilisation of capital within the industry it is unable to determine whether this is due to financial incentives or as a result of companies' preference for capital solutions. To determine the extent of which the capex bias is as a result of the financial incentives it would be desirable to re-examine the presence of the capex bias at the end of the 2014 price review period. If Ofwat's approach to the eliminating the capex bias has been successful one would expect the over-utilisation of capital to fall over the period. However if the over-capitalisation of capital is still prevalent then the capex bias may exist as a result of the nature of the industry preferring capital solutions.

8.4 Future Research

A limitation of the three-stage DEA is the separability condition in the second-stage regression. The separability condition only allows for the environmental variables to impact the distribution of inefficiencies, which effects their mean and variance. The separability condition, however, does not influence the optimal choices between the discretionary inputs. Badin et al (2012) highlight that the environmental variables may influence the distribution of the efficiency scores, and influence the shape of the frontier, or the environmental variables may be completely independent. To allow for the environmental variables to influence both the shape of the frontier and the distribution of the efficiency scores, Cazals et al (2002) and Darario and Simar (2005) introduce conditional DEA. Conditional DEA measures efficiency where the input and output set may depend on the value of environmental variables. The research could be extended by applying conditional DEA for the measurement of efficiency within the English and Welsh water and sewerage industry and contrasting the results found within the three-stage DEA approach.

The current approach to modelling efficiency determines the level of efficiency for each firm for each time period. In a perfectly competitive environment firms cannot exist in the presence of long-run technical efficiencies. Tsionas (2006) states that technical inefficiency may be as a result of factors under the influence of the firm that cannot be adjusted without incurring adjustment costs. If the adjustment cost is sufficiently high, this may result in persistent technical efficiency. Chapter 7 highlights the persistent over-utilisation of capital inputs over the period. Tsionas and Kumbhakar (2012) propose a model that decomposes the overall error term into a noise component, a persistent technical efficiency component, a short-run technical efficiency and a firm specific random effects component to capture heterogeneity. The decomposition of the error term is informative for the English and Welsh water and sewerage industry to separate the company-specific effects and the persistent inefficiency. This allows the regulator to analyse the level of inefficiency which can be eliminated with relative ease.

The measurement of dynamic DEA implies that there is a preference for capital expenditure over operating expenditure. The 2014 price review aims to eliminate the potential capex bias by using a totex approach. The efficient level of expenditure is examined for operating expenditure and capital expenditure together as much is feasible. This should level out the incentive rates between opex and capex. Within the 2014 price review, companies earn a rate of return on a proportion of their total operating expenditure rather than on capital expenditure. For future research, it would be of interest to examine the impact of the changes in the regulatory framework in eliminating the capex bias or whether the capex bias exists due to the nature of the industry building for future demand.

Dynamic DEA examines relative efficiency, the results therefore imply that five of the WaSCs have a capex bias relative to the frontier companies. This result would therefore indicate that five of the companies do not exhibit a capex bias. As the regulatory incentives are the same amongst the firms then one may expect that the frontier companies exhibit a capex bias. Therefore it is expected that the absolute capex bias will be larger than the relative capex bias. It would be of interest for future research to determine the level of the absolute capex bias. Alongside measuring the absolute capex bias, the thesis can be extended to reduce the issue of dimensionality. Dynamic DEA exhibits a dimensionality problem as the data is required to be examined for each period to take into account the intertemporal nature of capital. The dimensionality issue biases the efficiency score upwards, therefore biasing the estimation of the capex bias downwards. An extension of this work would be to examine a window analysis approach whereby several years of data are pooled together and examined under a meta-frontier and the window is rolled forwards. This approach would reduce the dimensionality problem, however it makes the implicit assumption that the capital stock remains fixed over the length of the window.

The future of the industry faces substantial change due to the facilitation of Ofwat for upstream and downstream competition. Within the PR14 price review, Ofwat is facilitating the introduction of retail competition for non-household customers by 2017, following the introduction of retail competition in Scotland. The introduction of competition introduces scope for additional research to evaluate firstly whether non-households switch their suppliers, and secondly the effectiveness of competitive pressure to stimulate efficiency improvements, and therefore lower bills.

8.5 Take home message

There are three take home messages from this study:

- Firstly, the elimination of the capex bias could reduce the average company's costs by 4.5% a year
- Secondly, the three-stage DEA model should be implemented within the price review as a robustness check of existing methods
- Thirdly, the modelling of the optimal level of capital for efficiency analysis is best determined within a dynamic framework.

Appendix1:Literature Summary

Study	Data Sample	Inputs	Ouputs	Environmental/Non-Discretionary
Hunt and Lynk (1995)	English and Welsh water and sewerage Industry	Operating Expenditure, Labour Costs	Distribution Input, Trade Effluent, Environmental Services	
Cubbin and Tzanidakis (1998)	English and Welsh water and sewerage Industry	Opex	Water Delivered, Length of Mains, Proportion of distribution to non-households	Proportion of water delivered to non-households
Thanassoulis (1999)	English and Welsh water and sewerage Industry	Opex less power costs	Properties, Length of Mains, Water Delivered	
Ashton (2000a)	English and Welsh Water and Sewerage Industry	Labour, consumables, other	Number of connected households	
Saal and Parker (2000)	English and Welsh Water and Sewerage Industry	Labour, consumables, other	Number of connected households	
Garcia and Thomas (2001)	55 French Water Municipals	Labour, Electricity, Materials and Other	Volume sold to final customers, Water losses	Number of customers, number of municipals served, Network length, production capacity, stocking, pumping capacity
Saal and Parker (2001)	English and Welsh water and sewerage industry. WaSCs	Labour, Capital and Other Costs	Water supply population and population connected to sewerage treatment work.	
Thanassoulis (2002)	English and Welsh Water and Sewerage Industry	Sewerage OPEX	Length of Sewers, Area served, capacity of pumping in the sewerage network	
Stone and Webster (2004)	English and Welsh Water and Sewerage Industry	Labour, Power, Capital and Other	Volume of water delivered, Properties connected for water supply, number of	

			properties connected for sewerage, Equivalent population served	
Tupper and Resende (2004	Brazil water industry	Labour Expenses, Operational costs, Other operational Costs	Water Produced, Treated Sewage, Population served water, Population served-treated sewage	Water Density, Sewerage Density, Water loss,
Woodbury and Dollery (2004)	73 Australian Water municipals	Management Costs, Maintenance and operation costs, energy and chemical costs, capital replacement costs	Number of assessment, Annual water consumption	Water Service index, Population, properties per km of mains, location, large seasonal variations, rainfall, percentage of residential assessment, water is filtered or not, groundwater is used
Aubert and Reynuad (2005)	Wisconsin Water Utilities	Labour, energy, chemicals, operation supplies and expenses and maintenance, capital	Volume of water, Number of customers served	Dummies for firms which purchase water from another utility and use surface water, average depth of pumping wells, dummies for rate of return regulation and hybrid regime.
Coelli and Walding (2005)	Australian Water Companies	Operating Expenditure, Capital (Length of Mains)	Number of properties connected, volume of water delivered	
Saal and Parker (2005)	English and Welsh Water and Sewerage Industry	Capital stock, opex	Water supply, water connections	Network density, Aph, water quality, Dummy variable for WaSCs
Garcia-Sanchez (2006)	Spanish Water Municipals	Total Staff, Treatment Plants, Pipes	Water Supplied, Connections, controls or analyses carried out.	Population, tourist index, number of houses, average people per house, meter squared of greenbelt, economic activity, average temperature, municipals area and population density
Erbetta and Cave (2007)	English and Welsh Water and Sewerage Industry (WaSC)	Capital, Labour, Other	Potable and non potable water* DWI Quality , Water properties, Sewerage	Proportion of boreholes, Water Loss, Water Density, Sewerage Density,

			properties, Physical amount of waste water*River quality	Proportion of trade effluent, Reg94, Reg99
Saal et al (2007)	English and Welsh water and sewerage Industry. WaSCs	MEA, Employees, other costs	Water Population* DWI Quality, Sewerage Population* (Weighted average Bathing water and river quality), Water supplied, Equivalent Population	% Underground water, Trade Effluent, Bathing Water Intensity, % Metered Properties
Picazo-Tadeo et al (2008)	38 Spanish Water municipals	Delivery network, sewer network, labour(number of workers), operational costs	Population served, water delivered and treated sewage	unaccounted-for water
Bottasso and Conti (2003)	English and Welsh Water and Sewerage Industry	Opex, labour	Water Delivered, length of main, average pumping head, proportion of river sources on total sources	Average Pumping Head, Share of water delivered to non-households, Population density, Distribution Input, MEA, proportion of river sources on total sources
Saal and Reid (2004)	English and Welsh Water and Sewerage Industry	Labour, Other, MEA Water, MEA Sewerage	Water Delivered, Equivalent Population Served	Drinking Water Quality, Population receiving secondary sewerage treatment, Properties with water pressure below the reference level, Water Properties, Sewerage Properties, Sewerage Density, Water Density
Bottasso and Conti (2009)	English and Welsh Water and Sewerage Industry (WoCs Only)	Labour, capital, other	Water Delivered, Connected Properties, area size of WoC	Aph, Distribution input from River, Water Quality, % of properties did not experience pressure problems and % of properties that did not experience service interruptions.
Bottasso and Conti (2011)	English and Welsh Water and Sewerage Industry	Labour, Other, Capital	Water Delivered, Equivalent Population served.	

Saal et al (2011)	English and Welsh Water and Sewerage Industry	Capital, Labour, Energy and Other	DI, Water Connected Properties, Sewerage Connected Properties, Equivalent Population	Water Density, Sewerage Density, DI from Boreholes, per capita water consumption, Distribution loss, Water Quality, Sewerage Quality
Portela et al (2011)	English and Welsh Water and Sewerage Industry	Opex	Number of billed Properties, Adjusted distribution input surface water, Adjusted distribution input non-surface water, Number of sources, Number of billed properties adjusted	
Maziotis et al (2012a)	English and Welsh Water and Sewerage Industry	Capital, Labour and other	Quality Adjusted Water Properties, Quality Adjusted Sewerage Properties	
Maziotis et al (2012b)	English and Welsh Water and Sewerage Industry.	Capital, Labour and Other	Water connected properties, Sewerage connected properties	
Maziotis et al (2013)	English and Welsh Water and Sewerage Industry	Capital, Labour, Other	Quality Adjusted Water Connected Properties, Quality Adjusted Sewerage Connected Properties	Water Quality, Sewerage Quality
Cherchy et al (2013)	English and Welsh Water and Sewerage Industry	Resources, power, capital, labour and other	Water Properties, Sewerage Properties, Volume of Water, Volume of Sewerage	Service area, Leakage, Prop of River water, Prop of ground water, Bulk supply imports, bulk supply exports, Water Connections, Sewerage Connections

Appendix 2- SYS-GMM

This appendix reports the results for the SYS-GMM of Blundell and Bond (1998) model in chapter 5 for equation 5.7. The SYS-GMM is a dynamic panel model which overcomes the problem of weak instruments in the GMM-DIFF model. However, due to data limitations the number of groups is less than the number of instruments, therefore the results are not robust.

	Total Costs GMM-SYS	Variable Costs GMM-SYS
Intercept	-0.0156	-0.0144073
	-0.0136	-0.0092708
$\ln \theta_{t-1}$	0.8869***	0.8775684***
	(0.0870)	(0.0707451)
Convergence Speed	-0.1131	-0.1224
Number of Observations	140	140
Sargan Test	31.00	37.04**
AR(1)	-2.54**	-2.36**
AR(2)	0.65	6.77
Half life	5.78	5.31

Notes: Significant levels *, 10%, **5%, ***1%. Standard erros in parentheses.
Estimations are perfomed by xtabond2 in STATA by Roodman (2009)

References

Afrait, S.N., 1972. Efficiency Estimation of Production Functions. *International Economic Review* 13, pp. 568–598.

Aigner, D.J., Lovell, C.A.K. and Schmidt, P., 1977. Formulation and Estimation of Frontier Production Function Models. *Journal of Econometrics* 6(1), pp. 21–37.

Ali, I. and Seiford, L., 1990. Translation invariance in data envelopment analysis. *Operational research Letters* 9, pp. 403–405.

Anwandter, L. and Ozuna, T., 2002. Can public sector reforms improve the efficiency of public water utilities? *Environment and Development Economics* 7, pp. 687–700.

Arbia, G. and Piras, G., 2005. *Convergence in per-capita GDP across European regions using panel data models expended to spatial autocorrelation effects.* Instituto Di Studi E Analisi Economica Working Paper no. 51.

Arenallo, M. and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, pp. 277–297.

Ashton, J. K., 2000a. Cost efficiency in the UK water and sewerage industry. *Applied Economic Letters* 7, pp. 455–458.

Ashton, J. K., 2000b. Total Factor Productivity Growth and Technical Change in the Water and Sewerage Industry. *The Service Industries Journal* 20(4), pp. 121–130.

Aubert, C. and Reynaud, A., 2005. The Impact of Regulation on Cost Efficiency: An Empirical Analysis of Wisconsin Water Utilities. *Journal of Productivity Analysis* 23(3), pp. 383–409.

Averch, H. and Johnson, L., 1962. Behaviour of the Firm Under Regulatory Constraint. *The American Economic Review* 52(2), pp. 1052–1069.

Avkiran, N.K. and Rowlands, T., 2008. How to better identify the true managerial performance: State of the art using DEA. *Omega* 36(20), pp. 317–324.

Badin, L., Darario, C. and Simar, L., 2012. How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research* 223(1), pp. 818–833.

Banker, R.D., 1993. Maximum likelihood consistency and data envelopment analysis: A statistical foundation. *Management Science* 39, pp. 1265–1273.

Banker, R.D., Charnes, A. and Cooper, W.W., 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30(9), pp. 1078–1092.

Banker, R. D. and Morey, R. C., 1986. Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research* 34(4), pp. 513–521.

Barro, R.D., 1991. Economic growth in a cross section of countries. *Quarterly Journal of Economics* 106, pp. 407–443.

Barro, R.D. and Sala-i-Martin, X., 1991. Convergence across states and regions. *Brooking Papers on Economic Activity* 1, pp. 107–182.

Barro, R.D. and Sala-i-Martin, X., 1992. Convergence. *Journal of Political Economy* 100, pp. 223–251.

Barro, R.D. and Sala-i-Martin, X., 1995. *Economic Growth.* New York: McGraw Hill.

Barro, R.J. and Sala-i-Martin, X., 2004. *Economic Growth*. 2nd ed. London: MIT Press.

Barro, R.J., 2012. *Convergence and Modernization Revisited.* NBER Working Paper No. 18295.

Battese, G.E. and Corra, G.S., 1977. Estimation of a Production Frontier Model: With Application to the Pastoral Zone off Eastern Australia. *Australian Journal of Agriculture Economics* 21(3), pp. 169–179.

Beesley, M., 1994. *Regulating Utilities: the Way Forward.* Institute of Economic Affairs. London.

Beesley, M. E. and Littlechild, S. C., 1989. The Regulation of Privatized Monopolies in the United Kingdom. *The RAND Journal of Economics* 20(3), pp. 454–472.

Berger, A. N., 1993. Distribution Free Estimates of Efficiency of the U.S. Banking Industry and Tests of the Standard Distribution Assumptions. *Journal of Productivity Analysis* 4(3), pp. 261–292.

Bernstein, J.I. and Sappington, D.E., 2000. How to determine the X in RPI-X regulation: a user's guide. *Telecommunication Policy* 24, pp. 63–68.

Blank, J. and Valdmanis, V., 2005. A Modified Three-Stage Data Envelopment Analysis: The Netherlands. *The European Journal of Health Economics* 6(1), pp. 65–71.

Blundell, R. and Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, pp. 115–143.

Bond, S., Hoeffler, A. and Jonathan, T., 2001. *GMM Estimation of Empirical Growth Models.* Centre for Economic Policy Research Working Paper No. 3048.

Bottasso, A. and Conti, M., 2003. *Cost Inefficiency in the English and Welsh Water Industry: An Heteroskedastic Stochastic Cost Frontier Approach.* Economic Discussion Papers, University of Essex.

Bottasso, A. and Conti, M., 2009. Scale economies, technology and technical change in the water industry: Evidence from the English water only sector. *Regional Science and Urban Economics* 39, pp. 138–147.

Bottasso, A., Conti, M., Piacenza, M. and Vannoni, D., 2011. The appropriateness of the poolability assumption for multiproduct technologies: Evidence from the English water and sewerage utilities. *International Journal of Production Economies* 130(1), pp. 112–117.

Bowlin, W.F., 1998. Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *Journal of Cost Analysis* 15(2), pp. 3–27.

Casu, B. and Girardone, C. 2010. Integration and efficiency convergence in EU banking markets. *Omega* 38, pp. 260–267.

Caves, D. W., Christensen, L. R. and Swanson, J. A., 1981. Productivity growth, scale economies and capacity utilization in US railroads, 1955–74. *American Economic Review* 71, pp. 994–1002.

Cazals, C., Forens, J. P. and Simar, L., 2002. Nonparametric Frontier estimation: A robust approach. *Journal of Econometrics* 106, pp. 1–25.

CEPA., 2012. *Incentives and Menus*. Cambridge Economic Policy Associates Ltd.

CEPA, 2014. *Cost Assessment Advantaged Econometric Models.* Cambridge Economic Policy Associated Ltd.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L. and Stutz., 1985. Foundations of data envelopment analysis and Pareto-Koopmans empirical production functions. *Journal of Econometrics* 30, pp. 91–107.

Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M., 1994. *Data Envelopment Analysis: Theory, Methodology, and Applications.* Boston: Kluwer Academic Press.

Charnes, A., Cooper, W. W. and Rhodes, E., 1978. Measuring Efficiency of Decision Making Units. *European Journal of Operational Research* 3, pp. 429–444.

Cherchye, L., Demuynck, T., De Rock, B. And De Witte, K., 2013. Non-parametric Analysis of Multiple-Output Production with Joint Inputs. *The Economic Journal* 124, pp. 735–775.

Christensen, L., Jorgenson, D. and Lau, L., 1973. Transcendental Logarithmic Production Frontier. *The Review of Economic and Statistics,* 55, pp. 28–45.

Cobb, C.W. and Douglas, P.H., 1928. A Theory of Production. *American Economic Review,* 18, pp. 177–191.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G. E., 2005. *An Introduction to Efficiency and Productivity Analysis.* 2nd Edition. New York: Springer Science+ Business Media.

Cooper, W.W., Seifor, K.M. and Tone, K., 2006. *Introduction to Data Envelopment Analysis and Its Uses.* Springer, New York.

Cordero-Ferrera, J. M., Pedraja-Chaparro, F. and Santin-González, D., 2010. Enhancing the inclusion of non-discretionary inputs in DEA. *Journal of the Operational Research Society* 61, pp. 574–584.

Cowan, S., 2002. Price-cap regulation. *Swedish Economic Policy Review* 9, pp. 167–188.

Cowan, S., 2003. 'The Water Industry' in Helm, D. and Jenkinson, T. (eds.), *Competition in Regulated Industries.* New York: Oxford University Press.

Cowing, T. G. and Howing, A. A., 1983. Multiproduct short run hospital cost functions: empirical evidence and policy implication from cross section data. *Southern Economic Journal* 49, pp. 637–653.

Cubbin, J., 2005. Efficiency in the water industry. *Utilities Policy* 13, pp. 289–293.

Cubbin, J. and Tzanidakis, G., 1998. Regression versus data envelopment analysis for efficiency measurement: an application to the England and Wales regulated water industry. *Utilities Policy* 7, pp. 75–85.

Daraio, C. and Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis* 24(1), pp. 23–121.

Debreu, G., 1951. The Coefficient of Resource Utilization. *Econometrica* 19(3), pp. 273–292.

Dieter, H. and Rajah, N., 1994. Water regulation: the periodic review. *Fiscal Studies* 15, pp. 74–94.

Diewert, W.E., 1971. An Application of the Shephard Duality Theorem: A Generalised Leontief Production Function. *Journal of Political Economy,* 79(3), pp. 481–507.

Dyson, R., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., and Shale, E.A., 2001. Pitfalls and protocols in DEA. *European Journal of Operational Research* 132(2), pp. 245–259.

Emrouznejad, A. and Thanassoulis, E., 2005. A mathematical model for dynamic efficiency using data envelopment analysis. *Applied Mathematics and Computation* 160, pp. 363–378.

Erbetta, F. and Cave, M., 2007. Regulation and Efficiency Incentives: Evidence from the England and Wales Water and Sewerage Industry. *Review of Network Economics* 6(4), pp. 425–452.

Fare, F. and Lovell, C.A.K., 1978. Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1), pp. 409–428.

Fare, R. and Grosskopf, S., 2000., Network DEA. *Socio-Economic Planning Sciences* (34), pp. 35–49.

Fare, R., Grosskopf, S. and Logan, J., 1983. The Relative Efficiency of Illinois Electric Utilities. *Resources and Energy* 5, pp. 349–367.

Fare, R., Grosskopf, S., and Lovell, C.A.K., 1985. *The Measurement of Efficiency of Production.* Boston: Kluwer- Nijhoff Publishing.

Fare, R., Grosskopf, S., Norris, M. and Zhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review* 84(1), pp. 66–83.

Fare, R. and Primont, D., 1995. *Multi-Output Production and Duality Theory and Applications.* Boston: Kluwer Academic Publishers.

Farrell, M.J., 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society series A* 120(3), pp. 253–290.

Fried, H. O., Schmidt, S. S. and Yaisawarng, S., 1999 Incorporating of the Operating Environment Into a Nonparametric Measure of Technical Efficiency. *Journal of Productivity Analysis* 12 (3), pp. 249–267.

Fried, H. O., Lovell, C. A. K., Schmidt, S. S. and Yaisawarng, S., 2002. Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. *Journal of Productivity Analysis* 17, pp. 157–174.

Fried, H. O., Lovell, C. A. and Schmidt, S. S., 2008. 'Efficiency and Productivity' in Fried, H.O., Lovell, C. A. and Schmidt, S.S. (eds.), *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

Friedman, M., 1992. Do old fallacies ever die? Journal of Economic Literature 30 (4), pp. 2129–2132

Fung, M. K., 2006. Scale economies, X-efficiency, and convergence of productivity among bank holding companies. *Journal of Banking & Finance* 30, pp. 2857–2874

Garcia, S. and Thomas, A., 2001. The Structure of Municipal Water Supply Costs: Application to a Panel of French Local Communities. *Journal of Productivity Analysis* 16, pp. 5–29.

Garcia-Sanchez, I. M., 2006. Efficiency Measurement in Spanish Local Government: The Case of Municipal Water Services. *Review of Policy Research* 23(2), pp. 355–372.

Geymueller, P., 2009. Static versus dynamic DEA in electricity regulation: the case of US transmission system operators. *Central European Journal of Operational Research* 17(4), pp. 397–413.

Gluschenko, K., 2012. *Myths about Beta-Convergence.* Mitchigan: University of Michigan. William Davidson Institute Working Paper Series 1040.

Golany, B. and Roll, Y., 1989. An application procedure for DEA. *Omega* 17(3), pp. 237–250.

Greene, W., 1980a. Maximum Likelihood Estimation of Econometric Frontier Functions. *Journal of Econometrics* 13, pp. 27–56.

Greene, W., 1980b. On the Estimation of a Flexible Frontier Production Models. *Journal of Economics* 3, pp. 101–115.

Greene, W.H., 2008. ‘The Econometric Approach to Efficiency Analysis’ in Fried, H.O., Lovell, C. A. and Schmidt, S.S. (eds.), *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

Helm, D. and Jenkinson, T., 2003. *Competition in Regulated Industries.* Oxford University Press, New York.

Helm, D. and Rajah, N., 1994. Water Regulation: The Periodic Review. *Fiscal Studies* 15(2), pp. 74–94.

Hunt, L. C., and Lynk, E. L., 1995. Privatisation and efficiency in the UK water industry: An empirical analysis. *Oxford Bulletin of Economics and Statistics* 57, pp. 371–388.

Islam, N., 1995. Growth empirics: A panel data approach. *Quarterly Journal of Economics* 110, pp. 1127–1170.

Kneip, A., Park, B.U., and Simar. L., 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* 14, pp. 783–793.

Koopman, T.C., 1951. ‘An Analysis of Production as an Efficient Combination of Activities’ in Koopman, T.C, (ed.), *Activity Analysis of Production and Allocation.* Cowles Commission for Research in Economics, Monograph No. 13, New York: Wiley.

Korostelev, A., Simar, L. and Tsybakov, A.B., 1995. On estimation of monotone and convext boundaries. *Publications de l’Institut de Statistiques de l’Universite de Paris* 39(1), pp. 3–18.

Kumbhakar, S. and Lovell, K., 2000. *Stochastic Frontier Analysis.* Cambridge: Cambridge University Press.

Littlechild, S., 1983. *Regulation of British Telecommunications' profitability*. Report to the Secretary of State. London: Department of Industry

Littlechild, S., 1986. *Economic Regulation of Privatised Water Authorities.* HMSO: London.

Littlechild, S., 1988. Economic Regulation of Privatised Water Authorities and Some Further Reflections. *Oxford Review of Economic Policy* 4(2), pp. 40–68.

Lovell, C., Travers, P., Richardson, S. and Wood, L., 1994. 'Resources and functionings: a new view of inequality in Australia' in Eichorn, W. (eds.), *Models and Measurement of Welfare and Inequality*. Berlin: Springer- Verlag.

Lovell, C. and Pastor, J., 1995. Unit invariant and translation invariant DEA models. *Operational Research letters* 18(3), pp. 103–103.

Lynk, E. L., 1993. Privatization, joint production and the comparative efficiencies of private and public ownership: the UK water industry case. *Fiscal Studies* 14(2), pp. 98–116.

Mass-Colell, A., Whinston, M. and Green, J., 1995. *Microeconomic Theory*. Oxford: Oxford University Press.

Maziotis, A. M., Saal, D.S. and Thanassoulis, E., 2013. *A Methodology to Propose the X-Factor in the Regulated English and Welsh Water and Sewerage Companies*. FEEM Working Paper, Nota Di Lavoro, No. 14.

Maziotis, A., Saal, D. S. and Thanassoulis, E., 2012a. *Profit, Productivity and Price Performance Change in the English and Welsh Water and Sewerage.* FEEM Working Paper, Nota Di Lavoro, No. 84.

Maziotis, A., Saal, D. S. and Thanassoulis, E., 2012b. *Output Quality and Sources of Profit Change in the English and Welsh Water and Sewerage Companies.* FEEM Working Paper, Nota Di Lavoro, No. 85.

Meeusen, W. and Van Den Broeck, J., 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18(2), pp. 435–444.

Mester, L.J., 1993. Efficiency in the Savings and Loan Industry. *Journal of Banking and Finance* 17(2), pp. 267–286.

Mester, L.J. 1996. A study of bank efficiency taking into account risk-preferences. *Journal of Banking and Finance* 20, pp. 1025–1045.

Nemoto, J. and Goto, M., 1999. Dynamic data envelopment analysis: modelling intertemporal behaviour of a firm in the presence of productive inefficiencies. *Economic Letters* (64), pp. 51–56.

Nemoto, J. and Goto, M., 2003. Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities. *Journal of Productivity Analysis* (19), pp. 191–210.

Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, pp. 1417–1426.

Ofwat, 1994., *Future Charges for Water and Sewerage Services: The outcome of the Periodic Review*. Birmingham.

Ofwat, 2004. *Future water and sewerage charges 2005-10 Final Determinations*. Birmingham.

Ofwat, 2008. *The Development of the Water Industry In England and Wales*. Birmingham. Available at: <http://www.ofwat.gov.uk/publications/commissioned/rpt_com_devwatindust270106.pdf > [Accessed 30 August 2014]

Ofwat, 2009. *Future Charges for Water and Sewerage Charges 2010–15: Final Determinations.* Birmingham.

Ofwat, 2011a. *Financeability and financing the asset base: a discussion paper.* Office of Water Services, Birmingham.

Ofwat, 2011b. *Capex bias in the water and sewerage sectors in England and Wales: substance, perception or myth? A discussion paper.* Office of Water Services, Birmingham.

Ofwat, 2012. *Water Service Regulation Authority (Ofwat) Annual report and accounts 2011-12.* Office of Water Services, Birmingham.

Ofwat, 2013. *Setting price controls for 2015–20: final methodology and expectations for companies' business plans.* Office of Water Services, Birmingham.

Ouellette, P. and Yin, L., 2008. Investment and dynamic DEA. *Journal of Productivity Analysis* (29), pp. 235–247.

Oxera, 2005. Fine-tuning RPI-X: the impact of change the incentives mechanism. Oxford: Oxera.

Parikh, A. and Shibata, M., 2004. Does trade liberalization accelerate convergence in per capita incomes in developing countries?. *Journal of Asian Economics* 15, pp. 33–38.

Pastor, J.T., 1996. Translation invariance in DEA: A generalization. *Annals of Operational Research* 66, pp. 93–102.

Picazo-Tadeo, A.J., Saéz-Ferńandez, F.J. and González-Gómez, F., 2008. Does service quality matter in measuring the performance of water utilities?. *Utilities Policy* 16, pp. 30–38.

Portela, M., Thanassoulis, E., Horncastle, A. and Maugg, T., 2011. Productivity change in the water industry in England and Wales: application of meta-Malmquist index. *Journal of Operational Research Society* 62, pp. 2173–2188.

Quah, D., 1993. Galton's fallacy and test of convergence hypothesis. Scandinavian Journal of Economics 95(4), pp. 427–443.

Quah, D., 1996. Twin peaks: growth and convergence in models of distribution dynamics. *Economic Journal* 106, pp. 1045–1055.

Ray, S. C., 1988. Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-Economic Planning Sciences* 22 (4), pp. 167–176.

Ray, S. C., 1991. Resource-use efficiency in public schools: a study of Connecticut data. *Management Science* 37, pp. 1620–1628.

Ray, S.C. and Desli, E., 1997. Productivity Growth Technical Progress and Efficiency Change in Industrialized Countries: Comment. *American Economic Review* 87(5), pp. 1033–1039.

Renzetti, S. and Dupont, D.P., 2009. Measuring the Technical Efficiency of Municipals Water Suppliers: The Role of Environmental Factors. *Land Economics* 85(4), pp. 627–636.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in stata. *The Stata Journal* 9(1), pp. 86–136.

Saal, D.S. and Parker, D., 2000. The Impact of Privatization and Regulation on the Water and Sewerage Industry in England and Wales: A Translog Cost Function Model. *Managerial and Decisions Economics* 21(6), pp. 253–268.

Saal, D. S. and Parker, D., 2001. Productivity and Price Performance in the Privatized Water and Sewerage Companies of England and Wales. *Journal of Regulatory Economics* 20(1), pp. 61–90.

Saal, D. S. and Parker, D., 2005. *Assessing the performance of water operations in the English and Welsh water industry: A panel input distance function approach.* Working Paper RP0502, Aston Business School.

Saal, D. S., Arocena, P. and Maziotis, A., 2011. *The cost implications of alternative vertical configurations of the English and Welsh water and sewerage industry.* Working Paper 8, Aston Centre for Critical Infrastructure and Services, Aston University.

Saal, D. S., Parker. D. and Weyman-Jones, T., 2007. Determining the contribution of technical change, efficiency change and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985–2000. *Journal of Productivity Analysis* 28(1-2), pp. 127–139.

Saal, D. S. and Reid, S., 2004. *Estimating opex productivity growth in English and Welsh water and sewerage companies: 1993–2003.* Working Paper PR0434, Aston Business School, Ashton University.

Sala-i-Martin, X., 1996. The classical approach to convergence analysis. *Economic Journal* 106, pp. 1019–1036.

Sawkin, J. W., 1995. Yardstick competition in the English and Welsh water industry. *Utilities Policy* 5(1), pp. 27–36.

Sawkin, J. W., 2001. The development of competition in the English and Welsh water and sewerage industry. *Fiscal Studies* 22(2), pp. 189–215.

Sexton. T. R., Taggart Jr., R.E., 1994. Improving pupil transportation in North Carolina Interfaces 24, pp. 87–103.

Sengupta, J., 1995. *Dynamics of Data Envelopment Analysis.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shleifer, A., 1985. A theory of yardstick competition. *Rand Journal of Economics* 16(3), pp. 319–327.

Silva, E. and Stefanou, S., 2003. Nonparametric Dynamic Production Analysis and the Theory of Cost. *Journal of Productivity Analysis* 19, pp. 5–32.

Silva, E. and Stefanou, S., 2007. Dynamic Efficiency Measurement: Theory and Application. *American Journal of Agricultural Economics* 89 (2), pp. 398–419.

Simar, L. and Wilson, P.W., 1998. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science* 44, pp. 49–61.

Simar, L. and Wilson, P.W., 2000a. A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics* 27(6), pp. 779–802.

Simar, L. and Wilson, P.W., 2000b. Statistical Inference in Nonparametric Frontier Models: The State of the Art. *Journal of Productivity Analysis* 13, pp. 49–78.

Simar, L. and Wilson, P. W., 2002. Non-parametric tests of returns to scale. *European Journal of Operational Research* 139 (1), pp. 115–132.

Simar, L. and Wilson, P. W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136, pp. 31–64.

Simar, L. and Wilson, P., 2008. ‘Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives’ in Fried, H.O., Lovell, C. A. and Schmidt, S.S. (eds.), *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

Stanton, K.R., 2002. Trends in relationship lending and factors affecting relationship lending efficiency. *Journal of Banking and Finance* 26, pp. 127–152.

Simar, L. and Wilson, P. W., 2007. Estimation of inference in two-stage, semi-parametric model of production processes. *Journal of Econometrics* 136, pp. 31–64.

Stevenson, E., 1980. Likelihood Functions for Generalized Stochastic Frontier Estimation. *Journal of Econometrics* 13, pp. 58–66.

Stone and Webster Consultants. 2004. *Investigation into Evidence for Economies of Scale in the Water and Sewerage Industry in England and Wales.* Report for Ofwat.

Thanassoulis, E., 2000a. The use of data envelopment analysis in the regulation of UK water utilities: Water distribution. *European Journal of Operational Research* 126, pp. 436–453.

Thanassoulis, E., 2000b. DEA and its use in the regulation of water companies. *European Journal of Operational Research* 127(1), pp. 1–13.

Thanassoulis, E., 2001. *Introduction to the Theory and Application of Data Envelopment Analysis.* Dordrecht: Kluwer Academic Publishers.

Thanassoulis, E., 2002. Comparative Performance Measurement in Regulation: The Case of English and Welsh Sewerage Services. *The Journal of Operational Research Society* 53(3), pp. 292–302.

Tone, K., 2001. A slack-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130, pp. 498–509.

Tone, K. and Tsutsui, M., 2009. Tuning Regression Results for use in Multi-Stage Data Adjustment Approach of DEA. *Journal of Operations Research* 52(2), pp. 76–85.

Tone, K. and Tsutsui, M., 2010. Dynamic DEA: A slacks-based measure approach. *Omega* (38), pp. 145–156.

Tsionas, E., 2006. Inference in Dynamic Stochastic Frontier Models. *Journal of Applied Econometrics* 21, pp. 669–676.

Tsionas, E. and Kumbhakar, S., 2012. Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True Random-Effects Model. *Journal of Applied Econometrics* 29, pp. 110–132.

Tupper, H.C., and Resende, M., 2004. Efficiency and regulatory issues in the Brazilian water and sewage sector: an empirical study. *Utilities Policy* 12, pp. 29–40.

Utility Week. 2014. *25 Years of Water Privatisation.* Press Release 27 November 2014. Available at: <http://www.uw-twentyfiveyears.net/30043> [Accessed 28 November 2014]

Varian, H.R., 1992. *Microeconomic analysis.* New York: Norton.

Weill, L., 2009. Convergence in banking efficiency across European countries. *Journal of International Financial Markets, Institutions and Money.* 19, pp. 818–833.

Wilson, P. W., 2004. *A preliminary non-parametric analysis of public education and health expenditures developing countries.* Unpublished Working Paper, Department of Economics, Clemson University, Clemson, SC.

Wheelock, D.C. and Wilson, P.W., 2003. *Robust nonparametric estimation of efficiency and technical change in the U.S. commercial banking.* Unpublished Working Paper, Department of Economics, Clemson University, Clemson, SC.

Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias. *Journal of the American Statistical Association,* 57, pp. 348–368.