DEVELOPING PANEL BASED METHODS FOR PROFIT AND PERFORMANCE MEASUREMENT IN THE WATER AND SEWERAGE SECTOR IN ENGLAND AND WALES

ALEXANDROS MAZIOTIS
Doctor of Philosophy

ASTON UNIVERSITY
January 2011

ASTON UNIVERSITY

DEVELOPING PANEL BASED METHODS FOR PROFIT AND PERFORMANCE MEASUREMENT IN THE WATER AND SEWERAGE SECTOR IN ENGLAND AND WALES

ALEXANDROS MAZIOTIS
DOCTOR OF PHILOSOPHY
2011

The aim of the thesis is to evaluate the impact of regulatory price cap schemes on the financial performance of the Water and Sewerage Companies (WaSCs) even when the number of observations is small. We applied an index number approach to allow for cross sectional comparisons of relative profitability, productivity and price performance of WaSCs during the years 1991-2008. We also applied a panel index approach across WaSCs over time to decompose unit-specific (temporal) index number based profitability growth as a function of the profitability, productivity and price performance growth achieved by benchmark firms, and the catch-up to the benchmark firm achieved by less productive firms. We also employed both index numbers and DEA techniques to evaluate various profit drivers such as price changes, productivity changes and activity effect levels on the financial performance of WaSCs over time. Exogenous characteristics like water and sewerage quality were also included in a profit decomposition analysis. The results showed that during 1991-2000 price caps were "weak" as prices were high enough for the firms to achieve economic profits despite their low productivity levels. However, after 2000 prices became "catch up promoting" as they required less productive companies to eliminate at least some excess costs in order to eliminate economic losses. The steady decline in average price performance, gains in productivity and relatively stable economic profitability after 2000, suggest that Ofwat is now more focused on passing productivity benefits to consumers, and maintaining stable profitability than it was in earlier regulatory periods. The positive impact on profit changes came from substantial improvements in technical change, the cost efficient allocation of resources by substituting labour with capital and small improvements in efficiency gains and output mix. The input price and scale effect had a significant negative impact on profit changes. We suggest that our approach should be of great interest to researchers who are interested in evaluating the effectiveness of regulation and/or developing effective comparative performance techniques when sample sizes are limited.

Keywords: Profits, productivity, price performance, activity, index numbers, DEA, regulation, water industry

ACKNOWLEDGEMENTS

I would like to express my heartiest gratitude to my supervisors Dr David Saal and Professor Emmanuel Thanassoulis, for their invaluable advice, guidance and constructive suggestions throught the duration of this research. Their support, encouragement and expertise in this field have been immensely appreciated. David, I will never forget our endless meetings, your continuous concern and support all and I would not be able to finish my Phd without your help. I am deeply grateful to you for your help and you are more than my supervisor to me. I would also like to thank Professor Pablo Arocena and Dr Ozren Despic for their encouraging and constructive comments made during my Phd viva and reflected in their suggested corrections. I would also like to thank Liz Blackford, Kelly Moore and Jeanette Ikuomola for their continuous assistance throughout these years of my studies. Finally, I am grateful to the Economic and Social Research Council (ESRC) and Office of Water Services (OFWAT) for funding of this research. I am also grateful to OFWAT for providing data and occasional advice on the research. The views expressed in the thesis are those of author and no representation is made that they are shared by ESRC or OFWAT.

I would like to take this opportunity to thank those who helped and supported me throught the duration of this research. First and foremost, I am indebted to my family for their continuous concern and support. My father would be very proud of me. A special appreciation is due to my "brother" Manolis Liapis and his family, and my best friend from my Msc studies, George Gazetas who always believed in me and my abilities and encouraged me all these years. Your friendship means the world to me. I would like to thank three important persons who I had the chance to meet them during my Phd studies and supported me through both the highs and lows of the past four years, George Kourogiorgas, George Kladis and Mike Smyrnakis. You are more than friends to me and I am deeply indebted to your friendship during the darker days of this Phd. I would also like to thank my flatmate, Farhan Al-Husban, Jan Duracz and Rish for their concern and support throughout my studies. Moreover, I would like to thank Dimitris Lolos for his patience and understanding throughout the last year of my Phd study. I am deeply indebted to you for tolerating me and for this reason you are a very important person in my life. A special appreciation is due to Alexandros Iatridis for his understanding and patientce the past few months of my research and his efforts for psychoanalysing me, most of the time without success.

Finally, I would like to take this opportunity to thank the loves of my life, Margarita Kokkinaki and Dimitra "Mika" Karachaliou. Margarita, I have known you for more than ten years and you are one of the few persons in my life who are always there for me. I am deeply grateful to you for your understanding, support and patience all these years and my life would mean nothing to me without you. Mika, I met you during your Msc studies and you have a special place in my heart thanks to your wonderful personality. I am grateful to you for your understanding and patience and although I do not see you very often (my fault), it has not been a single day without thinking of you. I will always love you and my life would mean nothing to me without you.

My heart belongs to Manolis, George, Dimitris, Margarita and Mika.

To *Manolis, George, Dimitris,*
Margarita and Mika

"I do not know if I am a good writer,
but anyone reading this thesis is a good reader"

CONTENTS

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION

1.1 Background

The water and sewerage industry in England and Wales was privatized in 1989 and before privatization there were 10 Regional Water Authorities responsible for the water and sewerage supply in England and Wales and 29 Statutory Water companies, which were already privatized companies that were only responsible for the supply of water. After 1989, the 10 Regional Water Authorities were privatized and formed the Water and Sewerage Companies (WaSCs) and the 29 Statutory Water Companies became Water Only Companies (WoCs). Today there are 10 WaSCs whose duties include, the supply of water in areas that are not supplied by the WoCs, and the collection, treatment and disposal of sewerage in all areas. However, there are now only 11 WoCs, after mergers and takeovers. The WaSCs supply drinking water to 80% of the population in England and Wales with WoCs supplying the rest. There are three regulatory bodies in the water and sewerage industry. The Office of Water Services (Ofwat), which is the economic regulator and sets the price limits for each company every five years, the Environment Agency (EA), which is responsible for pollution control, licensing and regulation of water abstraction, and the Drinking Water Inspectorate (DWI), which is responsible for controlling and monitoring drinking water quality.

The method of regulation in the UK water and sewerage sector is price cap regulation and is designed to both give firms incentives to increase profits by reducing costs and eliminating the potential to manipulate output prices. In the UK water and sewerage sector the price cap scheme has the form of RPI ± K, where RPI is the retail price index and K includes two components; X which reflects the beliefs of the regulator about potential improvements in productivity (reduction in costs) that the regulated companies can achieve over a specific period and Q, which reflects the allowed capital expenditure for mandated quality investment projects to improve water and sewerage quality and environmental standards.

At privatization in 1989, price limits were set by the Secretary of State for a period of ten years and were, on average, RPI +5.2 per annum for the industry, RPI+5 per annum for WaSCs and RPI+6.1 per annum for WoCs. The K factor was set at a high level in order to make up for years of underinvestment before privatization and to ensure that the shares of the public companies would be attractive to potential investors. However, as documented in past studies, the first price caps were relatively lax allowing the firms to gain extraordinary profits, and as a result Ofwat exercised its right to reset price caps in 1994. Thus, the average K factor after the 1994 reviews was RPI+0.9 for the industry, RPI+1.0 for the WaSCs and RPI-0.4 for the WoCs, representing a considerable tightening of price caps. This continued in the price review of 1999 with an average K factor of RPI-2.1 for the industry, RPI-2.0 for the WaSCs and RPI-2.8 for the WoCs. In the price review of 2004 the K factor increased again to an average of 4.2% per annum, whereas in 2009 Ofwat published its final price determinations suggesting an average K factor of RPI+0.5 per annum for WaSCs ,and RPI+0.3 per annum for WoCs for the next five years. Thus, Ofwat's most recent price limits would lead to a reduction in the water and sewerage bill of 3 pounds in real terms over the five year period. This was substantially lower than in companies' final proposed plans which suggested an average increase of 31 pounds. The regulated companies would continue to invest about 22 billion pounds over the next five years to maintain assets and services like reduce pressure problems and sewer flooding and improve drinking water and sewerage treatment quality and environmental standards.

The determination of X-factors in the UK water and sewerage industry, and therefore of price limits, is determined through benchmarking techniques, which provide information about the relative performance of companies. As there are companies that are regulated under the same framework, the regulator can compare the performance of each company against the performance of the others in the industry. Ofwat had developed econometric and unit cost analysis to measure the relative efficiency of WaSCs and WoCs after taking into account factors that are outside a company's control and may influence differences between companies' costs.

In its previous price reviews, Ofwat has gradually developed various techniques to measure the relative efficiency of water companies and the potential continuing productivity gains that could be made by water companies in the future. These comparative performance measurements are of great significance for water regulation because they allow Ofwat to set company specific price caps, which are designed to encourage both "catch-up" by relatively inefficient firms and "continuing productivity improvements" (technical change) by relatively efficient firms. Moreover, in principle, if not in practice, any efficiency gains made by the firms during the five years period need to pass to the customers in the first year of the new price control in the form of lower prices. Therefore, in principle, the regulator must achieve a balance between the benefits of consumers and the regulated companies. Moreover, Ofwat assesses water and sewerage companies' overall relative efficiency by aggregating several cross section models, which estimate relative efficiency for different water and sewerage company functions. The use of cross section models has some potential weaknesses since the number of observations (companies) is small (currently 10 WaSCs and 11 WoC). Since WaSCs supply drinking water to 80% of the population in England and Wales and WoCs to the rest, our thesis is concerned with the development of alternative models to assess the performance of the Water and Sewerage Companies (WaSCs).

1.2 Aims and Objectives of the Study

The research aim of the thesis is to evaluate the impact of the regulatory price cap scheme on the financial performance of the Water and Sewerage Companies (WaSCs) when the number of observations is small. In order to achieve this, we investigate the relationship between profits, productivity and price performance across firms and over time (panel data) based on recent methodological developments in efficiency and productivity measurement, Index Number and Data Envelopment Analysis (DEA) techniques. In particular, the first objective is to measure changes in profitability, productivity and price performance across firms (relative comparative performance) and indentify whether price caps are consistent with the achievement of productive and allocative efficiency. The second objective is to measure changes in profitability, productivity and price performance by less productive firms and benchmark firms over time. The third objective is to measure changes in profitability across firms over time caused by other factors except for productivity and price performance changes such as the resource mix, product mix and scale effect. Finally, the fourth objective is to include exogenous factors such as quality in a profit decomposition analysis since the UK water and sewerage sector has carried out substantial capital investment programs to improve water and sewerage quality and environmental standards since privatization. The following section will elaborate the structure of the thesis and provide an overview of the methodologies adapted in attaining the previously mentioned four research objectives.

1.3 Structure of the Thesis

This thesis is divided into eight major chapters. Following this introduction, chapter 2 provides a detail background on privatization and regulation in the water and sewerage sector in England and Wales. It discusses the different types of privatization, the principal aims of a privatization programme and the status of the UK privatized water and sewerage industry. Since WaSCs and WoCs were privatized as natural monopolies and Ofwat was set up as the economic regulator, chapter 2 further provides a discussion of the regulation of a natural monopoly followed by a detailed analysis of the legitimacy of a regulatory system. The method of price cap regulation applied in the UK water and sewerage industry and the concept of the X-factor are also discussed in this chapter.

Chapter 3 provides an overview of comparative efficiency and productivity measurement techniques. It starts with a theoretical presentation of a production technology where the concept of input/output distance functions, technical and allocative efficiency, overall cost and revenue efficiency and scale efficiency are discussed. After providing the necessary conceptual productivity framework the chapter proceeds with the popular techniques that can be used to measure relative efficiency and productivity. A detailed analysis of non-parametric methods is presented, which includes non-frontier methods, index numbers and frontier methods, data envelopment analysis which are applied in this thesis. The discussion on index numbers includes the presentation of the most popular quantity indices that can be used in productivity measurement, followed by a discussion on the approaches that are used to assess the quality of an index number. The usefulness of the index numbers for panel data analysis is also included in this section. Following that, there is a discussion on the concept of measuring efficiency and productivity change by employing data envelopment analysis techniques. The linear programming models used in DEA efficiency analysis, the concept of the Malmquist productivity index which allows the measurement of productivity change over time and the usefulness of DEA for panel data analysis are included in this section. Finally, the next section provides a brief presentation on parametric methods, which includes non-frontier methods, simple regression analysis and frontier methods, stochastic frontier analysis, followed by a discussion on Ofwat's approach to comparative efficiency analysis and productivity measurement.

Chapter 4 presents the reader with a critical review of the existing literature on efficiency and productivity measurement in the UK water and sewerage sector. It presents studies that had employed regression analysis and stochastic frontier analysis, index number techniques and data envelopment analysis methods to measure relative efficiency and productivity change in the water and sewerage sector in England and Wales before and after privatization. It summarizes the key findings from the studies, e.g. the impact of privatization and regulation on the performance of the regulated companies, the inputs, outputs and exogenous factors used, the evidence of economies of scale and scope in the UK water and sewerage industry, followed by a detailed discussion of these studies. Then, the chapter proceeds with a detailed presentation of studies that have carried out economic analyses of the performance of water and sewerage industries in countries other than the UK employing empirical modelling approaches such as regression analysis, data envelopment analysis, stochastic frontier analysis and index numbers.

The methodology and results are presented in chapters 5, 6 and 7. Chapter 5 investigates the first research objective by employing a cross sectional (spatial) index number technique to measure differences in the level of productivity, price performance and profitability across WaSCs (relative comparative performance) over the period 1991-2008. This spatial approach firstly allows comparative performance assessment when the number of observations is extremely limited, which makes the approach directly applicable by regulators in setting price caps. Secondly, and more significantly, it also allows the development of the theoretically consistent model of price cap regulation, e.g. it facilitates an analysis of whether price caps are consistent with the achievement of productive and allocative efficiency. The chapter proceeds with the methodology, the decomposition of a firm's actual economic profitability into two sources: a spatial multilateral Fisher productivity index (TFP) and a newly developed regulatory total price performance (TPP) index. The former is calculated using theoretically consistent relative productivity comparisons across companies in any given year (multilateral spatial comparisons) after assuming the most productive company is the base or benchmark firm. Moreover, we demonstrate that the inverse of a spatial multilateral TFP index can be interpreted as a regulatory excess costs index, which measures the excess of a firm's actual costs relative to benchmark costs. The regulatory TPP index is derived as a function of this regulatory excess cost index and the actual economic profitability index, and measures the excess of regulated revenues relative to benchmark costs. As such, it provides a direct measure of how tight price caps are, measured by the proportional deviation between allowed revenues and benchmark costs. Further consideration of the theoretical relationship between actual economic profitability, regulatory excess costs and regulatory price performance allows a characterisation of the power of regulatory price caps. Moreover, in order to examine the fourth research objective, output and output prices were also adjusted for quality allowing thus the measurement of the implicit impact of quality in the spatial productivity (excess costs) and regulatory TPP measures. The chapter ends with a summary of the key findings and policy implications. This chapter concludes that the period 1991-2000 can be characterised as a period of "weak" regulation since allowed regulatory revenues almost always exceeded regulatory excess costs, thereby demonstrating that price caps during this period allowed firms to maintain economic profitability regardless of whether they made any progress in catching up to benchmark productivity levels. Since 2001 Ofwat has implemented "catch up promoting" price caps since average regulated revenues were always below average regulatory excess costs indicating that the firms were required to eliminate at least some excess costs in order to regain economic profitability.

Chapter 6 investigates the second research objective and presents a panel index number technique to allow for differences in economic profitability, productivity and price performance across WaSCs over the period 1991-2008. It firstly illustrates the measurement of temporal (unit-specific) profitability, productivity and price performance across time for each firm. Secondly, we allow profitability, productivity and price performance comparisons across companies at any given year (multilateral spatial comparisons) calculated by using a multilateral Fisher index. Thirdly, by reconciling together the temporal and spatial profitability, productivity and price performance into relative profitability, productivity and price performance measures, we provide a single index that consistently measures performance change between both firms and over time. Finally, the reconciliation of the spatial, temporal and relative profitability, productivity and price performance measures allows us to decompose the unit-specific index based number profitability growth as a function of the profitability, productivity, price performance growth achieved by benchmark firms, and the catch-up to the benchmark firm achieved by less productive firms. This panel index number technique allows a more comprehensive decomposition of a firm's performance changes and is highly relevant in regulatory and other applications, where comparative performance measurement is appropriate. The inclusion of quality in our analysis allows us to investigate our first research objective and therefore we are able to further decompose the unit-specific profitability growth as a function of the catch-up in quality regarding productivity and price performance achieved by less productive firms and the quality growth of the benchmark firm. The chapter ends with a summary of the main findings. The results indicate that average economic profitability exceeded benchmark economic profitability during the years 1991-1994 and 1999-2008, showing high levels of catch-up relative to benchmark economic profitability after 2001 which was mainly explained by the relative decline in the economic profitability of the benchmark firm. The quality unadjusted productivity results indicated that until 1995 average and benchmark firms did not have strong incentives to achieve high productivity levels. Significant productivity gains for the average firm relative to the benchmark occurred after 1995 which also continued after 2000. Our results suggested that all of this catch-up can be attributed to the post 1995 period, after Ofwat first tightened price caps, and most of it can be attributed to the post 2000 period, following the even more stringent 1999 price review. Moreover, looking at the average and benchmark quality unadjusted price performance we concluded that in the post 1999 price review period, the price performance of all firms was substantially lower than in the first 10 years after privatisation. By 2000, there was a convergence in average and benchmark TPP and during the years 2001-2004, there was little or no difference between average and benchmark TPP and during the years 2005-2008 average TPP exceeded benchmark TPP showing the highest levels of price performance catch-up in 2006 and in 2008.

Turning our discussion now in the quality adjusted results for productivity and price performance changes, we concluded that while quality improvements have contributed to the productivity performance of the WaSCs, they have also contributed negatively to their price performance. The quality adjusted TFP results indicated that after 1997 and until 2002, average quality adjusted TFP increased more rapidly than benchmark quality adjusted TFP, therefore allowing average company to catch-up in quality adjusted productivity to the benchmark quality adjusted productivity, especially during the years 2000-2005. Even after 2002 the average company achieved still significant levels of catch-up in quality adjusted productivity until 2005, which must be attributed to input usage reductions. Furthermore, the considerable increase in average profitability relative to the benchmark firm must be attributed to this catch up effect. Moreover, the quality adjusted TPP results suggested that until 1994, average TPP exceeded benchmark TPP but after 1998, there was a steady erosion of average price performance relative to benchmark price performance suggesting that there was a considerable rebalancing of regulatory price decisions in favour of the benchmark firm, which was even more dramatically extended with the implementation of the 1999 price review in 2001. Moreover, after 2001 average quality adjusted TPP fell more than benchmark quality adjusted TPP suggesting that the broad convergence after 2000 between average and benchmark firm price performance which was observed in the quality unadjusted TPP results was no longer present.

In examining the third and fourth research objectives, profit changes decompose into several factors such as price effect, productivity effect (technical change and efficiency change) and activity effect (resource mix, product mix and scale effect) in Chapter 7 by employing both index number techniques and Data Envelopment Analysis (DEA). This profit decomposition approach shows other determinants except for prices and productivity that can explain the changes in profits and can be useful for industry regulators and managers for performance evaluation and effectiveness of price cap scheme. We also include the impact of exogenous characteristics like quality in a profit decomposition analysis, captured as output for high and low quality and assuming that consumers pay the same price for high and low quality of output. Our sequential DEA technique (using data from all the previous periods) allows us to estimate the productivity and the activity effect and their components when the number of observations is extremely limited. We apply our profit decomposition approaches, without and after controlling for quality to investigate the sources of profit change within the Water and Sewerage Companies (WaSCs) in England and Wales over the period 1991-2008. The chapter ends with a summary of the main findings. The results demonstrate minor differences when we do not control for differences in output quality but in both cases, the policy implications for the UK water and sewerage industry are significant. In both cases, the major determinant on the negative aggregate profitability is explained by the overall negative price effect which outstripped the overall positive quantity effect. The difference between the results from the two types of profit decomposition is on the magnitude of the productivity and activity effect. Without and after controlling for quality, the major determinants on the quantity effect and eventually, on profit change came from the technical change whose magnitude reduced after 2000 and especially during the years 2005-2008, the resource mix effect, a shift to a more cost efficient allocation of resources by substituting labour with capital, and the negative scale effect. The magnitude of the scale effect substantially increased after 2000, indicating that the mergers occurred in 2000/01 did not lower costs and therefore had a negative impact on aggregate profitability. Furthermore, in both types of the profit decomposition, efficiency change was found to have a small positive impact on profit change, whereas after controlling for high and low quality, the product mix effect became positive but small contributing positively to profit changes. Finally, the decomposition of the output effect into high quality and low quality output effect showed that over the whole period the water and sewerage companies moved to the production of more high quality of output than low quality of output contributing positively to the overall output effect and therefore profit changes.

Finally, the concluding chapter (chapter 8) summarises the main findings of the thesis and presents the conclusions. The implications of the study are then offered for both academic and practitioner audience before emphasizing the limitations of the study and suggesting directions for future research. It is concluded that using our cross section index number methodology, regulators and policy makers can determine if past regulatory decisions have not only promoted productive efficiency by providing appropriate efficiency incentives to firms, but also whether they have led to increased allocative efficiency by aligning consumer prices more closely with efficient costs. Moreover, our panel index number methodology facilitated a backward-looking approach that allowed conclusions to be drawn with regard to the impact of the price cap regulation on the productivity, price performance and profitability of the benchmark and less productive firms. Furthermore, using index numbers and DEA techniques, we were able to take into account the impact of quality in a profit decomposition analysis and to decompose profits into several factors that are important for the regulators and regulated companies for performance evaluation and effectiveness of price cap scheme such as price effect, productivity effect, resource mix, product mix and scale effect. Therefore, chapters 5, 6 and 7 allowed us to give a meaningful answer to our research aim, the impact of the price cap regulation on the financial performance of the Water and Sewerage Companies (WaSCs) when the number of observations is extremely limited. Moreover, it is concluded that the simultaneous measurement of firm specific productivity growth, as well as the spatial relative productivity measures in Chapters 5 and 6 would further aid regulators wishing to determine appropriate X-factors for regulated firms, as it would not only provide evidence for potential productivity catch-up, as in the current approaches, but would also provide evidence for further potential productivity improvements by benchmark firms (forward-looking).

Furthermore, chapter 7's findings can be summarised as follows. Firstly, the substantial capital investment programs carried out by the water and sewerage companies since privatization leaded to the production of high quality output and the reduction of low quality output. Secondly, significant productivity improvements which contributed positively to profit changes were mainly attributed to technical change, whereas gains in efficiency were small. This finding is consistent with Cave's review (2009) findings which suggested that since privatization the main driver on productivity growth for the UK water and sewerage sector was attributed to technical change, however, our findings also suggest that technical change was falling over time. Finally, the results from the profit decompositions showed that the resource mix effect was significantly large and positive over the whole period indicating that the water and sewerage industry moved to a cost efficient allocation of resources by substituting labour with capital and therefore contributing positively to profit changes. However, any substantial savings occurred by the resource mix effect were lost due to excessive mergers. The scale effect was negative over the whole period and substantially increased after 2000 indicating that the mergers occurred in 2000/01 had a negative impact on aggregate economic profitability. Therefore, this finding suggests that mergers were not profitable for WaSCs which is in contrast to Cave's review (2009) recommendations which suggested further mergers in the UK water and sewerage industry. We strongly believe that this finding is of great significance as it will allow further analysis on developing methodologies to explore the issue of economies of scale and scope and conclude about the most economically efficient structure and the existence of vertical integration economies in the UK water and sewerage industry (forward-looking).

Finally, our main goal is that the developed models in this thesis will not only serve the need of Ofwat to assess water and sewerage companies but will generally serve the regulated sector, which typically needs to assess the scope for efficiency savings of very large public organizations that have been privatized.

Moreover, the developed models may benefit other multi-function entities with limited data sets such as electricity and gas or health services.

CHAPTER 2 PRIVATIZATION AND REGULATION IN THE WATER AND SEWERAGE INDUSTRY IN ENGLAND AND WALES

2.1 Introduction

This chapter provides a background on the privatization in the UK, the types and aims of privatization and ends with a discussion on the status of the privatized water and sewerage industry in England and Wales. It further provides an analysis of the regulation of a natural monopoly followed by a discussion of the legitimacy of a regulatory system. The method of price cap regulation applied in the UK water and sewerage industry, the concept of the X-factor and therefore the price caps in the previous price reviews are also discussed in this chapter.

2.2 Privatization in the United Kingdom

Privatization was not considered to be one of the major policies in Thatcher's election victory in May 1979. However, the performance of the economy and especially the public industries at the beginning of the two decades 1980s and the 1990s, which were mainly characterized by economic recessions, promoted the phenomenon of privatization. High inefficiency, overmanned and bad management were considered to be the main reasons of the poor performance of the publicly owned industries. Vickers & Yarrow (1988) and Bishop & Kay (1988) provided information about the financial and productivity performance of the public industries for the years 1970-1985 using labour or total factor productivity indices. The bad performance of the publicly owned industries burdened the government budget and the public sector deficit and therefore, privatization appeared to be the appropriate solution.

There are several definitions of the word "privatization". It is the precise reverse of nationalization, the partial or total transfer of an enterprise from public to private ownership (Bos, 1991). Privatization is often justified as an effort to reduce the costs of the government or to improve the performance and effectiveness of the government (Gormley, 1994). Vickers & Yarrow (1991) defined three types of privatization: 1) the privatization of competitive firms, transfer to the private sector of public firms operating in competitive product markets free from substantial market failures, 2) privatization of monopolies that is transfer to the private sector of state-owned enterprises with market power and 3) contracting out of publicly financed services, previously operated by the public sector to private business. The first type of privatization took place in UK before 1984 where British Ports, British Aerospace, Britoil, Cable and Wireless, National Freight and British Petroleum, were the first to be privatized. The second type of privatization occurred in the UK after 1984 and referred to network utilities in telecommunications, electricity and gas and water and sewerage industry. The key difference between the two types of privatization was the presence of monopoly power and thus the need for some form of regulation. The third type of privatization, contracting out of publicly financed services to the private sector, refers to a franchise agreement. Several services such as the street cleaning, health provision or education, although staying publicly organized and financed, were conducted by private organizations.

Yarrow (1986) provides a list of the principal aims of a privatization programme. The two main objectives of privatization are concerned with the improvement in efficiency by increasing competition and allowing companies to borrow from the capital market and the reduction in the public sector borrowing requirement (PSBR). The privatized firms have the incentive to improve efficiency by reducing their costs and the goods and services are provided to the consumers at a low economic cost. Moreover, the privatized firms have the opportunity to borrow freely from the capital markets in order to finance their activities without having to comply with the borrowing constraints faced in the public sector. The revenue from the sale of the assets of the public enterprises is a significant factor to cover the heavy public sector deficit. Ricketts (2004) reports figures that have to do with the state revenues from privatization in other countries like China (US$ 100bn in 2001). Furthermore, Parker (1999, 2004) refers to the effectiveness of the privatization in many ways. The service quality does not seem to have been undermined due to large cost savings and the shares gave a high return to those who held them due to the high profits that those companies had right just after privatization. Nevertheless, privatization is still under investigation and a lot is to be examined before extracting the right conclusion.

In the second phase of privatization in the UK, more public network utilities were privatized such as British Telecommunications (BT) in 1984, British Gas (BG) in 1986, the Water and Sewerage Industry in England and Wales in 1989, the Electricity Supply Industry (ESI) in 1990. Since monopoly power was present in these industries, regulatory bodies were set up in order to protect the consumers from monopoly abuse and facilitate competition. The first regulator body was the Office of Telecommunications (Oftel, now part of Ofcom), then Ofgas (for British Gas), Ofwat (for the water and sewerage industry) and Offer (for the electricity industry). Later Ofgas merged with Offer and formed the new regulator Ofgem for the electricity and gas industry.

The water and sewerage industry was the third utility to be privatized in 1989 in the UK. Before privatization there were 10 Regional Water Authorities responsible for the water and sewerage supply in England and Wales and 29 Statutory Water companies, which were already privatized companies and were responsible for the supply of water. After 1989, the 10 Regional Water Authorities were privatized and formed the Water and Sewerage Companies (WaSCs) and the 29 Statutory Water Companies became the Water Only Companies (WoCs). Today there are 10 WaSCs whose duties are, except for the supply of water in areas that are not supplied by the WoCs, the collection, the treatment and disposal of sewerage. Also, there are 11 WoCs, after mergers and takeovers, which ensure the extraction and treatment of water for the domestic and industrial supplies. The WaSCs supply drinking water to 80% of the population in England and Wales and WoCs to the rest. There are three regulatory bodies in the water and sewerage industry. The Office of Water Services (Ofwat), which is the economic regulator and sets the price limits for each company every five years, the Environment Agency (EA), which is responsible for pollution control and licensing and regulation of water abstractions and the Drinking Water Inspectorate (DWI), which is responsible for controlling the drinking water quality conditions.

Comparative competition is a key characteristic of regulation in the UK water and sewerage industry. As there are companies that are regulated under the same framework, the regulator can compare the performance of each company against the performance of the others in the industry. Ofwat had developed econometric and unit cost analysis to measure the relative efficiency of WaSCs and WoCs after taking into account factors that are outside a company's control and may affect differences between companies on costs. Since privatization there have been a number of take-overs and mergers in the industry. Proposed mergers between water companies where the companies each hold assets used in the water enterprise valued in excess of 30 million pounds were referred to the Monopolies and Merger Commission (MMC). If the proposed merger involved a company with assets valued at under 30 million pounds, the proposal should be considered under the normal provisions of the Fair Trading Act (CRI, 1998). The ability of Ofwat to make comparisons between different companies is also relevant before a merger occurs.

Since water and sewerage services companies have a monopoly for most of the services they provide opportunities for direct competition are limited due to the nature of the water industry (CRI, 1998). However, there is some scope for competition to supply water and/or sewerage services within a license area. This is known as inset appointment, an agreement whereby one company would be given the right to supply services within another's statutory area (Cowan, 1997). Under the Water Act, 1989 there are three circumstances under which a new appointment can be made: (1) if the existing Appointee consents, (2) where premises in the area in question are not currently supplied by the existing Appointee and (3) where the circumstances for varying the existing appointment are provided for in the terms and conditions of the appointment (CRI, 1996). Common carriage is another form for competition in the UK water industry except for inset appointments and occurs when one service provider shares the use of another's assets.

2.3 Regulation in the United Kingdom

Parker (2004) provides a summary about the objectives of the regulator. Economic regulation tries to achieve a balance between promoting the consumers' interests in terms of lower prices and good quality of service and guaranteeing adequate returns to the investors by earning a normal return on capital invested. The primary duty of the regulator is to ensure that the regulated companies carry out and can finance their functions. It protects customers from monopoly abuse, promotes economy and efficiency and facilitates competition. Two popular methods of regulation are *rate of return* and *price cap* regulation. Berg and Sanford (1997) provide a list of other types of regulation for network utilities such as the performance based, franchise and yardstick regulation. The UK privatized utilities follow the price cap regulation, which will be discussed in detail later in this section after presenting the regulation of natural monopolies.

The implications for pricing of services provided by a natural monopoly can be tackled by asking two questions: what price would emerge in the absence of government regulation, and what prices should regulators try to attain? (Baldwin and Cave, 1999). The two questions can be answered in relation to Figure 2.1, which depicts the demand curve D, the average total cost curve AC and the marginal cost curve MC, which is assumed to be constant. Since the natural monopolies have large investment in pipelines such as a gas company or water and sewerage company, fixed costs are high and thus the average total cost curve has a downward slope, meaning that average total cost decreases as output increases. If the natural monopoly is not regulated, then it produces at Q_m where marginal revenues equal marginal cost and sets a price at the level of P_m resulting in excessive profits. The monopoly regulator intervenes by setting a price ceiling, a rule that specifies the highest price that the firm is permitted to set (price cap regulation).

Ideally the prices of goods and services sold in the economy should be set at their marginal costs because at a price where the demand curve cuts the marginal cost curve (P_{mc} in Figure 2.1), output has been expanded up to the point where the consumers' willingness to pay for an additional unit of the good and service provided, shown by the demand curve, exactly equals the marginal cost to the economy of producing that final unit of output. The output provided at a price P_{mc} is shown by Q_{mc} in Figure 2.1. At a price greater than this, the consumers' willingness to pay would exceed the marginal cost of providing an extra unit. At a price lower than this, the marginal cost of providing the last unit of output is higher than the consumers' willingness to pay for it. Hence, in theory, an economically efficient allocation requires a price equal to marginal cost. However, a price equal to P_{mc} would fail to cover firm's average cost (AC) and therefore the firm would make a loss and not be able to survive in the long-run. The lowest price consistent with the firm breaking even is P_{ac}, where the demand curve cuts the average cost curve. The output provided at a price P_{ac} is shown by Q_{ac} in Figure 2.1. Thus, if the regulator must ensure that the firm breaks even, this is the best price available. This is more satisfactory than the natural monopoly price P_m, however, less efficient than a price equal to marginal cost, P_{mc}.

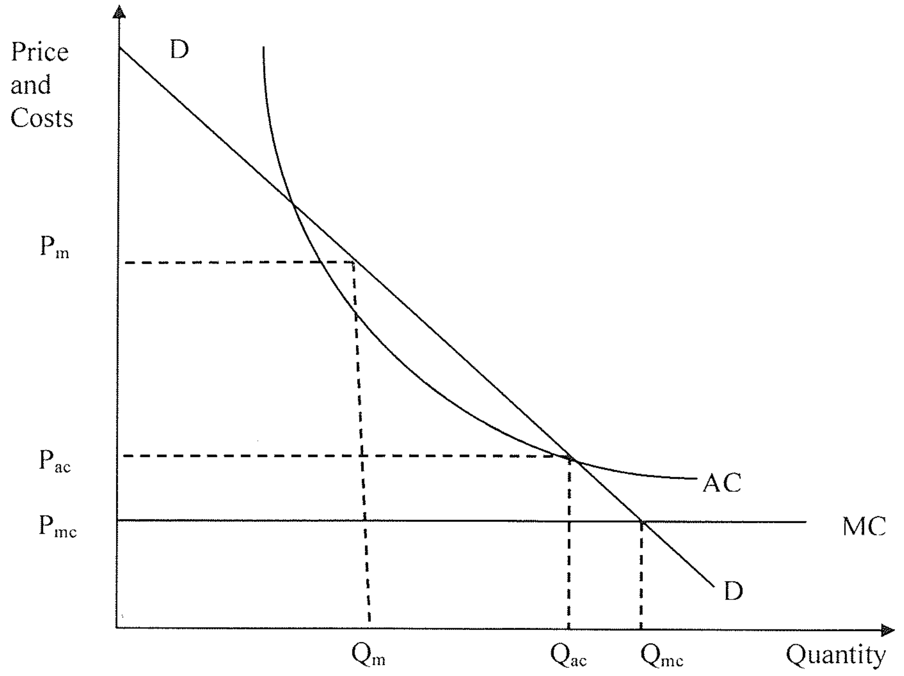


Figure 2-1 Regulation of a Natural Monopoly

The establishment of a regulatory body is necessary to serve the consumers' and producers' interests. The legitimacy of a regulatory body depends on public trust and is related to accountability, transparency, proportionality, targeting and consistency. Figure 2.2 depicts the legitimacy of a regulatory system. Accountability implies that regulators work within clearly agreed rules and are democratically accountable for their actions, whereas proportionality suggests that the regulation should be proportional to the market failure to be tackled. A transparent regulatory regime allows the public to appreciate the grounds for regulatory decisions and facilitates public consultation and challenge, whereas a targeted regulatory system is one in which the regulations introduced to correct market failure are not so loosely drafted that they impact unintentionally on other parts of the economy. Consistency suggests that the regulator should achieve a balance between the benefits of the consumers and the shareholders and develop a trust between the regulator and the regulated companies.

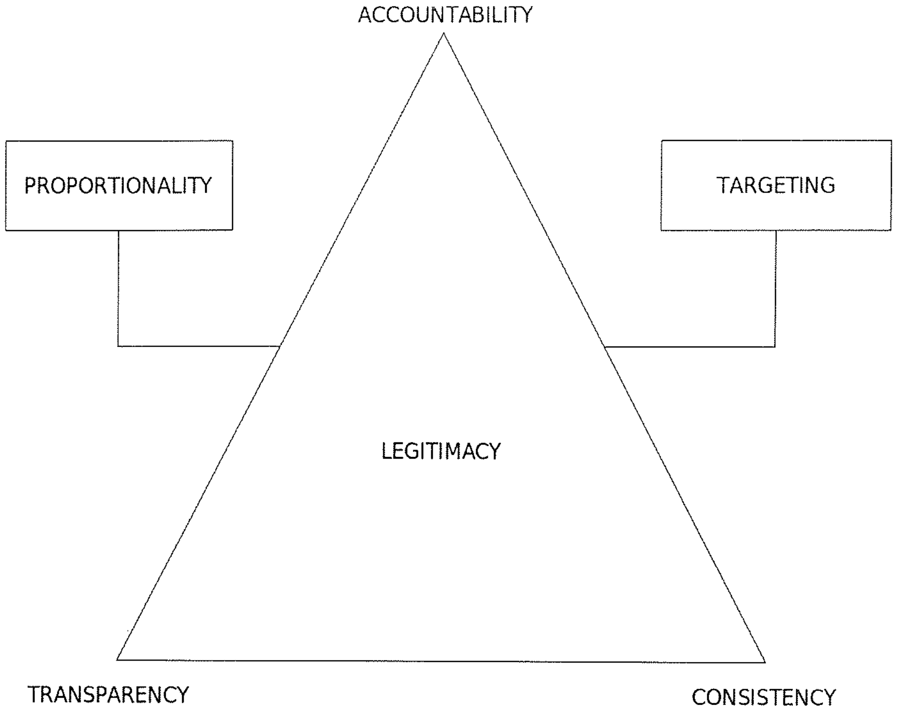


Figure 2-2 Regulatory Legitimacy (Parker, 2003)

Two popular methods of price regulation that are applied in the US and UK network utilities are rate of return and price cap regulation. Under rate of return regulation, the regulated firm is permitted to charge prices that would cover its operating costs and give it a fair rate of return on the fair value of its capital. When the prices move out of line with the company's costs, it could ask for a new set of prices (Green and Pardina, 1999). By imposing this type of regulation, the company has the opportunity to cover both the operating and capital cost and gain a return on capital. The formula used for the calculation of the firm's required revenue per year for the targeted rate of return based on its cost, which is the basis for charging prices to customers, is the following:

$$RR = OE + D + T + (RB \times ROR) \tag{2.1}$$

Where RR is the required revenue, OE is the operating cost, D are the depreciation expenses, T are the tax expenses, RB is the rate base (the value of the plant and equipment) and ROR is the rate of return. The regulated company files a tariff (required revenue) when it wishes to revise its price after calculating the operating expenses, the capital employed and the cost of capital. The advantage of the rate of return is the assurance the investors get that they will earn a fair rate of return. The regulator evaluates an allowable rate of return on the assets of the regulated firm to guarantee that it makes an appropriate level of capital investment. In imposing the rate of return the cost of capital to the firm is measured. Once the regulator has agreed an investment decision the regulated firm is assured a full cost recovery. However, the strong disadvantage of the rate of return regulation is that it does not provide any incentives to the firm to reduce its costs and improve efficiency. Indeed, the firm has the incentive to over-invest in capital (Averch-Johnson, 1962).

In contrast, under price cap regulation, the regulated companies are allowed to increase the price of the services by no more than the annual rate of increase in the Retail Price Index (RPI) less a negotiated factor X to represent the annual productivity improvements in excess of the economy-wide average that might be reasonably anticipated (Ricketts, 2004). During the regulatory period, usually four to five years, the regulated company cannot increase the annual average price that it charges to the consumers for its goods and services faster than RPI-X, where RPI is the Retail Price Index and X is an efficiency factor which measures productivity improvements. At the end of the regulatory period, the regulator sets the X factor again based on the cost savings, the efficiency improvements that the firm achieved during the pre-specified period. The reduction in costs passes to the customers with the form of lower prices and the process is repeated again. Efficiency gains under price cap regulation are fairly shared between investors, in terms of higher profits and consumers, in terms of lower prices.

The price cap regulation, the system RPI-X, was proposed by Littlechild (1983) in his report to the Government for the regulation of British Telecommunications (BT). He suggested that the new method of regulation should protect the consumers against the monopoly power, promote efficiency, innovation and competition and reduce the burden of the regulation. Littlechild rejected the rate-of-return regulation for applying it for BT because it did not provide any incentives for cost reductions and did not encourage efficiency. The RPI-X corresponds to the criteria that Littlechild set in his report because it is based on capping prices rather than profits so it encourages the firm to minimize its costs. The regulated firm has the incentive to combine its resources efficiently in order to produce a given level of output and to maintain any cost reductions. Regulated firms can increase their profits by reducing costs and not by exploiting prices. This can be explained by the following example.

The regulator sets a price limit of 100 million pounds the first year for the firms and anticipates that in the end of the regulatory period (usually five years) they can deliver their outputs to the consumers for costs of 90m pounds plus a return on capital to investors of 10m pounds. Since the increase in the return of capital cannot be gained by raising prices, the firms reduce their costs. If we assume, that each year the firm manages to achieve additional savings of 2m pounds than the regulator anticipates, at the end of the regulatory period, the rate of return will be 20m pounds thanks to the additional reduction in costs. So the higher returns on capital to the investors include the allowable rate of return (anticipated gains), which the regulator sets in the first year (10m) plus the additional savings of 10m (2m of each year) in the end of the regulatory period (unanticipated gains) by delivering outputs with a reduction in costs of 80m. The firms operate in lower cost and so the regulator can reduce the prices even further by presuming that the firm can reduce its costs by a further 10m by the end of the next five years period. That is the regulator in the next review period will set a target cost of 80m as demonstrated.

The regulator has to pass the efficiency gains to the customers in the form of lower prices. This can be done in two ways, "Po adjustment" and "glidepath". In the previous example the investors gain an excess rate of return of 20% above the cost of capital (10%). The regulator by using the Po adjustment technique requires an immediate cut in prices in the first year of each regulatory period, such that the firm's rate of return is reduced to the cost of capital at once. In this case the efficiency gains made by the firms during the five years period are passed to the customers in the first year of the new price control in the form of lower prices. A different technique to the "Po adjustment" is by making phased price reductions through a *glidepath* mechanism such that the rate of return is reduced to the cost of capital by the end of the price control period rather than in the first year of the new price control period (National Audit Office, 2002).

The determination of the X factor is an important issue for the regulator. Armstrong, Cowan and Vickers (1999) emphasize six factors that enter the process when the regulator has to decide about setting the X factor; the cost of capital, the value of the existing assets, the future investment programs, the expected future changes in the productivity and demand growth, and finally the effect of the X-factor on potential competitors. Bernestein & Sappington (1999, 2000) provide an example on how the X-factor can be determined in a regulated industry. The authors suggested that the X factor reflects the extent to which the total factor productivity in the regulated industry exceeds the total factor productivity elsewhere in the economy and the increase in input prices in the regulated industry do not exceed the increase in the input prices elsewhere in the economy. This means that the regulator estimates the costs of the firms in the future and sets the price cap so that the firms will cover those predicted costs. If they manage to increase their efficiency and reduce costs more than the regulator anticipates, the profits will also increase. Further, the regulator will pass the benefits to the consumers in the form of lower prices in the next regulatory period as the lower cost base will become the reference for the regulator in the next periodic review. If firms' efficiency is less than the regulator anticipated, then the profits will go down. However, if the regulator realizes that the firms have gained excess profits, then he/she will adjust the prices in the next price review by a once-and-for all adjustment to bring prices back into line with the costs (Po adjustment) or by a gradual process of erasure of excess profits over the next period.

In very broad terms, the X factor consists of two components, the catch-up and continuing improvement factors. The catch-up (efficiency change) urges the firms to improve their performance towards the top performing firms in the industry. The continuing improvement (technical change or frontier shift) factor indicates that the top performing companies should continue to improve their performance. The catch-up factor and technical change are depicted at Figure 2.3. A series of isocost lines, IC_1, IC_2, IC_3 are drawn representing different levels of cost using the combination of two factors, capital (K) and labour (L), whereas the isoquant, $I(Y)$ shows the different combinations of capital and labour that can produce a given level of output, Y. The least combination of capital/labour to produce a given level of output is at the point where the isoquant is tangent to the isocost line. Let's assume that the firm is efficient and operates at the point A by using K1 units of capital and L1 units of labour for a given level of output. In this case, the efficient factor X will reflect only technical change, that is the firm will not be required to 'catch up' with the efficient boundary, but will be required to have costs which decline as the efficient boundary moves over time to a more productive position. In the next period the regulator will expect that the firm will be able to move from point A to C and reduce its costs.

However, if the firm is not efficient and operates in the point B instead of C, the X factor will reflect both efficiency change and technical change, by moving from point B to point A, while the price cap will be imposed under the assumption that the costs can reduce even further from B to point C.

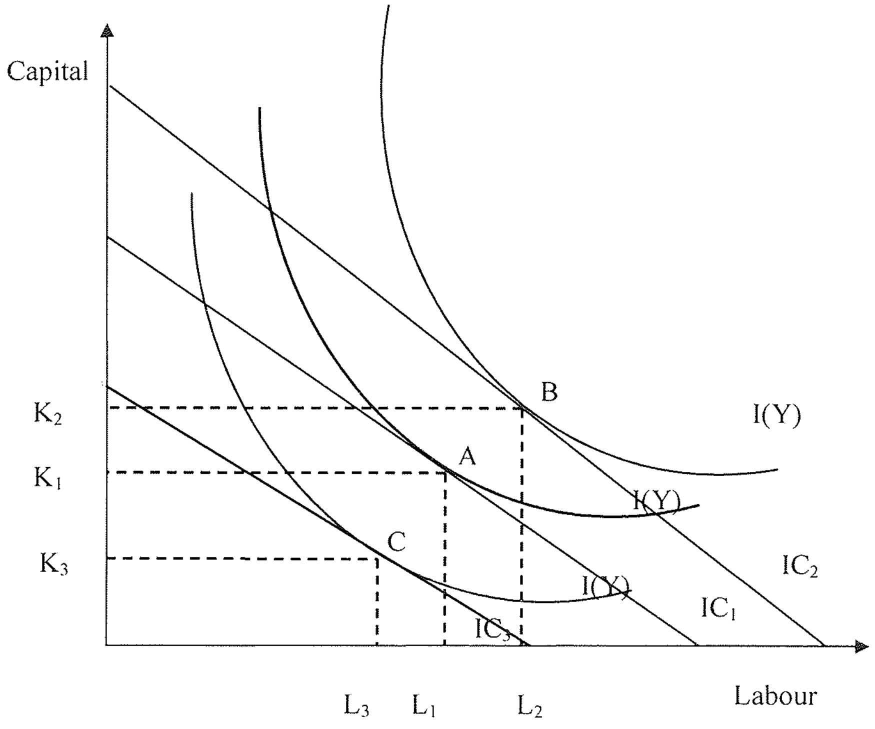


Figure 2-3 Catch-Up and Efficiency Change

In the UK water and sewerage sector the price cap scheme had the form of RPI+/-K, where the K factor reflects the improvement in efficiency and also the capital investment projects for improving the environmental standards and drinking water quality. Each company needs to collect sufficient revenue in order to finance its operating costs and the capital investment project. In addition, an allowance is made for tax and a return on capital. The percentage change between the revenue requirement and the revenue from the customers is the price limit (Emery, 2003). Price limits are determined separately for operating expenditure (OPEX) and capital expenditure (CAPEX) for water supply and sewerage services providing incentives for management to economise on OPEX and CAPEX (Parker et al, 2006).

At privatization in 1989, the price limits were set by the Secretary of State for a period of ten years and had the form of RPI +5.2. The K factor was set at a high level in order to make up for years before privatization of underinvestment and to ensure that the shares of the public companies would be attractive to potential investors. However, the first price cap was too lax allowing the firms to gain extraordinary profits and so the price cap needed to be reset. Parker (1997) reports that after privatization water and sewerage industry's profitability in terms of profit margins had been in excess of 30% although the return on capital employed had been much more modest 4.5%-15%, reflecting the large capital stock in water and sewerage services. The increasingly unacceptably high profits gained by the regulated firms made the regulator intervene and reset the price cap in 1994. Thus, the new K factor was based on the re-examination of the excessive returns to the investors and the further need to meet environmental obligations and drinking water quality standards. The weighted average K factor was RPI+0.9 for the industry, RPI+1.0 for the WaSCs and RPI-0.4 for the WoCs. The need for water and sewerage companies to meet the environmental obligations resulted in an increase in capital expenditure in the first five years and so the average K factor was higher in the first five years up to 2000. The price limits set in 1994 differed from the previous ones in 1989 because any further increase in the customer's bills would be unaffordable. Companies would need to provide existing standards of service at lower prices and to fund improvements in standards of service through greater efficiency rather than higher prices (Ofwat, 1994).

The customers expressed their satisfaction about the improvement in overall service provided by the water and sewerage companies and water only companies over the recent years. However, their main concern was the increase in the bills. The regulator set a new price limit in 1999 for the water and sewerage industry where there would be a reduction in the customers' bills compared to the previous years. On average, the K factor for the industry over the period 2000-2005 was RPI-2.1 and the price formula for the WaSCs had the form of RPI-2.0, while for the WoCs took the form of RPI-2.8. There would be a reduction in the customers' bills in the first year, on average, 13% and by 2004/05 bills would be on average around 12% lower in real terms than 1999/00 (Ofwat, 1999).

In 2004 Ofwat announced its final determinations with respect to the price limits that the companies would face for the period 2005-10. The annual average price limit for the water and sewerage industry was RPI+4.2%. The reasons for the increase in the prices charged to customers were associated with an increase in capital investment program to meet additional quality standards and reduce the risk of sewer flooding and new information in energy and pension costs (Ofwat, 2004). Finally, in 2009 Ofwat published its final price determinations suggesting that WaSCs should increase their prices charged to customers by RPI+0.5% and WoCs by RPI+0.3% on average in the next five years (Ofwat, 2009). Ofwat's price limits would lead to a reduction in the water and sewerage bill of 3 pounds in real terms over the five year period. This is substantially lower than in companies' final proposed plans which suggested an average increase of 31 pounds. The regulated companies would continue to invest about 22 billion pounds over the next five years to maintain assets and services like reduce pressure problems and sewer flooding and improve drinking water and sewerage treatment quality and environmental standards.

In RPI-X price regulation, the regulated companies are allowed to increase the price of the services by no more than the annual rate of increase in the Retail Price Index (RPI) minus an X-factor, which measures productivity improvements ie, it reflects the degree to which the regulator believes that the regulated companies can increase their productivity or decline their costs over a specific period (usually five years). The setting of X-factors includes the measurement of industry-level annual productivity growth using historic data and firm-level relative efficiency using average-based and frontier-based methods (benchmark techniques). The next chapter will provide an overview of comparative efficiency and productivity measurement techniques such as index numbers, data envelopment analysis, simple regression analysis and stochastic frontier analysis. We will also discuss Ofwat's approach to comparative efficiency analysis and productivity measurement. Furthermore, in chapter 6 will discuss a panel index number technique which can provide evidence with regard to the potential productivity catch-up of less productive firms and the potential for further improvements in benchmark productivity levels, a technique which can further aid regulators to determine appropriate X-factors for regulated firms. Finally, chapter 7 will combine both index numbers and DEA techniques to allow the estimation of productivity effect, technical change and efficiency change, productivity improvements achieved by most and less productive firms, scale effect, resource mix and product mix effect and their impact on profit changes.

2.4 Conclusions

This chapter provided a detail background on privatization and regulation in the water and sewerage sector in England and Wales. It discussed the different types of privatization, the principal aims of a privatization programme and the status of the privatized water and sewerage industry in England and Wales. Since WaSCs and WoCs were privatized as natural monopolies and Ofwat was set up as the economic regulator, this chapter further provided a discussion of the regulation of a natural monopoly followed by a detailed analysis of the legitimacy of a regulatory system. The method of price cap regulation, the concept of the X-factor and therefore the price caps in the previous price reviews applied in the UK water and sewerage industry were also discussed in this chapter.

The next chapter will provide a detailed analysis of non-parametric methods, including non-frontier methods, index numbers and frontier methods, data envelopment analysis which are applied in this thesis, followed by a brief discussion on parametric methods, simple regression analysis and frontier methods, frontier analysis and on Ofwat's approach to comparative efficiency analysis and productivity measurement. Index numbers and data envelopment analysis techniques will further allow us to provide a robust and comprehensive performance analysis of the Water and Sewerage companies (WaSCs) when the number of observations is extremely limited, thereby serving the needs of Ofwat in assessing the performance of water and sewerage companies. Furthermore, index numbers and data envelopment analysis methods will allow us to evaluate the impact of the regulatory price cap scheme on the financial performance of the WaSCs (backward-looking), which is the research aim of this thesis, by investigating the relationship between profits, productivity and price performance across firms and over time. The subsequent chapter will discuss alternative profit decompositions that are of great importance for the regulator and the regulated companies based on recent methodological developments in efficiency and productivity measurement, index number and DEA techniques.

Finally, chapter 6 discusses a panel index number technique that gives evidence not only with regard to the potential productivity catch up of laggard firms, but also the potential for further improvements in benchmark productivity levels. Such an approach would further aid regulators wishing to determine appropriate X-factors for regulated firms, as it would not only provide evidence for potential productivity catch-up, as in the current approaches, but would also provide evidence for further potential productivity improvements by benchmark firms (forward-looking). Finally, chapter 7 will combine both index numbers and DEA techniques to allow the decomposition of profit change into several determinants that are of great significance for the regulator and the regulated companies such as the quantity and price effect, the productivity effect which decomposes into technical change and efficiency change and the activity effect which decomposes into the scale effect, resource mix and product mix effect.

CHAPTER 3 OVERVIEW OF COMPARATIVE EFFICIENCY MEASUREMENT TECHNIQUES

3.1 Introduction

The production economics textbooks generally assume that producers are efficient. However, it is common knowledge that there are producers that are not fully efficient (however defined) and some producers are more efficient than others. Where there is regulation, the regulator's duty is to encourage the less efficient firms to improve their performance towards the most efficient firms, while the most efficient firms need to continue to improve their performance. This is achieved through comparative efficiency analysis, which is then used to determine price controls for each regulated firm in a price cap scheme such as the UK water and sewerage sector. This chapter gives a brief discussion of production concepts and then analyzes performance measurement techniques.

3.2 Theoretic Representation of a Production Technology

A production technology describes what is technically feasible in terms of converting inputs (X) into outputs (Y). A production set gives the set of all feasible production points and the upper boundary of this set is called production surface (or frontier). Points below the production frontier are technically inefficient, whereas points on the production frontier are technically efficient. Thus, the production technology set, S, can be described as the set of all inputs and outputs vectors (X,Y) such that X can produce Y.

$$S = \{(X,Y): X \text{ can produce} Y \} \tag{3.1}$$

The production technology set S can be equivalently defined by the output set O(X), which represents the set of all output vectors that can be produced by a given input vector, X. Alternatively, the production set S can be defined by the input set L(Y), which represents the set of all input vectors used to produce a given output vector, Y.

$$O(X) = \{Y : X \text{ can produce} Y\} = \{Y : (X,Y) \in S\} \quad \text{and}$$
$$L(Y) = \{X : X \text{ can produce} Y\} = \{X : (X,Y) \in S\} \tag{3.2}$$

The lower boundary of an input set is its input isoquant, which gives the several combinations of inputs that can be used to produce a given level of output, whereas the outer boundary of an output set is its output isoquant, or production possibility curve, which gives the various combinations of outputs that can be produced by a given level of input.

Distance functions are of great significance in describing the technology set and thus in measuring efficiency and productivity. Distance functions can be used in the case of multi-output and multi-input technologies without the need to include prices for outputs and inputs. An input distance function is defined as the minimal proportional contraction of the input vector given an output vector, whereas an output distance function is defined as the maximal proportional expansion of the output vector given an input vector. The input and output distance functions were introduced by Shephard (1970) and were given by:

$$D_I(Y,X) = \max\{\gamma : (X/\gamma) \in L(Y)\} \text{ and } D_O(Y,X) = \min\{\delta : (Y/\delta) \in O(X)\} \qquad (3.3)$$

If X belongs to the input set, then $D_I(Y,X) \geq 1$ and if X belongs to its input isoquant, then $D_I(Y,X) = 1$. If Y belongs to the output set, then $D_O(Y,X) \leq 1$ and if Y belongs to its output isoquant, then $D_O(Y,X) = 1$. Under *constant returns to scale*, the input distance function is the reciprocal of the output distance function for any (X,Y):

$$D_I(Y,X) = 1/D_O(Y,X), \text{ for all X and Y} \qquad (3.4)$$

Farrell (1957) suggested that the total economic efficiency of a firm includes two components, technical efficiency and allocative efficiency. Figure 3.1 gives a representation of input and output oriented measures of technical efficiency using ratios and distance functions. Technical efficiency is defined as the ability of a firm to obtain maximal output from a given set of inputs (output-oriented) or reduce inputs to minimal levels for a given set of outputs (input oriented). In Figure 3.1, the output oriented measure of technical efficiency of a firm operating at A can be expressed in terms of ratio and output distance function as:

$$TE_O = DA / DB = D_O(Y, X) \tag{3.5}$$

The input oriented measure of technical efficiency of a firm can be expressed in terms of ratio and input distance function as:

$$TE_I = EC / EA = 1 / D_I(Y, X) \tag{3.6}$$

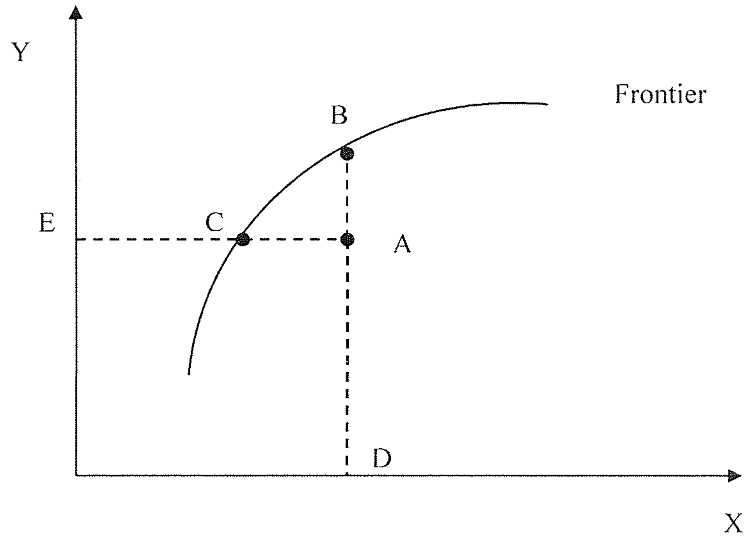


Figure 3-1 Input and Output Oriented Measure of Technical Efficiency

In the presence of input/output price information, it would be possible to measure the allocative efficiency of a firm, the ability of a firm to use the inputs/outputs in optimal proportions given their input/output prices and the production technology. Therefore, the measures of technical and allocative efficiency are then combined to provide a measure of overall cost or revenue efficiency. Figures 3.2 and 3.3 display the measurement of overall cost and revenue efficiency.

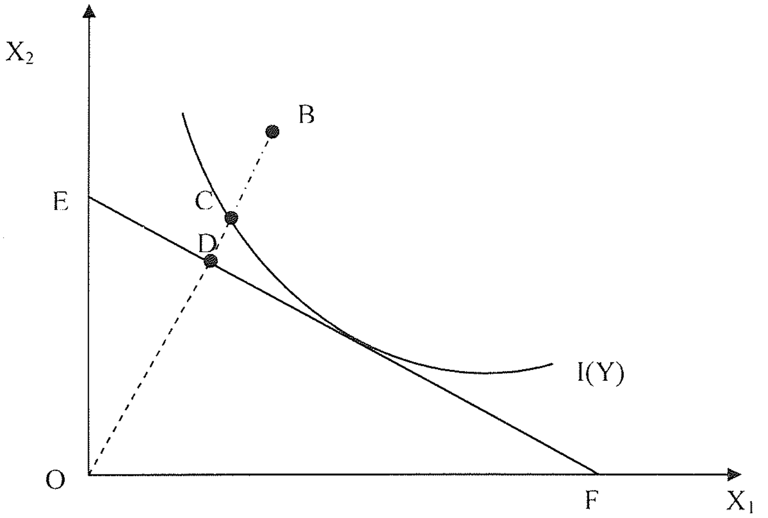


Figure 3-2 Overall Cost Efficiency

In Figure 3.2 I(Y) depicts the efficient input boundary, the minimum input requirements to produce a given level of output. If a given firm uses quantities of inputs, defined by the point B, to produce a unit of output, then the technical inefficiency of that firm could be represented by the distance BC, which is the amount by which all inputs could be proportionally reduced without a reduction in output. This is usually expressed in percentage terms by the ratio OC/OB, which represents the percentage by which all inputs need to be reduced to achieve technical efficient production (Coelli et al, 2005). The technical efficiency (TE) of the firm is measured by the ratio:

$$TE = OC/OB \tag{3.7}$$

If the input price ratio, given by the isocost line EF is also known, then allocative efficiency can be measured by the ratio:

$$AE = OD/OC \tag{3.8}$$

A firm that operates at point C is technically efficient but allocatively inefficient, whereas a firm that operates at point G is both technically and allocatively efficient. Therefore, the distance DC shows the decline in production costs that would occur if production were to occur at the point G instead of point C. Finally, overall cost efficiency is measured as the product of technical and allocative efficiency measures:

$$CE = TE \times AE = (OC / OB) \times (OD / OC) = (OD / OB) \qquad (3.9)$$

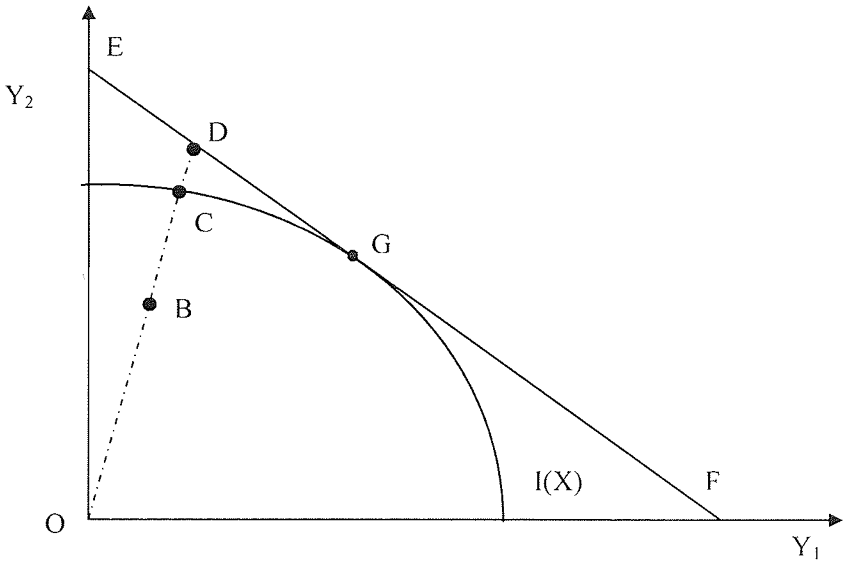


Figure 3-3 Overall Revenue Efficiency

In Figure 3.3, I(X) represents the efficient output boundary, the maximal outputs that can be produced with a given level of input. If a given firm produces quantities of outputs, defined by the point B, using a single input, then the technical inefficiency of that firm can be represented by the distance BC, which represents the amount by which outputs could be increased without requiring extra input (Coelli et al, 2005). The technical efficiency (TE) of a firm is measured by the ratio:

$$TE = OB / OC \qquad (3.10)$$

If the output price ratio, given by the isorevenue line EF is also known, then allocative efficiency can be measured by the ratio:

$$AE = OC / OD \qquad (3.11)$$

Finally, overall revenue efficiency is measured as the product of technical and allocative efficiency measures:

$$RE = TE \times AE = (OB/OC) \times (OC/OD) = (OB/OD) \qquad (3.12)$$

However, it is possible that a firm is both technically and allocatively efficient but the scale of operation of the firm may not be optimal. Figure 3.4 illustrates the concept of the effect of scale on efficiency, where a-one input, one output CRS and VRS technology are depicted. A production technology exhibits constant returns to scale (CRS) if a A% increase/decrease in inputs results in A% increase/decrease in outputs and increasing returns to scale (IRS) if a A% increase/decrease in inputs results in a more than A% increase/decrease in outputs. Finally, a production technology exhibits decreasing returns to scale (DRS) if a A% increase/decrease in inputs results in a less than A% increase/decrease in outputs. In Figure 3.4, firms B and C are technical efficient because they are operating on the production frontier but firm A is technical inefficient. Firm B is technical efficient under VRS technology and firm C is technical efficient under CRS technology. The technical efficiency of firm A relates to the distance from observed data point to the VRS technology and is equal to the ratio $TE_{VRS} = DB/DA$, whereas its technical efficiency under CRS is equal to the ratio $TE_{CRS} = DC/DA$. The *scale efficiency* of firm A relates to the distance from the technically efficient data point B to the CRS technology and is equal to the ratio $SE = TE_{CRS}/TE_{VRS} = DC/DB$. The efficiency of firm A can improve by moving from point A to point B on the VRS frontier (removing technical inefficiency) and it can further improve by moving from the point B to the point F (removing scale inefficiency). The firm operating at point F is unable to become more efficient by changing its scale of operation and it is said to be operating at the *most productive scale size* (MPSS) or at the *technically optimal productive scale* (TOPS) (Coelli et al, 2005).

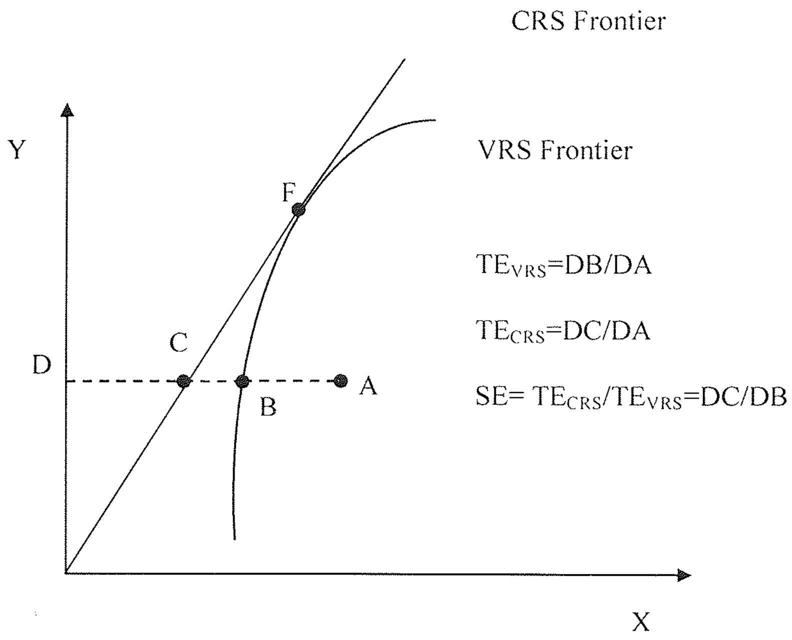


Figure 3-4 Scale Efficiency

After providing the necessary conceptual productivity framework we can proceed with the popular techniques that can be used to measure relative efficiency and productivity. These techniques can be separated into two categories: parametric and non-parametric methods. Parametric methods include non-frontier methods, simple regression analysis and frontier methods, stochastic frontier analysis, whereas the non-parametric methods include non-frontier methods, index numbers and frontier methods, data envelopment analysis.

3.3 Index Numbers and Productivity Measurement

Productivity measures the ability of inputs to produce outputs. If we have a single output and single input, then productivity is a ratio of output to input. If we have multiple inputs and outputs, then we have to derive an aggregate output and input index in order to measure productivity. Growth in productivity or total factor productivity (TFP) growth is the change in "the ratio" of outputs and inputs between two time periods.

Index numbers are used when we have information about prices for inputs and outputs. We can construct both price and quantity indexes. In our case we focus on quantity indices, which can be used to measure productivity. The Laspeyres, Paasche, Tornqvist and Fisher quantity indices are the most widely used index formulas given by:

$$Laspeyres \ = Q_{st}^{L} = \frac{\sum_{k=1}^{K} p_{ks} q_{kt}}{\sum_{k=1}^{K} p_{ks} q_{ks}} \qquad Paasche \ = Q_{st}^{P} = \frac{\sum_{k=1}^{K} p_{kt} q_{kt}}{\sum_{k=1}^{K} p_{kt} q_{ks}}$$

$$Fisher \ = Q_{st}^{F} = \sqrt{Q_{st}^{L} \times Q_{st}^{P}}$$

$$Tornqvist \ = Q_{st}^{T} = \prod_{k=1}^{K} \left[\frac{q_{kt}}{q_{ks}} \right]^{\frac{\omega_{ks} + \omega_{kt}}{2}}$$

$$\ln Q_{st}^{T} = \sum_{k=1}^{I} \left(\frac{\omega_{ks} + \omega_{kt}}{2} \right) \left(\ln q_{kt} - \ln q_{ks} \right)$$

p_{kt}, p_{ks} : Price of k - th good in t - th and s - th period

q_{kt}, q_{ks} : Quantity of k - th good in t - th and s - th period

$$\omega_{ks} = p_{ks} q_{ks} / \sum_{k=1}^{K} p_{ks} q_{ks} \tag{3.13}$$

$$\omega_{kt} = p_{kt} q_{kt} / \sum_{k=1}^{K} p_{kt} q_{kt}$$

ω_{ks}, ω_{kt} : Value shares of k - th good in s - th and t - th period

K : total number of goods

In equation (3.13), the Laspeyers quantity index uses base period prices to weight quantity changes, whereas the Paasche quantity indexes use comparison period prices to weight quantity changes. The Fisher is the geometric mean of Laspeyers and Paasche quantity index, while the Tornqvist quantity index is a weighted geometric average of the quantity ratios between the two periods s and t, with weights given by the average of the value shares in periods s and t represented by those quantities. The log-change form of the Tornqvist in equation (3.13) is widely used to measure changes in output produced and inputs used in production over two time periods. The popular Tornqvist TFP index expressed in log-change form is as follows:

$$\ln TFP_{st} = \ln \frac{OuputIndex_{st}}{InputIndex_{st}} = \ln(OutputIndex_{st}) - \ln(InputIndex_{st}) = \\ = \frac{1}{2}\sum_{m=1}^{M}\left(R_{ms} + R_{mt}\right)\left(\ln Y_{mt} - \ln Y_{ms}\right) - \frac{1}{2}\sum_{n=1}^{N}\left(S_{ns} + S_{nt}\right)\left(\ln X_{nt} - \ln X_{ns}\right) \qquad (3.14)$$

where s and t are two time periods, M and N are total number of outputs and inputs respectively, R_{ms} and R_{mt} are output revenue shares, S_{ns} and S_{nt} are input cost shares, Y_{ms} and Y_{mt} are output indexes and X_{ns} and X_{nt} are input indexes.

The economic theory (or functional approach) and the axiomatic (or test) approach are used to assess the quality of an index number discussed above. The economic-theoretic approach to index numbers assumes that the firms are technically and allocatively efficient and involves revenue maximization and cost minimization (Coelli, et al 2005). Diewert (1976) showed that the Tornqvist output price index is derived by a translog revenue function and hence, it is considered to be exact for the translog revenue function. Also, it is considered to be superlative since the translog function is a flexible functional form (provides a second-order approximation to any arbitrary function). The Fisher output price index is exact for a quadratic function and thus superlative. Moreover, the Tornqvist input price index is derived by a translog cost function and therefore, it is considered to be exact for the translog cost function and hence superlative. The Fisher input price index is exact for a quadratic function and thus superlative. Furthermore, using distance functions, the Tornqvist output quantity index is derived by a tranlog output distance function and is considered to be exact for the translog output distance function and therefore superlative, whereas the Tornqvist input quantity index is derived by a tranlog input distance function and hence, it is considered to be exact for the translog input distance function and superlative. Finally, the Fisher output and input quantity indices are exact and superlative for a quadratic distance function.

The axiomatic approach chooses the index number formula based on a number of several properties, tests or axioms that the index should satisfy. Diewert and Nakamura (2003) and Diewert and Lawrence (1999) refer to four axiomatic tests to choose among alternative index number formulas. Assuming only quantity indexes the four tests (or desirable properties) that an index number should satisfy (or exhibit) are the followings:

- Identity or Constant Quantities Test, which means that if the quantities are the same in two periods, then the quantity index should be the same in both periods irrespective of the price values in both periods.
- Constant Basket Test, which means that if the prices are constant over the two periods, 0 and 1, then the level of quantities in period 1 compared to period 0 will be equal with the value of the constant basket of prices evaluated at the period 1 quantities divided by the value of the constant basket of prices evaluated at the period 0 quantities.
- Proportional increase in quantity test, which means that if the quantities in period 1 are multiplied by a positive factor ξ, then the quantity index in period 1 should increase by this factor ξ, compared to the quantity index in period 0.
- Time reversal test, which means that if the prices and quantities in period 0 and 1 are interchanged, then the resulting quantity index should be the reciprocal of the original quantity index.

Fisher has all the desirable properties, whereas both the Laspeyres and Paasche indexes are not consistent with the time reversal test and the Tornqvist index does not satisfy the constant basket test (Mawson et al, 2003). Diewert (1992) showed that the Fisher index satisfies many more properties and thus justifies the label of "ideal index". Also, the Fisher index can handle zero values for inputs and outputs.

Two more terms regarding index numbers when we have longitudinal or *panel* data are used in this section. When comparing quantities and thus productivity over time, it is possible that we may want to compare the base period 0 with any period t quantities and productivity using any one of the index numbers discussed earlier. Such an index is called a *fixed-base index*. However, if we want to compare each year with the previous year and then combine annual changes in quantities and productivity to measure changes over a given number of periods, then the resulting index is defined as a *chain index* (Coelli et al, 2005).

In the case of *spatial* comparisons, i.e. when comparing quantities and thus productivity across firms in a single time period, we need to take into account comparisons across all pairs of firms (*multilateral)*. We would like these comparisons to be internally consistent, i.e., to satisfy the property of transitivity. Internal consistency (transitivity) requires that any direct comparison between any two firms should be the same as a possible indirect comparison between these two firms through a third firm. None of the four popular quantity indices discussed earlier satisfy the property of transitivity in a multilateral spatial context. Thus, we need to convert these indices into multilateral consistent transitive indices by using the EKS method developed by Elteto-Koves (1964) and Szulc (1964). Let's assume that we are working with Fisher quantity index numbers between firms. For all firms i and j, we use the EKS method to convert the Fisher indices into multilateral indices by calculating, $Q_{ij}^{F-EKS} = \prod_{k=1}^{I} \left[Q_{ik}^{F} \times Q_{kj}^{F} \right]^{\frac{1}{I}}$. These indices satisfy the following properties:

(i) Q_{ij}^{F-EKS}, for $i, j = 1,2,\ldots,I$, are transitive where I represents the total number of firms.

(ii) The new indices, Q_{ij}^{F-EKS}, deviate from the least from the original Fisher indices in a least-squares sense (Coelli et al, 2005).

Caves, Christensen and Diewert (1982a), applied the EKS method to derive multilateral Tornqvist indices that are transitive. The multilateral transitive version of a Tornqvist index is the geometric mean of all indirect comparisons between companies (or time periods) s and t through all possible link companies (or time periods). Caves et al (1982a) proved that the multilateral Tornqvist index that derives from a translog function has the following form:

$$\begin{aligned} LnTFP_{st}^{transitive} &= \left[\frac{1}{2} \sum_{m=1}^{M} \left(R_{mt} + \overline{R_m} \right)\left(\ln Y_{mt} - \overline{\ln Y_m} \right) - \frac{1}{2} \sum_{m=1}^{M} \left(R_{ms} + \overline{R_m} \right)\left(\ln Y_{ms} - \overline{\ln Y_m} \right) \right] \\ &- \left[\frac{1}{2} \sum_{n=1}^{N} \left(S_{nt} + \overline{S_n} \right)\left(\ln X_{nt} - \overline{\ln X_n} \right) - \frac{1}{2} \sum_{n=1}^{N} \left(S_{ns} + \overline{S_n} \right)\left(\ln X_{ns} - \overline{\ln X_n} \right) \right] \end{aligned} \quad (3.15)$$

Where the bar above the variables means arithmetic mean. In this case, a comparison between two firms (or time periods), s and t, is achieved by first comparing each firm with the average firm and then comparing the differences in firm levels relative to the average firm. Chapters 5 and 6 provide an application of index number techniques in a multilateral productivity context, across firms and over time and how the multilateral spatial productivity indices and temporal productivity indices can be combined together in order to measure relative productivity change over time.

3.4 Data Envelopment Analysis and Productivity Measurement

The previous section discussed index number techniques to measure productivity changes over time and space. This section analyzes techniques that measure efficiency and productivity relative to an estimated frontier. A very popular technique which is based on linear programming is Data Envelopment Analysis (**DEA**). DEA uses linear programming to construct a non-parametric piece-wise linear frontier over the data. Efficiencies of homogeneous decision making units (**DMU**) such as schools, bank branches or companies in a regulated industry are calculated relative to this frontier. Operating units on the frontier are 100% efficient, while units that do not operate on the efficient frontier are given an efficiency rating that reflects their distance from the frontier. DEA does not require the specification of a functional form or any assumption of inefficiency distribution so it does not run the risk of misspecification. Also, DEA can be less demanding of data points as it does not attempt to estimate a full "cost function" but only unit-specific results and a few "efficient units" would be enough to give a boundary reflection of an inefficient point. Finally, DEA assumes that deviations from the efficient frontier are attributable to inefficiency only and not to measurement error.

The usefulness of DEA analysis for assessing efficiencies and productivity measurement of DMUs is threefold. Firstly, we can assess the performance of units in an input/output orientation dealing with multiple-outputs and multiple-inputs without the need for including prices for output and inputs. Secondly, we can use value-based models in measuring efficiency among the units based on the importance of inputs and outputs for the production. The implicit values of inputs and outputs assigned by value-based DEA models give information on rates of substitution and transformation between the factors of production, which improves the usefulness of DEA analyses (Thanassoulis, 2001). Thirdly, DEA can be used to measure productivity change and efficiency change over time. This is done using the *Malmquist index* developed by Fare et al, (1994). Under constant returns to scale, the Malmquist index decomposes into two components, the *catch-up* (efficiency change) component and the *boundary shift* (technical change) component. Under variable returns to scale, the Malmquist index contains a third component, the *scale efficiency change*, which captures the impact on productivity of a change in scale size of the operating unit.

DEA was introduced by Charnes, Cooper and Rhodes (1978) where the authors provided an input oriented model under constant returns to scale (CCR model) and was further developed by Banker, Charnes and Cooper (1984) under variable returns to scale (BCC model). A graphical illustration of a DEA frontier is depicted in Figure 3.5, in the case where two inputs are used to produce a single output. DMUs (or firms) A, C and D are efficient since they lie on the DEA frontier whereas B lies inside the frontier and is inefficient. The projection of B to the frontier gives the inputs that B should use to get the output it already produces. The efficiency of B is the fraction to which it could reduce radially its input levels without detriment to its output. This means that the ratio of 0B' /0B gives the measure of technical efficiency for company B. The projected point of B, B', lies on the line joining points A and C. This means B' is a virtual firm created from A and C to act as benchmark for B. Thus, firms A and C are referred as *referent peers* or simply *peers* for firm B.

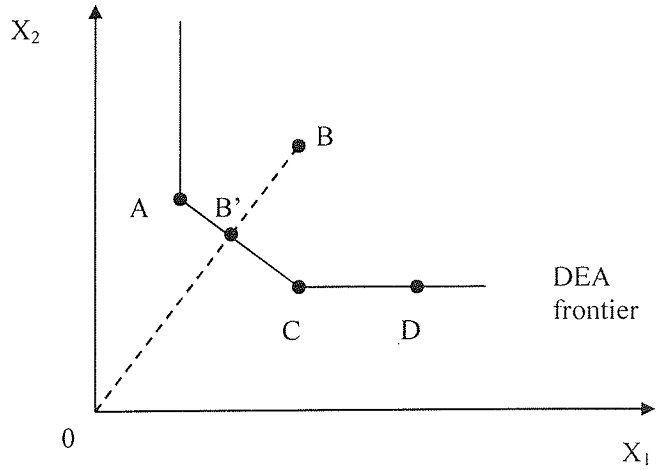


Figure 3-5 Measuring Efficiency Using DEA

In the general case, the Farrell measure of efficiency can be computed using linear programming as follows. Consider, for example, that we have I DMUs (i=1,2,..,I) using N inputs to produce M outputs. Let's define x_{ni} and y_{mi} the observed level of the nth input and mth output respectively for each DMU i. The following linear programming models can be solved to ascertain whether DMU i_o is DEA efficient and measure its efficiency:

$$Min\left\{\theta_o\left|\begin{array}{l}\sum_{i=1}^{I}\lambda_i x_{ni}+s_n^- = \theta_o x_{no}, n=1,\dots,N\\ \sum_{i=1}^{I}\lambda_i y_{mi}-s_m^+ = y_{mo}, m=1,\dots,M\\ \\ \lambda_i \geq 0, i=1,\dots,I\\ s_n^-, s_m^+ \geq 0\\ s_n, s_m; slacks\end{array}\right.\right\} \qquad (3.16)$$

$$Max\left\{z_o\left|\begin{array}{l}\sum_{i=1}^{I}\lambda_i x_{ni}+s_n^- = x_{no}, n=1,\dots,N\\ \sum_{i=1}^{I}\lambda_i y_{mi}-s_m^+ = z_o y_{mo}, m=1,\dots,M\\ \\ \lambda_i \geq 0, i=1,\dots,I\\ s_n^-, s_m^+ \geq 0\\ s_n, s_m; slacks\end{array}\right.\right\} \qquad (3.17)$$

The technical input efficiency is θ_o^*, the optimal value of θ_o in model (3.16) and the technical output efficiency of DMU i_o is $1/z_o^*$ where $z_o^*$is the optimal value of z_o in models (3.17). Charnes et al, (1978) considered only CRS technologies. However, the above linear programming models can considered for VRS technologies if we set the sum of lambdas equal to 1. If $\theta_o^* \prec 1$, then DMU i_o is not efficient in the models (3.16) because the model will have identify another unit that secures at least the output vector y_o but using no more than the reduced input vector $\theta_o^* x_o$. Similarly, if $z_o^* > 1$, then DMU i_o is not efficient in the models (3.17) because the model will have identified another unit that secures at least the augmented output vector $z_o^* y_o$ but using no more than the input vector x_o. If $\theta_o^* = 1$ or $z_o^* = 1$, then the DMU i_o is 100% efficient. However, DMU when 100% efficient is not necessarily Pareto-efficient because simultaneous improvements to inputs or alternatively to outputs may not be possible but improvements to the individual levels of some inputs or outputs may be possible. Such improvements are captured in the slacks of models (3.16) and (3.17) (Thanassoulis et al, 2008). Therefore, in an input oriented case, a DMU is Pareto-efficient if it is not possible to lower anyone of its input levels, without increasing at least another one of its input levels and/or without lowering at least one of its output levels (Thanassoulis, 2001). This implies that a DMU i_o is Pareto-efficient if and only if $\theta_o^* = 1, s_n^* = 0$ and $s_m^* = 0$ in (3.16). Similarly in an output oriented case, a DMU is Pareto-efficient if it is not possible to raise anyone of its outputs levels, without lowering at least another one of its outputs levels and/or without increasing at least one of its input levels (Thanassoulis, 2001). This a DMU i_o is Pareto-efficient if and only if in (3.17) we have $z_o^* = 1, s_n^* = 0$ and $s_m^* = 0$.

The duals of models (3.16) and (3.17) are shown in (3.18) and (3.19) respectively for the case where $\sum \lambda_i = 1$:

$$Max \left\{ h_o = \sum_{m=1}^{M} \rho_m y_{mo} + g \left| \begin{array}{l} -\sum_{n=1}^{N} \mu_n x_{ni} + \sum_{m=1}^{M} \rho_m y_{mi} + g \leq 0 \\ \sum_{n=1}^{N} \mu_n x_{no} = 1, n = 1, \ldots, N \\ \\ \rho_m, \mu_n \geq 0, \\ i = 1,..i_o., I \\ g \quad \text{is} \quad free \end{array} \right. \right\} \tag{3.18}$$

$$Min \left\{ \varphi_o = \sum_{n=1}^{N} \mu_n x_{no} + g \left| \begin{array}{l} \sum_{n=1}^{N} \mu_n x_{ni} - \sum_{m=1}^{M} \rho_m y_{mi} + g \geq 0 \\ \sum_{m=1}^{M} \rho_m y_{mo} = 1, m = 1, \ldots, M \\ \\ \rho_m, \mu_n \geq 0, \\ i = 1,..i_o., I \\ g \quad \text{is} \quad free \end{array} \right. \right\} \tag{3.19}$$

The duals in (3.18) and (3.19) are specified for a VRS technology but a CRS technology can be derived by setting g equal to zero. The above models are called "value-based" DEA models. The variables ρ_m and μ_n are the weights that a DMU "assigns" to each of its inputs and outputs respectively so that its efficiency will be maximized. Their optimal values, ρ_m^* and μ_n^* are in effect the imputed value per unit of output m and input n respectively. That is they yield information on rates of substitution and transformation between the factors of production, which improves the usefulness of DEA analyses (Thanassoulis, 2001).

As noted earlier the measurement of productivity change over time using DEA requires the calculation of a Malmquist Index of productivity change. The Malmquist productivity index, first introduced by Fare et al, (1989) and calculated by using distance function in an input/output oriented case, consists of two components, efficiency change (catch-up) and technical change (frontier shift), assuming constant returns to scale The authors in an another study (1994) proposed another decomposition of the Malmquist index by taking into account the impact of scale changes on productivity when the technology is not CRS. Figure 3.6 shows the graphical illustration of the Malmquist productivity index between two time periods.

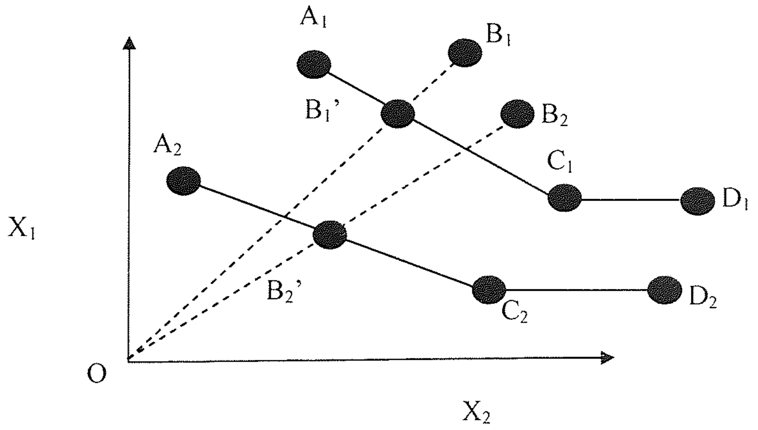


Figure 3-6 Malmquist Index of Productivity Change

The units A_1, B_1, C_1 and D_1 correspond to period 1, while the rest to period 2. The units B_1' and B_2' are the virtually efficient units to which B_1 and B_2 are compared respectively. However, in order to extract a useful conclusion about the change of the productivity of B over these two periods one must take into consideration the shifting down of the boundary, meaning that the technology allowed for a total reduction of the inputs needed for the given level of the outputs. That means that even though we can see that the technical efficiency of B_2 is less than that of B_1 the technological change (boundary shift) has favoured the industry as a whole.

Here we use the terminology in Thanassoulis (2001) to substitute distance functions with technical efficiency scores recalling that $TE_O = D_O(Y, X)$ and $TE_I = 1/D_I(Y, X)$ from equations (3.5) and (3.6). The general formula for the Malmquist Index of a firm i between time period t and t+1 in the input oriented case is the following (Thanassoulis, 2001):

$$MI_i = \left[\frac{CRS\_EF_{Tt}^{Dt+1} CRS\_EF_{Tt+1}^{Dt+1}}{CRS\_EF_{Tt}^{Dt} CRS\_EF_{Tt+1}^{Dt}} \right]^{1/2} \tag{3.20}$$

Where $CRS\_EF_{Tt}^{Dt}$ is the technical input efficiency of firm i calculated using its data of period t (Dt) relative to the efficient frontier (technology) of period t (Tt) under the assumption of constant returns to scale. Then, after some simple algebraic transformations we get that:

$$MI_i = \left[\frac{CRS\_EF_{Tt+1}^{Dt+1}}{CRS\_EF_{Tt}^{Dt}} \right] \times \left[\frac{CRS\_EF_{Tt}^{Dt+1} CRS\_EF_{Tt}^{Dt}}{CRS\_EF_{Tt+1}^{Dt+1} CRS\_EF_{Tt+1}^{Dt}} \right]^{1/2} \tag{3.21}$$

Where the first component outside the brackets measures the efficiency change (catch-up), which shows how much closer or further from the boundaries the firm has moved over the two periods. The second term in (3.21) is defined as *technical change (frontier shift)* which indicates the movement of the boundary itself over time (Thanassoulis 2001). Assuming that a firm i operates under variable returns to scale then the Malmquist productivity index at equation (3.21) is further decomposed into another factor known as scale efficiency catch-up , which captures the impact of any change in scale size of a firm and is calculated by the expression, $CRS\_EF = VRS\_EF * SC\_EF$. Thus, the Malmquist productivity index of any firm i under variable returns to scale can be written as:

$$MI_i = \underbrace{\frac{VRS\_EF_{Tt+1}^{Dt+1}}{VRS\_EF_{Tt}^{Dt}}}_{\substack{\text{Pure technical}\\ \text{Efficiency}\\ \text{Catch-up}}} \times \underbrace{\frac{SC\_EF_{Tt+1}^{Dt+1}}{SC\_EF_{Tt}^{Dt}}}_{\substack{\text{Scale}\\ \text{Efficiency}\\ \text{Catch-up}}} \times \underbrace{\left[\frac{CRS\_EF_{Tt}^{Dt+1}}{CRS\_EF_{Tt+1}^{Dt+1}} \times \frac{CRS\_EF_{Tt}^{Dt}}{CRS\_EF_{Tt+1}^{Dt}}\right]^{1/2}}_{\substack{\text{Boundary shift –}\\ \text{Technical change}}} \qquad (3.22)$$

Further decompositions of the Malmquist productivity index come from Fare et al, 1995) by including quality into the technology and Lovell (2004) by incorporating the activity effect, output mix and input mix effect, output bias and input bias of technical change. We shall define further and use the decompositions by Lovell in Chapter 7. Moreover, Ouellete & Vierstraete (2004) proposed a modified Malmquist productivity index introducing quasi-fixed inputs, whereas Pastor and Lovel (2005) suggested a circular Malmquist productivity index by computing the distances needed for the index relative to a single frontier rather than in relation to more than one frontier. The index was further extended by Portela and Thanassoulis (2008) allowing for the comparison in the productivities of two units either at the same or at two different points in time.

Finally, Tulkens & Vanden Eeckaut (1995) proposed the construction of three different frontiers using DEA for panel data analysis, the contemporaneous, sequential and intertemporal frontier. A contemporaneous frontier assumes the construction of a reference production set at each point in time t, from the observations made at that time only. Thus, DEA efficiency scores can be computed for each year by taking each cross-section of the observation period separately in the DEA model. A sequential frontier allows the current period technology set to be constructed from data of all the companies in all years prior to and including the current period. Thus, technologies in previous periods are "not forgotten" and remain available for adoption in the current period. Finally, an intertemporal frontier assumes the construction of a single production set from the observations made throughout the whole observation period. It merges the data for all periods into one set and DEA efficiency scores can be calculated for the entire data set.

3.5 *Parametric Methods for Efficiency and Productivity Measurement*

This section discusses the parametric approach to relative efficiency analysis, non-frontier methods, simple regression analysis and frontier methods, stochastic frontier analysis.

A graphical illustration of two parametric methods, ordinary least squares (OLS) and corrected ordinary least squares (COLS) can be seen at Figure 3.7, where the horizontal axis represents output and the vertical axis costs. The regression model can be written as follows:

$$C_i = f(y_i : \beta) + u_i \tag{3.23}$$

Where C_i denotes total cost, y_i is the output vector, β is a vector of parameters to be estimated, f characterizes the relationship between C_i and y_i and u_i is the estimate of inefficiency for firm i. The simple regression analysis consists of two steps. The method of ordinary least squares (OLS) identifies the relationship between cost and cost drivers (outputs) that fits the data best by minimizing the sum of the squared residuals, the residual being the difference between the predicted and the observed cost value. The residuals are used to measure inefficiency. Corrected ordinary least squares can be considered as an extension of ordinary least squares as it further shifts the estimated cost function downwards until all the residuals are positive and none is negative. The company or companies found to have zero residual are efficient. Both OLS regression and corrected ordinary least squares are subject to criticism due to their limitations, such as no assumption about measurement error, degrees of freedom, and multicollinearity problems. Advantages of theses methods can be considered the specification of a functional for cost or production, which can be tested via statistical tests and the inclusion of environmental factors that are out of control of management. For an illustration of OLS and COLS consider Figure 3.7. In OLS companies A and B lie above the regression "fitted" line and are considered relatively inefficient while companies C and D lie below the OLS regression line and are considered as relatively efficient. In contrast under COLS all bar company D are inefficient.

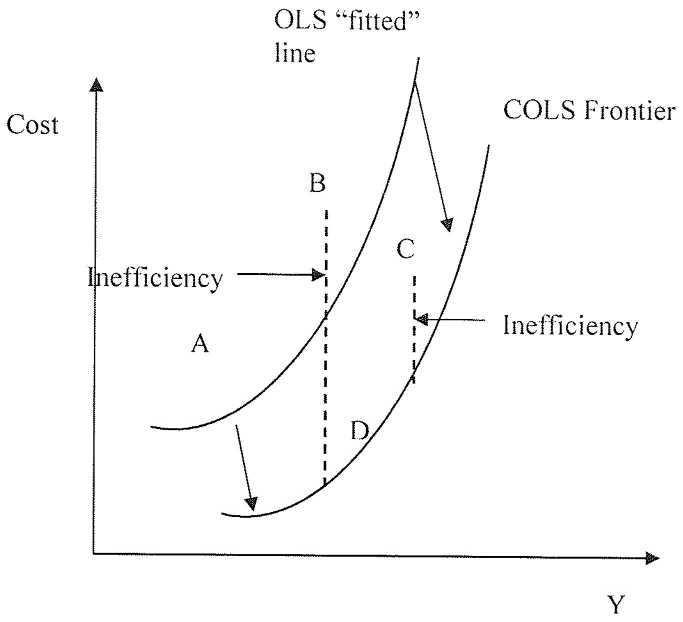


Figure 3-7 OLS and COLS Efficiency

Stochastic frontier analysis (SFA) differs from OLS and COLS in several ways. Stochastic frontier models allow for technical efficiency but they also acknowledge the fact that random shocks outside the control of producers can affect output. The most widely used stochastic frontier models include the stochastic production frontier model, stochastic cost frontier model, and stochastic distance function models. The stochastic frontier model can be written as follows:

$$C_i = f(y_i : \beta) + v_i + u_i \qquad (3.24)$$

Where $f(y_i : \beta)$ is the cost frontier, v_i denotes the error term due to noise which has a mean at zero and u_i is the inefficiency for firm i which takes non negative values. The inefficiency terms is assumed to follow the half-normal, exponential, truncated normal or gamma distribution. Stochastic frontier models are estimated using the maximum likelihood technique. SFA uses available data to estimate the cost function of a relatively efficient firm, known as the "frontier" and this cost function is assumed to be common for all firms and is used to obtain measures of inefficiency (Sarafidis, 2002). Any deviations from the frontier are attributed to both measurement error and inefficiency providing more accurate measures of relative efficiency. A graphical illustration of SFA is shown at Figure 3.8. Firm B's deviation from the SFA frontier is attributed to both inefficiency and measurement error (noise), whereas for firm D which lies below the SFA frontier, the noise is larger than the inefficiency and it is reducing costs so the firm appears 'super-efficient'.

SFA allows the inclusion of environmental factors either directly in the cost or production function or in the efficiency term. Also, it allows statistical inference on the specification of the functional form of the cost or production but due to small sample size (number of explanatory variables, quality and nature of data) it is likely that it will detect little or no inefficiency and collapse to simple OLS regression (residuals "wrong skew"). The strong advantage of the SFA method is that it can be used in a panel context (including more information and increasing the sample size than cross section techniques) through fixed or random effects models assuming time invariant or time varying inefficiency (Battese & Coelli, 1992 and Kumbhakar, 1990), thus allowing the measurement of productivity change over time and further decomposition into efficiency change, technical change and scale change.

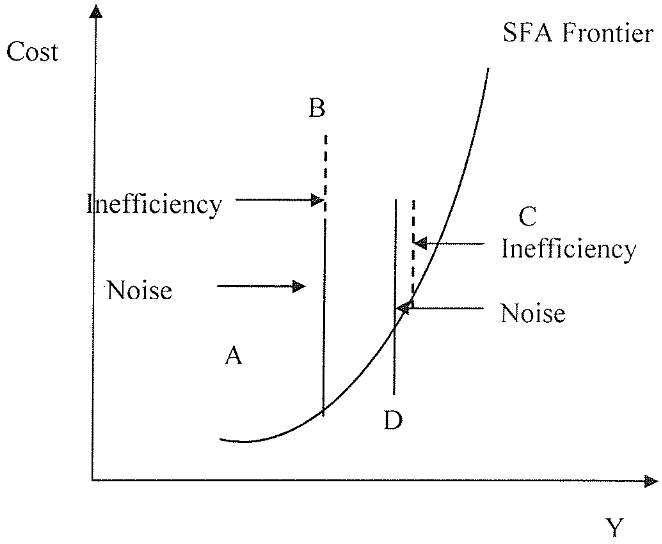


Figure 3-8 Stochastic Frontier Analysis

3.6 Ofwat's Approach to Comparative Efficiency Analysis and Productivity Measurement

As we will be assessing water company efficiency and productivity in chapters 5, 6 and 7 we conclude this chapter with a brief outline of how OFWAT uses the foregoing methods in assessing the comparative efficiencies of English and Welsh water companies.

Ofwat uses cross section econometric techniques and unit cost models to assess operating expenditure (OPEX) and capital expenditure (CAPEX) relative efficiency separately, for water and sewerage services. Operating expenditure (OPEX) relative efficiency for water services is derived by the use of four econometric models: *water distribution, resources and treatment, power and business activities.* Efficiency of sewerage services is derived using two econometric models, namely *network including power* and *large sewerage treatment works* and three unit cost models, namely *small sewerage treatment works*, *sludge treatment and disposal* and *business activities*. Efficiency assessment is carried out at the level of function (water and sewerage) and aggregated to give a separate aggregate measure for OPEX and CAPEX. The separability of each function and service (OPEX and CAPEX) was criticised by Stone & Webster (2004) where scope economies in water and sewerage industry were found. Saal et al (2005) emphasized that it was not appropriate to assume a common frontier for WaSCs and WoCs since this would lead to biased estimates in efficiency and productivity change. The source of the bias was the non-separability of the WaSC's water and sewerage operations.

The selection of the benchmark company that defines the efficient frontier is done under COLS. This is done after taking into account company specific factors and adjusting the residuals, a 10% reduction to the water services residuals for OPEX and CAPEX and 20% reduction for the sewerage services residuals (Ofwat, 2004) before placing the companies in efficiency bandings for OPEX and CAPEX separately. The relative efficient bandings for OPEX and CAPEX for water and sewerage services and the catch-up to the benchmark costs are applied as follows. We firstly define the relative efficient banding for OPEX efficiency for water and sewerage services. Ofwat defines the most efficient companies as band A. These are the companies that need to reduce their operating costs within 5% of benchmark costs (0%-5%). The next band is B. These are companies which need to reduce their costs up to 15% in order to achieve the efficiency level of a company in band A (between 5% and 15% of benchmark). The next band is C, where these are the companies which need to reduce their costs up to 25% relative to the benchmark costs (between 15% and 25% of benchmark). The next band is D, where the companies placed in this band need to reduce their costs up to 35% relative to benchmark costs (between 25% and 35% of benchmark) where the band E includes the companies that need to reduce their operating costs more than 35% relative to benchmark costs. Moreover, the relative CAPEX efficiency bands for water and sewerage services are applied as follows. Ofwat defines the most efficient companies as band A, where the companies need to reduce their costs up to 10% of benchmark costs, band B includes the companies whose costs need to be reduced up to 20% relative to benchmark costs (between 10% and 20% of benchmark). The next band is C, which requires a reduction in companies' costs up to 30% relative to benchmark costs (between 20% and 30% of benchmark), band D includes companies that need to reduce their costs up to 40% of benchmark (between 30% and 40% of benchmark) and finally the band E requires the companies to reduce their costs more than 40% of benchmark.

The catch-up factor for water and sewerage operating expenditure are calculated under the assumption that the inefficient companies should close 60% of the OPEX and 40% of the CAPEX gap they have with the most efficient companies (in band A). To illustrate this let us assume that a firm belongs to band B. Band B is separated into an Upper band B (5% - 10% away from the benchmark firm) and a Lower band B (10% - 15% away from the benchmark firm). We assume now that the firm is at the midpoint of the Upper band B (7.5% away from the benchmark firm). Then, the firm must reduce its operating costs by 7.5% * 60% = 4.5% over the regulatory period which is 0.9% reduction in operating costs per year over 5 years (catch-up). Companies that improve their performance are awarded + 0.5% on price limits while companies which did not perform well are penalized by – 1.0 down in the first year of the next price review (Ofwat, 2004).

3.7 Conclusions

This chapter provided an overview of comparative efficiency and productivity measurement techniques. It started with a theoretical presentation of a production technology where the concept of input/output distance functions, technical and allocative efficiency, overall cost and revenue efficiency and scale efficiency are discussed. Then, the chapter proceeded with the popular techniques that can be used to measure relative efficiency and productivity. A detailed analysis of non-parametric methods was presented, which included non-frontier methods, index numbers and frontier methods, data envelopment analysis which are applied in this thesis. Following that, there was a brief presentation on parametric methods, which included non-frontier methods, simple regression analysis and frontier methods, stochastic frontier analysis, followed by a discussion on Ofwat's approach to comparative efficiency analysis and productivity measurement.

CHAPTER 4 LITERATURE REVIEW OF ASSESSMENTS OF EFFICIENCY AND PRODUCTIVITY OF WATER COMPANIES

4.1 Introduction

This section reviews studies that had employed regression analysis and stochastic frontier analysis, index number techniques and data envelopment analysis methods to measure relative efficiency and productivity change in the water and sewerage sector in England and Wales before and after privatization and in countries other than the UK. Studies that were concerned with the cost structure of the water industry, the issue of economies of scale and scope are also included in this chapter.

4.2 UK Studies

There was a wide range of variables that were used when assessing productivity and efficiency in water industry and examining the cost structure of the UK water industry. In the determination of cost functions a number of variables needed to be used including some variant of cost as dependent variable such as total costs (see Hunt & Lynk, 1995, Ashton, 2000a, Saal & Parker, 2000, 2006, 2007, Stone & Webster Consultants, 2004, Erbetta & Cave, 2006), variable costs (Botasso & Conti, 2003, Stone & Webster, 2004) or only operating costs (Lynk, 1993, Cubbin & Tzanidakis, 1998, Ashton, 2000b) as well as the level of output and input prices as independent variables. There was a large number of variables that were used as outputs such as water supplied, volumes of trade effluent, water delivered, length of mains and the proportion of water delivered to non-households, water and sewerage connected properties, resident water supply population, equivalent sewerage treatment population, sewerage resident population served, sewerage treatment load. Also, Saal & Parker (2000, 2007) and Erbetta & Cave (2006) employed quality adjusted measures of water and sewerage output calculated as the product of water output and drinking water quality index and the product of sewerage output and sewerage treatment quality index (a weighted average of river quality and bathing water quality relative to the quality level of all England and Wales). Inputs were often used as a measure of physical capital stock based on the modern equivalent asset (MEA) estimation of replacement cost of water operations, labour and other costs defined as the difference between operating costs and total labour costs. Often other variables that might drive costs were also included such as drinking water quality, properties below the reference level for water pressure, number of properties with supply interruptions of more than 12 hours, number of properties regarded as at risk of sewer flooding, average pumping head, % of metered billed properties, % of water from river sources and waste-water from trade effluent customers, the proportion of water abstracted from underground sources, % of water losses, water and sewerage population density, bathing water intensity, the % of connected water properties that were metered, distribution input (Stone & Webster Consultants, 2004, Saal & Parker, 2006, Saal et al, 2007, Erbetta & Cave, 2006, Bottaso & Conti, 2003). Dummy variables for time, geography, ownership and price reviews were also included in the econometric specifications.

In contrast, the DEA analysis has relied on a small number of variables using operating expenditure as the sole input (Cubbin & Tzanidakis, 1998, Thanassoulis, 2000a and 2000b, 2002, Portela et al, 2009) or capital (MEA), labour and other costs as inputs (Erbetta & Cave, 2006). Outputs, when DEA analysis was undertaken, included water delivered, length of mains, the proportion of water delivered to non-households, pipe bursts, volume of water delivered, water and sewerage connected properties, water distributed (distribution input) from surface resources and non-surface resources multiplied with average pumping head, household and non-household billed properties adjusted to capture differences in the cost of service between urban and rural areas, the number of sources and quality adjusted outputs for water and sewerage services (Cubbin & Tzanidakis, 1998, Thanassoulis, 2000a and 2000b, 2002, Portela et al, 2009, Erbetta & Cave, 2006). Finally, index number techniques allowed for the construction of an aggregate quality unadjusted output index including resident water supply population and equivalent sewerage treatment population as outputs, an aggregate quality adjusted output index based on drinking water quality index and sewerage treatment quality index and the construction of an aggregate input index using capital (MEA), labour and other costs as inputs (Saal & Parker, 2001).

Looking at the results from the studies below we conclude that panel stochastic frontier techniques, DEA and index number methods can be employed to assess water and sewerage operations performance as the number of observations increases over time. Thanassoulis (2002) also underlined the usefulness of DEA in the regulation of water industry by using sub-company data, the potential division of water and sewerage companies' operations into smaller parts, making these parts, units of assessment. Moreover, Saal & Parker (2006) suggested that it is in appropriate to assess jointly the performance of both WaSCs and WoCs water operations since this leads to biases in the efficiency and productivity measures.

Ashton (2000a) concluded that privatization did not appear to have increased the level of productivity growth and technical change for WaSCs since 1989. Saal & Parker (2000) confirmed that privatization did not lead to increased efficiency for WaSCs although there was a significant reduction in the trend growth of costs which occurred after the 1994/95 price review. Saal et, al (2007) reported that a reduction in WaSCs' productivity was mainly attributed to reduced efficiency change, negative scale efficiency, whereas technical change substantially increased. Moreover, Saal & Parker (2001) employing an index number technique showed that privatization did not lead to a significant improvement in the overall productivity growth of WaSCs, whereas the 1994/95 price review reduced firm-specific economic profitability but there were not any significant improvements in productivity. However, in another study by Saal & Parker (2006) indicated that since 1993 there were productivity improvements for WaSCs and WoCs with technical change being the major determinant and efficiency and scale change having the smallest impact on productivity growth. This was confirmed by Portela et al (2009) where the authors employed DEA techniques to conclude that technical change was the main driver in the productivity improvement for WaSCs and WoCs when operating expenditure was used as the sole input. Moreover, Ashton (2000b) and Botasso & Conti (2003) showed that cost efficiency for WaSCs and WoCs increased over time, whereas Erbetta & Cave (2006) concluded that both technical and allocative efficiency improved since privatization, while the tightened 1999/00 price review had a positive impact on the technical efficiency of WaSCs, but the 1994/95 price did not.

Overall, the above studies showed that regulation had a positive impact of the performance of the UK regulated water and sewerage sector but privatization did not. Any improvements in productivity were mainly attributed to technical change. With respect to the issue of economies of scale and scope, Ashton (2000a and 2000b), Stone & Webster Consultants (2004), Saal & Parker (2000) and Saal et al (2007) found that WaSCs operated under decreasing returns to scale whereas WoCs under constant returns to scale. Economies of scope with the joint production of water supply and sewage, and water supply and environmental services within WaSCs was found by Lynk (1993) and Hunt and Lynk (1995), while Saal & Parker (2000) found significant quality adjusted economies of scope between water and sewerage services. Stone & Webster Consultants (2004) suggested that there were statistically significant diseconomies of scope from the overall integration of water and sewerage services, however, costs savings can be achieved between water production and distribution to connected properties for WoCs and between water treatment and sewerage treatment and disposal for WaSCs.

A detailed examination of studies in the UK water and sewerage sector is presented below. We begin with Lynk (1993) where a stochastic cost frontier model was used in order to estimate the efficiency of WaSCs and WoCs for the pre-privatization period 1979-80 to 1987-88. The authors used as outputs water supply measured in megalitres per day, trade effluent volume per day and environmental services defined as a turnover value and include components such as water quality regulation, pollution alleviation, recreation and amenity, navigation, fisheries and charges for environmental services. Operating cost was used as the cost variable and labour cost was defined as the ratio of total labour costs and total employment. Dummy variables to capture technical change for each year and the impact of regional location on operating costs were also included. The estimated results indicated that WaSCs showed lower levels of inefficiency than WoCs before privatization and economies of scope with the joint production of water supply and sewage, and water supply and environmental services within WaSCs were found. The presence of economies of scope between water supply, sewage and environmental services within WaSCs was confirmed in another study by Hunt and Lynk (1995). The authors used a multi-product cost function and a measure of total costs was regressed on the previous period's costs, each measure of output, labour price, technical change and the interaction of each combination of outputs. Outputs and costs were taken from Lynk's study discussed above.

.Cubbin and Tzanidakis (1998) used regression analysis, Corrected Ordinary Least Squares (COLS) and Data Envelopment Analysis (DEA) to assess the performance of the UK regulated water and sewerage industry by using data from the year 1994/95. Operating expenditure was used as the sole input variable and water delivered, length of mains and the proportion of water delivered to non-households as outputs. DEA and COLS showed different results since only one company appeared to be 100% efficient by using the econometric approach. DEA under constant returns to scale had three out of 29 companies fully efficient whereas under variables returns to scale the number of fully efficient companies increased. The authors concluded that DEA is best used if the number of observations is large although regression analysis does not put individual weights on variables and as such may not be fair to individual firms, if they face peculiar operating conditions not reflected in the model used.

Ashton (2000a) estimated economies of scale, productivity growth and technical change for WaSCs by using a translog cost function for the years 1985-1997. The number of households connected to the water distribution system was used as an output while capital, labour and consumables as inputs. Economies of scale were estimated as a measure of cost elasticity with respect to output. Technical change was estimated as a measure of the elasticity of cost with respect to time. Estimates of total factor productivity were measured as a function of technical change, economies of scale and the change in output relative to input. The results suggested that there were substantial diseconomies of scale (0.678) implied that a doubling of output would lead in a 67.8% increase in costs. The author concluded that privatization did not appear to have increased the level of productivity growth and technical change since 1989. Evidence of diseconomies of scale (0.466) was found in another study by Ashton (2000b) where the author used a fixed effects panel data model to estimate the cost efficiency among WaSCs during the years 1987-1997. Overall operating cost efficiency was estimated to be 84% suggesting that WaSCs could produce output with 16 % less operating costs. The range of efficiency was 23 percentage points with a standard deviation of 8 percentage points meaning that there was a dispersion in efficiency among WaSCs.

Thanassoulis (2000a and 2000b) used DEA to estimate potential savings through improved operating efficiency in water distribution in UK using data obtained from Ofwat. Operating expenditure was defined as an input, while five outputs were used to assess the relative efficiency of WaSCs in water distribution, number of properties connected, length of mains, volume of water delivered and pipe bursts. The results indicated that only three outputs, the number of properties connected, the length of mains and the water delivered gave the more accurate reflection of company cost efficiency. DEA results underlined that the companies whose level of efficiency was under 100% should reduce their operating costs by 26.6% in order to move closer to the most efficient firms. The results from DEA (26.6% in operating savings) confirmed the predictions of Ofwat in the price review in 1999 that the inefficient companies should reduce their operating costs by 25%-35% in order to improve their performance toward to the top performing companies.

Thanassoulis (2002) used DEA to estimate potential cost savings in sewerage companies in England and Wales by taking data from the year 1992/93. Several factors such as the resident population, the length of sewers, the capacity of pumping in the sewerage network and the size of area served influence operating costs in sewerage network. Tests with different output sets led to finally using only the length of mains and the resident population as outputs. The results showed that there was scope for efficiency savings in sewerage, ranging from 9.55% of observed expenditure to 61.5%. The DEA results were compared with those obtained from regression analysis, which used more explanatory variables than DEA. Three companies, which appeared to be more efficient using the DEA approach, were now less efficient and achieved lower efficiency ranking. However, seven out of ten companies did not change more than two places in their efficiency ranking although regression analysis used more independent variables than the DEA technique did. The author emphasized the use of sub-company data, the potential division of water and sewerage companies' operations into smaller parts, making these parts, units of assessment. By dividing water and sewerage companies in several functions, several observations can be obtained from one company in one given time period in one function. For instance, in the case of sewerage function the sub-units are the districts of each company and thus the 10 WaSCs yielded 60 sewerage districts.

Saal and Parker (2001) used a temporal index approach to measure the impact of privatization and regulation on productivity, price performance and profitability for WaSCs for the years 1985-2000. An index for aggregate output captured both the quantity and quality of water and sewerage services. Water output was proxied by the resident water supply population served by each WaSCs and sewerage output was proxied by equivalent sewerage treatment population for each WaSCs. Drinking water quality index was calculated as the ratio of the average percentage of each WaSC's water supply zones that are fully compliant with key water quality parameters, relative to the average compliance percentage for England and Wales in 1990. Sewerage quality was calculated as a weighted average river quality and bather water quality for each WaSCs. The product of water output and drinking water quality was defined as a quality-adjusted water output and the product of sewerage output and sewerage quality index was defined as a quality-adjusted sewerage output. An index for aggregate inputs was calculated by using the Tornqvist index of labor, capital and other costs like materials and fuel usage. A quality-adjusted output price and an aggregate input price were also calculated. The results indicated that privatization did not lead to a significant improvement in the overall productivity growth of WaSCs, whereas the tightened 1994/95 price review reduced firm-specific economic profitability but there were not any significant improvements in productivity.

Saal and Parker (2000) estimated a translog multiple output cost function for WaSCs for the years 1985-1999 including the impact of drinking water quality and sewerage treatment quality on total costs. The results suggested that significant quality adjusted economies of scope cannot be rejected between water and sewerage services since quality influences costs. Also, the scale elasticity took a value below one, mean range for WaSCs 0.83-0.88, providing evidence of decreasing returns to scale, while the estimated parameter associated with the impact of price review on productivity growth was statistically significant and positive. The authors concluded that privatization did not increase efficiency but the tightened 1994/95 price review led to significant efficiency gains for WaSCs.

Stone and Webster Consultants (2004) used a translog cost model for WaScs and WoCs separately and a generalized quadratic cost model, which deals with zero outputs pooling WaSCs and WoCs to estimate the economies of scale and scope in water and sewerage industry for the period 1992/93 to 2002/03. The outputs used were water and sewerage service connected properties, water delivered and sewerage resident population served. Capital, labour, energy prices and other costs were used as inputs. Also, exogenous variables were included in the econometric estimation, which were defined as drinking water quality, environmental standards, properties below the reference level for water pressure, number of properties with supply interruptions of more than 12 hours, number of properties regarded as at risk of sewer flooding, average pumping head, % of metered billed properties, % of water from river sources and waste-water from trade effluent customers. The results indicated that operating characteristics and quality affected companies' costs and thus their inclusion in the estimation of cost functions was necessary. Moreover, WaSCs showed diseconomies of scale but these diseconomies were declining over the sample period, whereas WoCs reported very small economies of scale throughout the whole period. Regarding economies of scope, both the translog and the generalised quadratic specifications suggested that there were statistically significant diseconomies of scope from the overall integration of water and sewerage services, implying that the joint production of water and sewerage businesses did not lead to lower costs. Further consolidation in the form of WaSCs and WoCs mergers would not lead to any real economic benefits. However, costs savings can be achieved between water production and distribution to connected properties for WoCs and between water treatment and sewerage treatment and disposal for WaSCs.

Saal and Parker (2006) used a panel stochastic frontier model to estimate the overall performance of companies in the UK water and sewerage industry .The results indicated that it was not appropriate to assume a common frontier for WaSCs and WoCs since this would lead to biased estimates of efficiency and productivity change. The source of the bias was the non-separability of the WaSC's water and sewerage operations. Thus, while Ofwat assesses its water models jointly for WaSCs and WoCs to allow a greater number of observations, the authors provided clear evidence that we should reject the hypothesis that WaSCs and WoCs operate with the same technology, thereby undercutting the validity of Ofwat's efficiency assessment work. Jointly estimation of WaSCs and WoCs showed that average WaSC's productivity improved by 29.4% which was attributed to an increase in technical change by 29.3%, a small increase in efficiency change by 0.1% and a small reduction in scale change by 0.2%, whereas average WoCs' productivity increased by 9.9% and it was mainly attributed to an increase in technical change by 11%, a reduction in efficiency by 1% and a small increase in scale change 0.6%. Separation estimation of WaSCs and WoCs indicated that average WaSC's productivity improved by 20.4% which was attributed to an increase in technical change by 21.2%, a small reduction in efficiency change by 0.7% and a small increase in scale change by 0.2%, whereas average WoCs' productivity increased by 31.2% and it was mainly attributed to an increase in technical change by 31.6%, a reduction in efficiency and scale change by 0.3% and 1.4% respectively.

Also, Saal, Parker and Weyman-Jones (2007) used a fixed effects input distance function in a panel context to assess the impact of privatization on the productivity growth of WaSCs using data from 1985-2000. Quality adjusted outputs for water and sewerage services were calculated as the product of water connected properties and drinking water quality index discussed above and the product of sewerage connected properties and sewerage quality index also discussed above. Quality unadjusted water and sewerage outputs were proxied as water supply and sewerage treatment load. A physical measure of capital stock based on the modern equivalent asset (MEA) estimation of the replacement cost of net tangible fixed assets, total employment and other costs were used as inputs. Moreover, the generalized Malmquist index introduced by Orea (2002) was decomposed into technical efficiency, technical change and scale efficiency. The results indicated that by 2000 the impact of privatization and price cap regulation had not resulted in improved productivity rates, there was evidence of diseconomies of scale, whereas technical change was significantly high.

Erbetta and Cave (2006) used both DEA and SFA to assess the impact of regulatory policy on the efficiency of WaSCs in England and Wales for the period 1992-93 to 2004-05. Estimates of technical and allocative efficiency were obtained by running DEA. Then, the authors calculated input specific excess utilization and allocative distortion measures and regressed them on a set of environmental variables using SFA. Quality unadjusted outputs were proxied by the total number of household and non-household water and sewerage service-connected properties. Following Saal and Parker and Saal et al, (2001, 2007) approaches quality adjusted outputs were calculated as the product of the total volume of delivered potable plus non-potable water and the drinking water quality index and the product of physical amount of waste water and the river quality index. A physical measure of capital stock based on the MEA values, total employment and other costs were used as inputs. Environmental and regulatory variables were proxied by the proportion of water abstracted from underground sources, % of water losses, water and sewerage population density and dummy variables for time and price reviews 1994 and 1999 to capture technical change and regulatory impact. The results indicated that the tightened 1999/00 price review had a positive impact on the technical efficiency of WaSCs, whereas the 1994/95 price did not. Allocative efficiency improved significantly during the observed period, showing over-utilization of labour and under-utilization of capital at the beginning of privatization but these distortions declined over time, moving to a labour-saving and capital-augmenting industry.

Bottaso and Conti (2003) estimated a variable stochastic cost frontier model treating capital as quasi-fixed input to assess the operating cost efficiency for WaSCs and WoCs during the years 1995-2001. In order to control for the large firm's size variation (heteroskedasticity) the authors parameterized the variances of the measurement error as exponential function of size and the inefficiency terms as exponential function of various hedonic variables such as average pumping head, the proportion of river sources on total sources, population density, distribution input. The results suggested that there was a reduction in average cost inefficiency of about 5% over time and inefficiency variations among WaSCs and WoCs had steadily declined.

Finally, Portela et al (2009) used a circular Malmquist productivity index to measure the productivity change of WaSCs and WoCs for the period 1993-94 to 2006-07. The circular Malmquist productivity index was used to compute the distances needed for the index relative to a single frontier rather than in relation to more than one frontier, satisfying the circularity property and allowing for the comparison of the productivities of two units either at the same or at two different points in time. The authors used only operating expenditure as input and five outputs. The water distributed (distribution input) from surface resources and non-surface resources was multiplied with average pumping head to capture differences in water distribution costs. Household and non-household billed properties were also adjusted to capture differences in the cost of service between urban and rural areas. The number of sources was also included as an output. The authors did not include capital expenditure or differences in the quality of output between firms in their model. The results indicated that there were significant productivity improvements among WaSCs and WoCs during the years 1993 to 2005 and were mainly attributed to technical change rather efficiency change (catch-up). After 2005, a declining trend in productivity was found. There was a big improvement in productivity between 2000 and 2001 and there was a big decline in productivity between 2005 and 2006. Efficiency change contributed positively to an improvement in productivity change from 1998 to 1999 and from 2004 to 2005, whereas it contributed negatively to the productivity change from 2006 to 2007.

4.3 International studies

This section discusses studies that have carried out economic analyses of the performance of water and sewerage industries in countries other than the UK employing empirical modelling approaches such as regression analysis, data envelopment analysis, stochastic frontier analysis and index numbers. There was a wide range of variables that were used when assessing productivity and efficiency in water industry and examining the cost structure of the water industry in countries other than UK. In the determination of cost functions a number of variables needed to be used including some variant of cost as dependent variable such as total costs (see Kirpatrick et al, 2004, Fillipini et al, 2008, Fraquelli & Moiso, 2006 and Lin, 2005), variable costs (Vitialiano, 2005, Aubert and Reynaud, 2005 and Mosheim, 2006) or only operating costs (Estache & Rossi, 2002, Corton & Berg, 2009) as well as the level of output and input prices such as wages, capital, electricity, chemical and material prices as independent variables. There was a large number of variables that was used as outputs such as the average production in gallons per day, the number of connections, the total cubic meters of water delivered, water billed, the number of customers, volumes of water sold, the amount of water distributed, water quality production measured as the difference in turbidity between finished and raw water. Often other exogenous variables that might drive costs were also included such as number of hours of water availability per day (water quality), population density, percentage water from surface and underground sources, network length, percentage of metered connections and other quality variables such as quality of service, coverage of service attained, management efficiency, and management finance efficiency. Dummy variables for time, ownership, water purchased by other water utilities, different types of water treatment, the presence or absence of filtration plant and health violations (water quality), water losses were also included in the econometric specifications.

In contrast, the DEA analysis has relied on a large number of variables using labour, operational expenditure, capital, total costs, treatment plants, delivery network, sewer network, electricity measured as the amount of kilowatt-hour, costs in materials, chemicals, outside services, other costs and wastewater treatment costs as inputs. Outputs, when DEA analysis was undertaken, included volumes of treated sewage, population served water and treated sewage, number of connected properties, volumes of water delivered, cubic of meters of water supplied, controls performed capturing quality and potability of water, population served, water supply, water primary treatment and secondary treatment, complaints per 1000 water connections, volumes of water billed. Often in DEA analysis other exogenous factors were included to test the impact of efficiency in water utilities such as higher density areas and reduction in water leakages, social indicators such as the municipal population, the number of houses in a municipality, average people per house, the average temperature of the municipality, customer's income, quality measured as the amount of water supplied lost from pipelines due to inadequate maintenance, population density, non-residential users, unaccounted for water proxied as water loss divided by water produced, other variables to capture the impact of decentralization and regulation, the effect of economies of scale, scope and density, treatment, pumping and infrastructure expenses and time period (Tupper & Resende, 2004, Garcia Sanchez, 2006, Picazo-Tadeo et al, 2008, Anwandter and Ozuna, 2002, Byrnes et al, 2009). The results indicated that the inclusion of exogenous factors in DEA analysis leads to significant changes in the efficiency of water utilities if they were not included.

With respect to index number analysis, Corton & Berg (2009) used the volumes of water billed and the number of connections as outputs and labour and energy as inputs to construct the total factor productivity index. Finally, the issue of the optimal cost structure of water industry in other countries rather in UK was examined by Antonioli and Fillipini (2001), Fabbri and Fraquelli (2000), Garcia and Thomas (2001) where labour, capital and energy were used as inputs and total cubic meters of water distributed, the volume of water sold to final customers and water network losses as outputs. Several exogenous characteristics were also included in their econometric specification such as the number of customers served, the network size, the percentage of water loss in distribution pipes, population density, the cost of water input purchased by the firm and treatment costs, the number of metered connections and the number of local communities serviced by the water utility. A detailed examination of international studies in the water industry is discussed below.

Vitaliano (2005) used a stochastic variable cost model to estimate the capital cost and overall cost efficiency for 75 small municipal US water systems in 1999. The variable cost function included one output, the average production in gallons per day, capital stock was treated as quasi-fixed input, whereas data for wage and electricity prices and water quality proxied by the presence or absence of filtration plant and health violations were also used. The results indicated that there were significant increasing economies of scale, while the cost efficiency was 0.73, meaning that the water systems could further reduce their costs by 27%. Median capital stock inefficiency due to overinvestment was $70,500 per water system, whereas median cost inefficiency was $24,300. The combined cost of these two types of inefficiency was $663.6 million per year.

Estache & Rossi (2002) used a Cobb-Douglas stochastic cost frontier model to compare the efficiency of 50 publicly and privately owned water utilities in 29 Asian and Pacific region countries in 1995. The data covered operational costs, annual salary, number of clients, daily production, number of connections, population density and various water quality variables. The results suggested that there were not any significant differences between public and private operators. Public firms that needed to compete with new private entrant who enjoyed the latest technology would often be expected to obtain not only industry gains but also specific gains to offset firm-specific inefficiencies. Thus, the authors concluded that competition mattered more than ownership.

Tupper and Resende (2004) used DEA to assess the efficiency of 20 water and sewerage companies in Brazil over the period 1996-2000, where the price cap regulation is employed. The outputs used were water produced, treated sewage, population served water and populations served treated sewage, while the inputs included labour, operational costs and other operational costs. The change in the efficiency scores over time among the companies were attributed to regional heterogeneities. Thus, a Tobit model was then employed to regress the DEA efficiency scores on three variables, density of water, sewerage network and water loss. The results underlined that higher density areas and reduction in water leakages had a positive influence on efficiency. Due to the statistical significance of regional heterogeneities on the technical efficiency scores, the authors generated adjusted DEA scores by taking into account these regional factors. The original and adjusted DEA technical efficiency scores showed important differences since the adjusted DEA efficiency scores showed that there was only one company that was 100% efficient over time, while the original results showed that there were 10 companies 100% efficient over time.

Kirkpatrick, Parker and Zhang (2004) used statistical, data envelopment analysis and stochastic cost frontier analysis to assess the impact of privatization and regulation in water utilities in Africa in 2000. Data were taken for 71 water utilities from which 63 were publicly owned and 8 were under private ownership. The statistical and DEA results indicated that privatization lead to performance gains, while the stochastic frontier model underlined that while the coefficient on the ownership was negative, consistent with lower costs under private ownership, the result was not statistically significant.

Aubert & Reynaud (2005) employed a stochastic variable cost frontier analysis to assess the impact of different regulatory policies on the cost efficiency of 211 Wisconsin water utilities in the US over the period 1998 to 2000. Rate of return, hybrid rate of return and interim price cap regulation were assessed. Hybrid rate of return regulation is a process when the water utilities want to increase water and/or sewage rates and combines some aspects of a rate of return regulation with an upper bound on water price increases. Interim price cap regulation is applied when the water utility does not wish to change water or sewage prices and thus, the maximum allowable prices are those set by the regulator at the last price increase. The results suggested that on average the most efficient utilities were under a rate of return regulation, whereas the least efficient water utilities were under a hybrid rate of return. Water utilities under interim price cap regulation were quite efficient but not as much as those under rate of return regulation.

Mosheim (2006) used a variable shadow cost function to estimate and decompose economic efficiency among community water systems in the US in 1996 including quality and ownership effects. The results indicated that there was overinvestment in capital and the marginal cost of water quality varies by community water system's size, water quality production level and organizational type with small community water systems having higher marginal cost than large community systems. Also, the cost of allocative inefficiency was significant with smaller firms being much more cost inefficient than larger firms, whereas technical inefficiency was found to be much smaller than allocative inefficiency. Finally, ownership effect was insignificant regarding technical and allocative inefficiency and economies of density's results suggested that cost savings could be obtained by merging small water systems into medium ones but not medium into large ones.

Coelli and Walding (2006) employed DEA techniques to measure the performance of the urban water supply industry in Australia and eventually propose X-factors using data from the period 1995/96 to 2002/03. The authors used two outputs, number of connected properties and volumes of water delivered and two input variables, operating expenditure and capital. Their results indicated that the average firm had a mean technical efficiency of 90.4% and scale efficiency of 90.3%. Then, a Malmquist productivity index was calculated suggesting that average annual productivity change was equal to 1.2% decline per year which was decomposed into a 2.2% technical regress per year and 1.2% increase in technical efficiency per year.

Combining these results with 1.1% annual productivity growth for the industry, the authors proposed an average X-factor of 2% per year, meaning that the average firm should reduce unit costs in real terms by 2% per year.

Garcia Sanchez (2006) used DEA to measure technical and scale efficiency of water supply in Spain provided by private and public water utilities. The sample consisted of 24 towns with over 50,000 residents and the technique used to collect data for the variables later used in the model was the questionnaire. Total costs, total employment, treatment plants and the net of pipes were used as inputs. Cubic meters of water supplied, number of connections and controls performed capturing quality and potability of water were used as outputs. The author also included a set of ten factors in the model that are outside the control of managers, called social indicators such as the municipal population, the number of houses in a municipality, average people per house, the average temperature of the municipality, customer's income. The results indicated that on average technical efficiency was 86.58% and scale efficiency was 91.23%. Also, 19 out of the 24 units assessed, were found pure technically efficient and 6 out of these 19 units, were found scale inefficient. Two municipalities were scale efficient but technically inefficient and three municipalities were technically inefficient (pure technical and scale). The author concluded that ownership did not have a significant influence on efficiency, meaning that private ownership did not imply higher levels of efficiency.

Filippini, Hrovatin and Zoric (2008) employed four panel data stochastic frontier techniques to measure cost efficiency and economies of scale of Slovenian water distribution utilities over the period 1997-2003. Output was proxied by the total cubic meters of water delivered and prices for labour, material and capital, the number of customers served, the size of the service area, dummy variables to capture water losses, surface and underground water and shift in technology (time) were included in the estimation of the translog cost function. Four panel data stochastic frontier models were employed, a pooled frontier model estimated with maximum likelihood estimation (MLE), random-effects (RE) model estimated with generalized least squares estimation (GLS), RE model estimated with MLE and the true fixed effect (TFE) model estimated with MLE. The differences between the models were related to the different assumptions imposed on the error term, cost inefficiency and firm-specific effects. The results indicated that the TFE model performed better than the other three models, the pooled stochastic model reported average cost inefficiency of

22.5%, the RE (GLS) and RE (MLE) models 66.3 % and 50% respectively. The average cost inefficiency based on the TFE model was estimated to be 19.1%. Also, economies of scale were found in small-sized utilities while large utilities showed diseconomies of scale based on the results from the four models.

Fraquelli and Moiso (2005) employed SFA technique to examine the cost efficiency and economies of scale of the water sector in Italy over a period of 20-30 years for a total of 407 observations. Output was proxied by the total cubic meters of water delivered, prices for labour, material and capital, the network length, an indicator of water losses and shift in technology (time) were also included in the estimation of the translog cost function. The cost inefficiency was expressed as a function of exogenous variables firm and time specific like density which was calculated a ratio of the number of customers and the network length. The results showed that the average cost inefficiency was 28%, partially explained by network characteristics like density. The presence of economies of scale suggested that the situation could be improved by a reduced fragmentation at a local level.

Picazo-Tadeo et al, (2008) used DEA to estimate conventional quantity-based (without quality) and quality-adjusted scores of technical efficiency for water utilities in Andalusia (Spain). Three outputs were considered, population served, water delivered and treated sewerage. Inputs were delivery network, sewer network, total employment and operational costs. Quality was measured as the amount of water supplied lost from pipelines due to inadequate maintenance and was calculated as the difference between water introduced in delivery pipelines and water billed to final customers. Quality was treated as bad attribute and maximization of output was constrained by the need of maintaining the level of quality, i.e. the highest increase in outputs without diverting resources from producing quality. The results indicated that on average the quantity-based technical efficiency of water utilities was 77.3%, while on average the quality-adjusted technical efficiency was 85.1%. The authors concluded that service quality played a major role in measuring the performance of water utilities which was also confirmed by the statistically significance in the mean difference between quantity-based and quality-adjusted technical efficiency scores.

Lin (2005) employed a stochastic cost frontier model to measure the efficiency of the Peruvian water utilities over the period 1996-2001. Outputs were proxied by the water billed and the number of customers and prices for labour and capital, and a time trend to capture technical progress/regress were also included in the estimation of the cost function. Quality variables such as quality of service, coverage of service attained, management efficiency, and management finance efficiency were considered either as environmental variables that might influence the efficiency of a firm or as additional outputs of the cost function. The results suggested that quality variables were necessary to include them as additional outputs rather than as environmental variables. The ranking correlation was high between the models with or without the quality variables, however, rankings could change significantly for specific water utilities.

Anwandter and Ozuna (2002) used DEA to evaluate the effect of public sector reforms, decentralization- the responsibility of providing water supply and sewerage operations from states to the municipal level- and autonomous regulator on the efficiency of Mexican water utilities in 1995. Data on operations (inputs, outputs and operating characteristics) of 110 water utilities were obtained from a questionnaire which was completed by the managers of water companies. Total employment, electricity measured as the amount of kilowatt-hour, costs in materials, chemicals, outside services, other costs and wastewater treatment costs were used as inputs. No data were available for capital stock. Outputs were proxied by water supply, primary treatment and secondary treatment. The DEA results indicated that 51 out of 110 firms were found 100% efficient, while the range of the technical efficiency for the inefficient companies was between 30% and 99%. Then the DEA technical efficiency scores were regressed on a set of explanatory variables such as population density, non-residential users, unaccounted for water proxied as water loss divided by water produced and the impact of decentralization and regulation was captured by dummy variables. The results from the Tobit regression suggested that unaccounted for water and non-residential users had a significant effect on efficiency, however, decentralization and regulation did not.

Byrnes et al, (2009) used DEA to measure the relative technical efficiency and productivity of 52 urban water utilities in regional New South Wales (NSW) and Victoria in Australia over a four-year period 2000-2004. Total operating costs were used as input and total potable water supplied and complaints per 1000 connections as outputs. The complaints variable was transformed, "undesirable output" to be used as an output in the DEA models such that maximizing was akin to minimizing actual complaints. The DEA results suggested that the NSW water utilities showed higher levels of overall technical efficiency than the Victorian water utilities over the observed period. Victorian water utilities showed higher levels of pure technical efficiency and NSW utilities higher levels of scale efficiency. There was a decrease in NSW utilities' productivity on average over the four-year period by 10.4%, which was attributed to a 2.6% decrease in efficiency change and 7.9% reduction in technical change. With respect to Victoria water utilities' productivity, on average there was a reduction by 8.8%, which was attributed to a 4.6% decline in efficiency change and 4.3% reduction in technical change. Then, the overall, pure technical and scale efficiency scores were regressed on a set of operating characteristics using Tobit regression. A total of 19 variables were used to capture the effect of economies of scale, scope and density, treatment, pumping and infrastructure expenses, institutional (whether a water utility is located in Victoria or NSW) and time period on the three sets of the DEA efficiency scores. The results indicated that 8 out of those 19 variables such as population density, residential consumption, groundwater, institutional effect, had a significant impact on the DEA efficiency scores. The authors underlined the importance of including operating characteristics in measuring water performance because after controlling for those variables, Victorian water utilities were found to be on average 5% more efficient regarding overall efficiency and 13% more efficient by pure technical efficiency.

Corton and Berg (2009) used TFP, DEA and SFA to measure the efficiency of water service utilities in six countries in Central America region over a four-year period 2002-2006. Cost Rica, Guatemala and Honduras were represented in this study by two service providers of different sizes, small and large providers, while Panama, El Salvador and Nicaragua by only one operator. With respect to TFP analysis, two sets of TFP measures were calculated, one using volumes of water billed as the output and another using number of connections as the output. Two inputs were employed, labour and energy. The results showed that the utility in Panama was the only service provider exhibiting improved productivity over the observed period when considering both number of connections and volumes of water billed, with range between 31% and 53%. Service providers in Nicaragua and El Salvador showed a small decrease in productivity. For the DEA analysis, the outputs were the same as before and labour, network length and GNI- the value of services of goods and services bought by citizens of a country irrespective of where they live- were employed. The DEA results in 2005 suggested that the utilities in Panama, El Salvador, Nicaragua, small utilities in Cost Rica, Guatemala and Honduras were 100% efficient. Large utilities in Cost

Rica, Guatemala and Honduras were 63%, 85% and 99% efficient. For the SFA analysis, operating costs were used as the dependent variable, while the independent variables were GNI, water billed and price for labour and energy. A stochastic cost panel data frontier model using exponential form for the inefficiency term to explain changes in efficiency over time was estimated. The results indicated that inefficiency had not changed over time (negative sign and statistically insignificant). Also, by 2005 Nicaragua and Panama needed to reduce their operating costs by 2%, El Salvador, large and small utilities in Cost Rica by 3%, small utility in Guatemala by 4%, while large utilities in Guatemala and Honduras by 12% and 14% respectively.

Antonioli and Fillipini (2001) used a Cobb-Douglas variable cost function to examine the cost structure of 32 water distribution companies in Italy over the period 1991-1995. Three inputs were used, labour, capital and energy and a single output, the total cubic meters of water distributed. The number of customers served, the network size measured by the length of the network pipes and the percentage of water loss in distribution pipes were included as network characteristic variables. The results indicated that there would be efficiency losses if individual customers were served by more than one distribution water company, economies of output and density took a value greater than one and there was no clear evidence that larger service areas resulted in any economies in water distribution, economies of scale took a value lower than one suggesting there were diseconomies of scale.

Fabbri and Fraquelli (2000) used several forms of cost function, translog and Cobb-Douglas to estimate the effect of exogenous (hedonic) factors on the water companies' costs in Italy by using data for 173 water companies observed in 1991. Four hedonic variables were used namely the number of customers, a proxy for density calculated as a ratio of population served and length of pipes, the cost of water input purchased by the firm and treatment costs. Volumes of water delivered were used as a proxy for output and three inputs were used, capital, labour and energy. The authors concluded that the translog cost hedonic function fitted the data better compared to the other three models. Among the hedonic variables, the cost of water input purchased by the firm and treatment costs were not statistically significant, whereas an increase in the number of customers increased costs and an increase in the population density resulted in cost savings (ceteris paribus). Finally, evidence of increasing economies of scale was found by the estimation of the translog hedonic model.

Garcia and Thomas (2001) used a translog variable cost function to estimate economies of scale, scope and density for 55 water utilities in the Bordeaux region, south of France, over the period 1995-1997, assuming that capital is quasi-fixed in the short run. The cost function and the input cost shares were estimated by applying the Generalized Method of Moments (GMM) technique. In France, local communities under public or private ownership are responsible for the water supply to local customers. Instead of price cap or rate-of-return regulation, there is a contract between the private company and the local community or the municipality board makes the decisions on the way the water utility is managed in the case of public operation. Outputs were defined as the volume of water sold to final customers and water network losses, which was the difference between volumes distributed and volumes sold to final customers. The authors stated that the water utilities will not produce and sell water to final customers (desirable output) without "producing" lost water in the form of water network losses (undesirable output). Capital, labour and energy were employed as inputs, and a set of environmental characteristics captured by the number of metered connections and the number of local communities serviced by the water utility were also included in the estimation model. The results indicated that the joint production of a desirable output with an undesirable one was more profitable than increasing efficiency in the production of the desirable output. Thus, minimizing water losses was not a priority since the repair of leaks was very costly. Moreover, the elasticity of cost with respect to capital was higher than zero and significant, indicating that water utilities were not located on the optimal, long-run equilibrium path (over-investment in capital). The study also found evidence of short-run economies of scale as a 1% increase in water supply resulted in an increase in variable costs by 0.65%.

4.4 Conclusions

This chapter presented the reader with a critical review of the existing literature on efficiency and productivity measurement in the UK water and sewerage sector based on regression and stochastic frontier analysis, index numbers and data envelopment analysis techniques. It summarized the key findings from the studies, e.g. the impact of privatization and regulation on the performance of the regulated companies, the inputs, outputs and exogenous factors used, the evidence of economies of scale and scope in the UK water and sewerage industry, followed by a detailed discussion of these studies. Then, the chapter proceeded with a detailed presentation of studies that had carried out economic analyses of the performance of water and sewerage industries in countries other than the UK. We finally conclude that although there are several studies in the past that evaluated the impact of regulation on the performance of the regulated water and sewerage sector using mainly econometric techniques and DEA, there is only one study by Saal & Parker (2001), where the measurement of firm-specific profitability, productivity and price performance indices over time is allowed. However, this methodology did allow them to measure differences in the level of productivity, price performance and profitability across firms (relative comparative performance). Furthermore, there is not any other study where profits can decompose into several factors that are important for the regulator and the regulated companies such as technical change and efficiency change, resource mix, product mix and scale effect using non-frontier and frontier techniques. The next chapters will discuss these issues further, present the sample and data and will show the effectiveness of the price cap scheme for the regulated companies and the consumers in the UK water and sewerage sector and more significantly, the assessment of meaningful comparative performance measurement in a regulated industry where the number of observations is extremely limited. Finally, it will also show the importance of including an exogenous factor like the water and sewerage quality in a profit decomposition analysis since the water and sewerage companies have been obliged to carry substantial capital investment projects in order to improve water and sewerage quality and environmental standards.

CHAPTER 5 INDEX NUMBER BASED PROFIT DECOMPOSITION INCLUDING ONLY SPATIAL PRODUCTIVITY MEASURES

5.1 *Introduction*[1]

The Water and Sewerage companies (WaSCs) in England and Wales were privatized as natural monopolies and thus they had strong incentives for monopoly pricing and weak incentives for operating efficiently. The Office of Water Services (Ofwat), the economic regulator of water and sewerage sector in England and Wales was set up in order to incentivize firms to achieve both productive and allocative efficiency. The method of regulation in UK water and sewerage sector is price cap regulation and is designed to give firms incentives to increase profits by reducing costs by eliminating the potential to manipulate prices and is preferred to rate of return regulation, which potentially leads to overcapitalization (Averch-Johnson, 1962).

In this chapter, we demonstrate that the effectiveness of a price cap scheme; e.g. whether it encourages regulated firms to achieve efficiency in production as well as appropriate allocatively efficient pricing; can be evaluated by determining the relationship between productivity, price performance and profits across firms. Moreover, the underlying index number techniques also allow for comparative performance measurement assessment even in cases where the number of observations is extremely limited. Thus, when compared to alternative methodologies such as DEA and SFA, which require a relatively large number of observations to specify an efficient frontier, index number techniques provide a considerable advantage. Previous studies that illustrated the relationship between profits, productivity and price performance using index number techniques include Water and Street (1998), Han and Hughes (1999), Saal & Parker (2001), Salerian (2003), and Diewert & Lawrence (2006).

Our approach could be naively seen as a minor development of Saal & Parker (2001), which employed a methodology that only allowed the measurement of firm-

---

[1] This chapter hs been published as Maziotis A, Saal D.S. and Thanassoulis E., "Regulatory Price Performance, Excess Cost Indexes and Profitability: How Effective is Price Cap Regulation in the Water Industry?", *Aston Business School Working Papers*, RP 0920 (2009)

specific profitability, productivity and price performance indices over time. Thus, our study instead employs a cross sectional (spatial) index number technique to measure differences in the level of productivity, price performance and profitability across firms (relative comparative performance). However, moving from the consideration of firm-specific indices to a spatial approach firstly allows comparative performance assessment, which makes the approach directly applicable by regulators in setting price caps. Secondly, and more significantly, it also allows the development of the theoretically consistent model of price cap regulation presented in this chapter. As a result, the approach employed in this chapter facilitates an analysis of whether price caps are consistent with the achievement of productive and allocative efficiency, which was simply not possible with the methodology employed by Saal and Parker (2001).

The key theoretical contribution that is allowed by the spatial orientation employed in this chapter is the decomposition of a firm's actual economic profitability into two sources: a spatial multilateral Fisher productivity index (TFP) and a newly developed regulatory total price performance (TPP) index. The former is calculated using theoretically consistent relative productivity comparisons across companies in any given year (multilateral spatial comparisons) after assuming the most productive company is the base or benchmark firm. Moreover, we demonstrate that the inverse of a spatial multilateral TFP index can be interpreted as a regulatory excess costs index, which measures the excess of a firm's actual costs relative to benchmark costs. The regulatory TPP index is derived as a function of this regulatory excess cost index and the actual economic profitability index, and measures the excess of regulated revenues relative to benchmark costs. As such, it provides a direct measure of how tight price caps are, measured by the proportional deviation between allowed revenues and benchmark costs. Further consideration of the theoretical relationship between actual economic profitability, regulatory excess costs and regulatory price performance allows a characterisation of the power of regulatory price caps, and we illustrate this by offering an analysis of changes in the estimated power of price cap regulation in the English and Welsh water industry over the period 1991-2008.

This chapter unfolds as follows. Section 5.2 discusses the potential application of index number techniques for measuring actual profitability, relative productivity and price performance and its adaption under price cap regulation. Section 5.3, then considers the methodology necessary to empirically apply this approach in a multilateral setting. The next section provides a discussion of data employed, and the following section details the empirical results. Section 5.6 then offers some conclusions, as well as suggesting a potential extension of the model.

5.2 Relative Profitability, Productivity and Price Performance

A firm's economic performance is commonly measured by its economic profitability (π). However, changes in profitability can be decomposed into changes in productivity and price performance. Total factor productivity (TFP) captures changes in performance attributable to increased physical production of outputs relative to inputs. In contrast, total price performance (TPP) captures the impact of changes in output prices relative to input prices. Comparing changes in TFP and TPP therefore allows determination of whether profit change is primarily explained by improvements in productivity or is simply attributable to an increase in output prices relative to input prices that has improved the firm's price mark up relative to actual costs.

Saal & Parker (2001) demonstrates an index number approach to decompose a firm's economic profitability change into TFP change and TPP change. For any given firm, this methodology allows identification of the relative contributions of productivity and price performance to observed profit change and the paper illustrates how changes in regulatory policy influenced both the productivity change and price performance change of regulated water and sewerage companies (WaSCs) in England and Wales (E&W). Nevertheless, while this methodological approach has the strong advantage of allowing the decomposition of profit change even if data is only available for a single firm, it only allows comparison of cross firm differences in the rate of change of TFP, TPP and profitability. Thus, the lack of any cross sectional link between firms' indices makes it impossible to measure differences in the level of TFP, TPP and profitability across firms. The implication of this limitation is highlighted if one notes that Saal & Parker (2001) considers an industry subject to price cap regulation in which prices are set using a comparative yardstick regime that measures firm performance levels relative to other regulated firms, but it does not in fact provide a methodology that allows for measurement of such performance differences. This chapter therefore proposes a spatial alternative to Saal & Parker (2001) that allows for measurement of a firm's TFP, TPP and profit performance relative to its peers at any give time.

In this section we first illustrate spatial indices of economic profitability and their decomposition in any given year of our sample and how we can employ spatial indices of productivity and price performance under an ideal incentive regulation regime. After this illustration, Section 5.3 will tackle the thornier issue of applying this concept in an empirical multilateral setting.

5.2.1 Spatial Profitability, Productivity, and Price Performance and its Adaptation to Price Cap Regulation

In this section we consider the relationship between profits, productivity and price performance for firm i relative to a base firm b at time t which we call a spatial index, thereby adopting the terminology employed in the price index literature (Hill, 2004). As a result of its definition, these indices only directly measure differences in performance in the spatial dimension (between firms) at any given time.

We define the economic profits of the base firm b at time t, $\Pi_{b,t}$, as a ratio of its total revenues, $R_{b,t}$ and total costs, $C_{b,t}$, at time t. Thus, the total revenues of the base firm b at period t are defined as $R_{b,t} = P_{b,t} \times Y_{b,t}$, where $P_{b,t}$ and $Y_{b,t}$ respectively present the output price index and the aggregate output index of the base firm b at time t. Its total costs at year t, $C_{b,t}$, are defined as $C_{b,t} = W_{b,t} \times X_{b,t}$, where $W_{b,t}$ and $X_{b,t}$ denotes the input price index and the aggregate input index respectively of the base firm at time t. Similarly, we can define the economic profitability of any firm i at time t, $\Pi_{i,t}$ as a ratio of its total revenues, $R_{i,t}$ and its total costs, $C_{i,t}$. We can thus define and decompose a spatial economic profitability index for any firm i relative to the base firm b at time t, $\pi^S_{b,t}$ as follows:

$$\pi^S_{i,t} = \frac{\Pi_{i,t}}{\Pi_{b,t}} = \frac{\dfrac{R_{i,t}}{C_{i,t}}}{\dfrac{R_{b,t}}{C_{b,t}}} = \frac{\dfrac{P_{i,t}Y_{i,t}}{W_{i,t}X_{i,t}}}{\dfrac{P_{b,t}Y_{b,t}}{W_{b,t}X_{b,t}}} = \frac{TFP_{i,t}}{TFP_{b,t}} \times \frac{TPP_{i,t}}{TPP_{b,t}} = \frac{\dfrac{Y_{i,t}}{Y_{b,t}}}{\dfrac{X_{i,t}}{X_{b,t}}} \times \frac{\dfrac{P_{i,t}}{P_{b,t}}}{\dfrac{W_{i,t}}{W_{b,t}}} = \frac{Y^S_{i,t}}{X^S_{i,t}} \times \frac{P^S_{i,t}}{W^S_{i,t}} = TFP^S_{i,t} \times TPP^S_{i,t} \qquad (5.1)$$

Thus, at time t, a spatial economic profitability index, $\pi_{i,t}^{S}$ can be expressed as a function of an index of spatial total factor productivity for firm *i* relative to the base firm *b*, $TFP_{i,t}^{S}$ and a spatial index of total price performance between firm *i* and the base firm *b*, $TPP_{i,t}^{S}$. As $TFP_{i,t}^{S} = Y_{i,t}^{S} / X_{i,t}^{S}$ and $TPP_{i,t}^{S} = P_{i,t}^{S} / W_{i,t}^{S}$ these indices can be further decomposed as functions of the spatial output ($Y_{i,t}^{S} = Y_{i,t} / Y_{b,t}$), input ($X_{i,t}^{S} = X_{i,t} / X_{b,t}$), output price ($P_{i,t}^{S} = P_{i,t} / P_{b,t}$) and input price ($W_{i,t}^{S} = W_{i,t} / W_{b,t}$) indices. This decomposition of spatial profitability highlights that, at any given time, observed differences in profitability between firms can be explained by differences in productivity, differences in price performance, or differences in both.

By definition spatial indices estimate firm *i*'s performance relative to any potential base firm *b*. However, one of the goals of an ideal incentive regulation system is to incentivize firms to improve their productivity so as to catch up to the productivity levels achieved by the frontier or most productive firm. Therefore, a natural candidate for the base firm is the firm that has the highest relative level of productivity at time t. This is because $TFP_{i,t}^{S}$ then becomes an easily interpretable index of any firm's TFP as a proportion of the best observed productivity level. We therefore henceforward define the base firm (*b*) as the firm with the highest productivity observed at time t.

However, allocative efficiency is also a goal of an ideal incentive regulation regime, as output prices should in principle be just sufficient to cover the efficient economic costs of production. Stated differently, if output prices are allocatively efficient, the most productive regulated companies should achieve a normal rate of return, or equivalently economic profits should be equal to zero. In contrast, less productive firms should in theory have output prices that would allow them to achieve a normal rate of return only if they achieved the productivity levels achieved by the most productive firm. We therefore choose to adapt our definition of spatial TPP so rather than being based relative to the actual output price index of the base firm ($P_{b,t}$), it is instead based to the output prices that are consistent with zero economic profits for the base firm ($P_{b,t}^{*}$).

Mathematically, this is easily accomplished by first noting that if the base (highest productivity) firm at time t makes zero economic profits, then

$\Pi_{b,t} = \frac{R_{b,t}}{C_{b,t}} = \frac{P_{b,t}Y_{b,t}}{W_{b,t}X_{b,t}} = 1$. By simply rearranging this expression, and after assuming that input prices are exogenous, we obtain the following expression for the optimal output price for the best practice firm, which is consistent with the regulatory goal of achieving allocative efficiency:

$$P_{b,t}^{*} = W_{b,t}\frac{X_{b,t}}{Y_{b,t}} \qquad (5.2)$$

$P_{b,t}^{*}$ is an obvious restatement of the condition that for a firm achieving the highest observed productivity, prices should be set such that total revenues are equivalent to economic costs, thereby making economic profits equal to zero. Moreover, it is illustrative to note that in the case of a single output, this is consistent with $P_{b,t}^{*}$ being equal to average long run total costs.

The simple substitution of $P_{b,t}^{*}$ for $P_{b,t}$ in $TPP_{i,t}^{S}$ completes the construction of a "regulatory" TPP index ($TPP_{i,t}^{R}$), which indicates the contribution to profitability that can be attributed to deviation of firm *i*'s output prices from those that would be consistent with achieving the same productivity as the base firm, and also realizing zero economic profits:

$$TPP_{i,t}^{R} = \frac{\dfrac{P_{i,t}}{P_{b,t}^{*}}}{\dfrac{W_{i,t}}{W_{b,t}}} = \left[\frac{\dfrac{P_{i,t}}{W_{i,t}}}{\dfrac{X_{b,t}}{Y_{b,t}}}\right] = \frac{P_{i,t}Y_{b,t}}{W_{i,t}X_{b,t}} \qquad (5.3)$$

As suggested by the term after the first equality, $TPP_{i,t}^{R}$ indicates the deviation of firm *i*'s output price from those that would be appropriate given its input prices and the assumption that it achieved the same productivity as firm *b*. It should also be clear that increases (decreases) in $TPP_{i,t}^{R}$ can be interpreted as loosening (tightening) of regulatory price caps, because this reflects an increase (decrease) in allowed revenues relative to benchmark costs. We can begin to further characterize regulation if we focus on the term after the second equality, if $TPP_{i,t}^{R}=1$, then $P_{i,t}=P_{i,t}^{*}=W_{i,t}X_{b,t}/Y_{b,t}$ and firm *i* will achieve a normal rate of return if it achieved the productivity level of the base firm. If $TPP_{i,t}^{R}<1$ then $P_{i,t}<P_{i,t}^{*}$, thereby suggesting that the regulatory price $P_{i,t}$ has been set "low" as the firm would make an economic loss even if it achieved the productivity levels of the best firm. In contrast, if $TPP_{i,t}^{R}>1$, then $P_{i,t}>P_{i,t}^{*}$, thereby suggesting that the regulatory price $P_{i,t}$ has been set "high" as the firm would make an economic profit if it achieved the productivity levels of the best firm.

Given the regulatory definition of $TPP_{i,t}^{R}$, multiplying it by $TFP_{i,t}^{S}$, no longer results in the spatial measure of profitability ($\pi_{i,t}^{S}$) detailed in (5.1). As (5.4) demonstrates the product of $TPP_{i,t}^{R}$ and $TFP_{i,t}^{S}$ has the advantage that it results in the direct measure of actual firm specific economic profitability $\Pi_{i,t}$ previously defined above:

$$\Pi_{i,t}=\frac{R_{i,t}}{C_{i,t}}=\frac{P_{i,t}Y_{i,t}}{W_{i,t}X_{i,t}}=\left[\frac{\frac{P_{i,t}}{W_{i,t}}}{\frac{X_{b,t}}{Y_{b,t}}}\right]\left[\frac{\frac{Y_{i,t}}{X_{i,t}}}{\frac{Y_{b,t}}{X_{b,t}}}\right]=TPP_{i,t}^{R}TFP_{i,t}^{S} \tag{5.4}$$

Interpretation of (5.4) also demonstrates several useful implications from a regulatory perspective. Focusing on the base firm and given our assumptions, $TFP_{b,t}^{S}=1$ thereby revealing the rather obvious finding that the base firm has achieved the regulatory target of achieving best practice TFP. As a result, for the base firm any economic profits (losses) result in $\Pi_{b,t}>1$ ($\Pi_{b,t}<1$) and imply that $TPP_{b,t}^{R}>1$ ($TPP_{b,t}^{R}<1$). Thus, from a regulatory perspective, for the base firm, economic profits (losses) can only result from "inappropriately" high (low) output prices such that $P_{b,t}>P_{b,t}^{*}$ ($P_{b,t}<P_{b,t}^{*}$). e.g. for the most productive firm, economic profits or losses can only be attributed to regulatory output prices that are not consistent with zero economic profits, and as a result revenues exceed economic costs. We would note that, in a regulatory context where regulators set price caps that include both "catch-up" and "continuing improvement factors", it is more than plausible that a regulator could set $P_{b,t}<P_{b,t}^{*}$ in order to incentive the base firm to further improve its underlying TFP in the future.

For any other firm *i*, if $TPP^{R}_{i,t} = 1$, then $P_{i,t} = P^{*}_{i,t}$ and the regulator has set prices so as to fully incentivize the firm to catch up to the productivity of the base firm. As a result, $\Pi_{i,t} = TFP^{S}_{i,t}$ and the firm's profitability index, will deviate from 1 in exact proportion to its spatial TFP. In this situation the firm will be making an economic loss because of its below par TFP performance. Moreover, it could be argued that such losses are appropriate as the firm's revenues have been set equal to an appropriate benchmark cost associated with the best TFP performance, but the firm's costs exceed these costs by the proportion $1/TFP^{S}_{i,t}$. Therefore, these economic losses are consistent with setting regulated prices so as to incentivize the firm to fully close the productivity gap between it and the base firm.

When $P_{i,t} \neq P^{*}_{i,t}$, output prices, and hence $\Pi_{i,t}$ are not consistent with the spatial productivity benchmark $TFP^{S}_{i,t}$. If, $P_{i,t} < P^{*}_{i,t}$ then $TPP^{R}_{i,t} < 1$ and as a result $\Pi_{i,t} < TFP^{S}_{i,t}$. In this situation regulated revenues are below the benchmark economic costs implied by $TFP^{S}_{i,t}$, and the economic losses of the firm are partially explained by low output prices that imply the firm would need to not only catch up to but also exceed the base firm's productivity level, if it wished to eliminate its economic losses. Therefore, if $TPP^{R}_{i,t} < 1$, then this could suggest evidence of "powerful" price caps, and/or price caps that are designed to stimulate both catch-up and continuing improvements in TFP.

In contrast, if $P_{i,t} > P^{*}_{i,t}$ then $TPP^{R}_{i,t} > 1$ and as a result $\Pi_{i,t} > TFP^{S}_{i,t}$, this is consistent with regulatory prices not having been set to incentivize a firm to fully close its productivity gap with the base firm. This could result for a variety of reasons. One potential reason is the common regulatory practice of setting price caps in a manner that allows required catch up productivity gains to be accomplished over several years rather than immediately. However, this situation could also be taken as evidence of "weak" regulation that does not fully penalize unproductive firms and cause them to suffer economic losses unless they improve their productivity. As UK regulators, for example, have a duty to maintain the financial viability of regulated companies as well as to improve their productivity, this could even be justified by regulators on the grounds that less productive firms would go bust if tough price caps were set and they were unable to sufficiently improve their productivity performance.

As the above paragraph illustrates, there are plausible and potentially appropriate reasons why regulators may choose to set $P_{i,t} > P^*_{i,t}$. Moreover, even if $P_{i,t} > P^*_{i,t}$, the regulator may still have set output prices in a manner that is designed to better incentivize laggard firms to improve their productivity performance, even if they are not required to close the full productivity gap with the base firm. It is therefore worthwhile to carefully define several critical values of $TPP^R_{i,t}$ that can be used to characterise the power of price cap regulation. As discussed above, if $TPP^R_{i,t} = 1$, then a firm could only eliminate its economic losses by fully catching up to the base firm's TFP level. Therefore, if $TPP^R_{i,t} < 1$, then we can characterize price regulation as "powerful" as $P_{i,t}$ is such that a firm could only eliminate its economic losses by catching up to and then exceeding the base firm's TFP.

In contrast, if $TPP^R_{i,t} > 1$, then regulation is somewhat dampened in its effect as economic losses can be eliminated without improving TFP to the level of the base firm. However, there is still a clear distinction between "catch-up promoting" regulatory price caps which still retain some incentives to improve productivity, and "weak" price caps, which allow a laggard firm to potentially earn economic profits regardless of whether its productivity is improved. Thus, if $TPP^R_{i,t} > 1$, but $TPP^R_{i,t} < 1/TFP^S_{i,t}$, then price caps are "catch-up promoting" as they require some, but not full catch up in TFP to eliminate economic losses. In contrast, if $TPP^R_{i,t} > 1/TFP^S_{i,t}$, then price caps are "weak" as prices are high enough for the firm to achieve economic profits despite its low productivity levels, thereby suggesting relatively weak incentives for the firm to improve its productivity

We finally note that it is worthwhile to define a regulatory excess cost index $E^R_{i,t} = 1/TFP^S_{i,t}$, which given the assumption of exogenous input prices, provides an index of the excess of a firm's costs relative to those that would be achieved if it achieved the productivity benchmark. If $TPP^R_{i,t} > 1$, but $TPP^R_{i,t} < E^R_{i,t}$, then the revenues achieved by a firm *i* when its output prices exceed optimal prices are lower than the "excess costs" relative to the benchmark costs resulting in economic losses $(\Pi_{i,t} < 1)$. However, if $TPP^R_{i,t} > 1$, but $TPP^R_{i,t} > E^R_{i,t}$, then the revenues achieved by a firm *i* when its output prices exceed optimal prices are greater than the "excess costs" relative to the benchmark costs resulting in economic profits $(\Pi_{i,t} > 1)$. Thus, it should be clear that if $TPP^R_{i,t} < E^R_{i,t}$, then the regulator has set prices that require laggard firms to improve their productivity/eliminate their excess costs, if they wish to return to economic profitability.

In sum, our discussion highlights that if both productive and allocative efficiency are the goals of price cap regulation, any firm in a regulated industry that has the productivity of the "best practice firm" should in principle have output prices that result in zero economic profits, while less productive firms should have output prices that would result in economic losses unless they improve their productivity. Therefore, the regulator needs to take into account the impact of both TFP and TPP in the overall performance of the companies when setting price caps. Moreover, at a theoretical level, systematic deviation of $TPP^R_{i,t}$ from a value of 1 can be seen as evidence of deviation from the goal of setting prices that are consistent with a strict interpretation of both the incentive and allocative efficiency based justification for price cap regulation. Similarly, if $TPP^R_{i,t} > E^R_{i,t}$, then regulated output prices are high enough to violate a looser interpretation of appropriate incentive regulation, which requires only partial productivity catch up to achieve economic profitability. Given this theoretical discussion, our next section therefore discusses a methodological approach that allows the development of these ideas in an empirical application.

5.3 Multilateral Productivity, Price Performance and Profitability Computations

In this section, we employ a multilateral Fisher index approach to measure profitability, productivity and price performance across companies at any given year (multilateral spatial comparisons). When the price and quantities across different companies are compared, it is important that such comparisons are undertaken for every pair of companies being considered (multilateral comparisons). However, in order to achieve consistency between all the pairs of comparisons we need to derive multilateral indexes that fulfill the property of transitivity. Internal consistency (transitivity) implies that a direct comparison between two firms gives the same result when comparing indirectly these two firms through a third firm.

Bilateral Fisher output and input indexes between two firms i and j in the case of m outputs and n inputs are respectively, $Y^F_{i,j}$ and $X^F_{i,j}$ where:

$$Y^F_{i,j} = [\frac{\sum_{m=1}^{M} P_j^m Y_i^m}{\sum_{m=1}^{M} P_j^m Y_j^m} \times \frac{\sum_{m=1}^{M} P_i^m Y_i^m}{\sum_{m=1}^{M} P_i^m Y_j^m}]^{\frac{1}{2}} \qquad X^F_{i,j} = [\frac{\sum_{n=1}^{N} W_j^n X_i^n}{\sum_{n=1}^{N} W_j^n X_j^n} \times \frac{\sum_{n=1}^{N} W_i^n X_i^n}{\sum_{n=1}^{N} W_i^n X_j^n}]^{\frac{1}{2}} \tag{5.5}$$

Y_i^m and Y_j^m denote the quantities for the *mth* output for firms *i* and *j* respectively, whereas X_i^n and X_j^n present the quantities for the *nth* inputs for firms *i* and *j* respectively. Moreover, P_i^m and P_j^m are prices for the *mth* output, while W_i^n and W_j^n denote input prices. The Fisher output and input indexes measure firm *i*'s output and input as a proportion of firm *j* and are the geometric means of Laspeyers and Paasche output and input indexes. For instance, Laspeyers output and input indexes use company *j*'s prices to weight quantity changes, whereas Paasche output and input indexes use firm *i*'s prices to weight quantity changes. The bilateral Fisher productivity index can then be constructed as a ratio of the Fisher output index relative to the Fisher input index:

$$TFP^F_{i,j} = \frac{Y^F_{i,j}}{X^F_{i,j}} \tag{5.6}$$

The above formula is a binary comparison that can be applied directly when we are only interested in making comparisons between two firms. However, when we are interested in making meaningful comparisons between more than two firms, the multilateral nature of spatial comparisons creates some difficulties, which arise from the fact that more than two firms are compared at the same time. Firstly, the number of comparisons may be quite large depending on the number of companies that we have in our sample so the calculation of productivity index can be quite difficult. Secondly, we need consistent comparisons between all firms such that the relative comparisons between any two firms are consistent with other comparisons (transitivity).

Following standard practice, the process of calculating a transitive Fisher output ($Y_{i,j}^{F}$) and input ($X_{i,j}^{F}$) indices begins by calculating all the possible binary comparisons, $i, j = 1,...,I$ where I is the total number of companies, and results in the following $I \times I$ matrices of binary comparisons:

$$\begin{bmatrix} \mathrm{Y}_{1,1}^{\mathrm{F}} \ Y_{1,2}^{F} \dots\dots Y_{1,I}^{F} \\ Y_{2,1}^{F} \ Y_{2,2}^{F} \dots\dots Y_{2,I}^{F} \\ \dots\dots\dots\dots\dots\dots \\ Y_{I,1}^{F} \ Y_{I,2}^{F} \dots\dots\dots Y_{I,I}^{F} \end{bmatrix} \qquad \begin{bmatrix} \mathrm{X}_{1,1}^{\mathrm{F}} \ X_{1,2}^{F} \dots\dots X_{1,I}^{F} \\ X_{2,1}^{F} \ X_{2,2}^{F} \dots\dots X_{2,I}^{F} \\ \dots\dots\dots\dots\dots\dots \\ X_{I,1}^{F} \ X_{I,2}^{F} \dots\dots\dots X_{I,I}^{F} \end{bmatrix} \tag{5.7}$$

These binary Fisher indices can be converted into multilateral consistent transitive indices by applying the EKS method developed by Elteto-Koves (1964) and Szulc (1964) to derive transitive Fisher indices (see Caves, Christensen and Diewert (1982a), Diewert and Lawrence (2006), Caves et al (1981), Ball et al (2001) for a discussion on multilateral transitive indices). We therefore derive transitive Fisher output and input indices using the EKS method, which is equivalent to taking the geometric mean of the *I* possible direct and indirect (through any possible 3rd firm *k*) binary Fisher comparisons of firms *i* and *j*. The resulting Fisher output and input indices, Y_{ij}^{S} and X_{ij}^{S} therefore fulfill the transitivity property:

$$Y_{ij}^{S} = \prod_{k=1}^{I} \left[Y_{ik}^{F} \times Y_{kj}^{F} \right]^{\frac{1}{I}} \qquad X_{ij}^{S} = \prod_{k=1}^{I} \left[X_{ik}^{F} \times X_{kj}^{F} \right]^{\frac{1}{I}} \tag{5.8}$$

Adopting the terminology of the price index literature (Hill, 2004) we refer to these multilateral output and input indices as spatial indices, as they provide spatially consistent measures across all firms. The matrix in (5.7) represents all multilateral comparisons involving I firms and ideally we would like these comparisons to be internally consistent, i.e., to satisfy the property of transitivity (Coelli et al, 2005).

The binary Fisher index has many desirable properties but it is not suited for multilateral comparisons. The EKS index extends it to the multilateral context since it equals the geometric mean of the I possible direct and indirect (through any possible 3rd firm *k*) binary Fisher comparisons of firms *i* and *j*. Thus, the EKS index is indeed an appropriate multilateral generalisation of the Fisher index (Neary, 2004). The multilateral EKS satisfies the properties of base-company invariance (not sensitive to the choice of base or reference company), transitivity and characteristicity (the comparison between two companies depend only on variables which characterise them and not on variables characteristic of other countries (Neary, 2000). As for characteristicity, the EKS index exhibits this to a high degree by construction, since it is the solution to the problem of finding a transitive index which minimises the sum of squared deviations from the bilateral (and non-transitive) Fisher indexes (Neary, 2004). The multilateral version of the Fisher index is the EKS index, which is widely used in international comparisons, especially by the OECD and it satisfies the form symmetry test and is transitive (Defra, 2003). Fox (2000) and (2004) proved that the EKS index can be derived directly from a flexible transformation function that is non-separable in inputs and outputs and permits non-neutral differences in productivity among countries. The author concluded that "in doing so, it is hoped that it will be clear that the EKS index should no longer be overlooked in favour of the CCD index on the basis of economic justification". (Caves, Christensen and Diewert (1982a)). As we demonstrate below, we apply the EKS procedure in equation (5.8) to yield consistent Fisher output, input and then productivity indices between all firms that satisfy the transitivity property by choosing one firm as a base firm. Therefore each consistent transitive measure, is a measure of any firm i relative to the chosen base firm.

The spatial total factor productivity Fisher index for a firm i relative to firm j, $TFP^S_{i,j}$, can then be constructed as a ratio of the spatial Fisher output index relative to spatial Fisher input index:

$$TFP^S_{ij} = \frac{Y^S_{ij}}{X^S_{ij}} \tag{5.9}$$

However, one can also derive fully equivalent transitive Fisher productivity indices using the EKS method by directly taking the geometric mean of all I possible direct and indirect (through any possible 3rd firm k) binary Fisher productivity comparisons of firms i and j:

$$TFP_{ij}^{S} = \prod_{k=1}^{I} \left[TFP_{ik}^{F} \times TFP_{kj}^{F} \right]^{\frac{1}{I}} \tag{5.10}$$

The resulting index fulfills the transitivity property since it is derived using the EKS method, so any direct comparison between two firms i and j is the same with an indirect comparison between these two firms with a third firm k:

$$TFP_{i,j}^{S} = TFP_{i,k}^{S} \times TFP_{k,j}^{S} \qquad \forall i, j \tag{5.11}$$

While we can generate the $I \times I$ possible transitive spatial output, input and productivity indexes between all firms, transitivity also implies that all meaningful information with regard to relative productivity is available in a subset of only I of these indices. Thus, if we arbitrarily choose one firm as a base firm and set $j = b$, then each spatial measure, is a measure of firm i relative to the chosen base firm and we can also simplify notation such that $TFP_{i,b}^{S} = TFP_{i}^{S}, Y_{i,b}^{S} = Y_{i}^{S}, X_{i,b}^{S} = X_{i}^{S}$. Therefore, productivity relative to the base firm's productivity can be expressed as:

$$TFP_{i}^{S} = \frac{Y_{i}^{S}}{X_{i}^{S}} \tag{5.12}$$

However, this simplification comes at no loss of generality as another spatial productivity measure between any given firms can simply be calculated as $TFP_{i,j}^{S} = TFP_{i}^{S} / TFP_{j}^{S}$. Similarly, $Y_{i,j}^{S} = Y_{i}^{S} / Y_{j}^{S}$ and $X_{i,j}^{S} = X_{i}^{S} / X_{j}^{S}$.

If spatial comparisons are available for each of T time periods indexed by t, and we assume the same base firm in all years, we can define the spatial productivity of firm i relative to firm b at time t as:

$$TFP_{i,t}^{S} = \frac{Y_{i,t}^{S}}{X_{i,t}^{S}} \tag{5.13}$$

These $I \times T$ measures then form the elements of a complete set of spatial comparisons indicating the productivity, output and input of firm i relative to the base firm at time t, and can be succinctly illustrated in the matrices:

$$TFP^S = \begin{bmatrix} TFP^S_{1,1} & \mathrm{TFP}^S_{1,2} \cdots\cdots \mathrm{TFP}^S_{1,T} \\ TFP^S_{2,1} & \mathrm{TFP}^S_{2,2} \cdots\cdots \mathrm{TFP}^S_{2,T} \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ TFP^S_{I,1} & \mathrm{TFP}^S_{I,2} \cdots\cdots \mathrm{TFP}^S_{I,T} \end{bmatrix} \qquad Y^S = \begin{bmatrix} Y^S_{1,1} & \mathrm{Y}^S_{1,2} \cdots\cdots \mathrm{Y}^S_{1,T} \\ Y^S_{2,1} & \mathrm{Y}^S_{2,2} \cdots\cdots \mathrm{Y}^S_{2,T} \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ Y^S_{I,1} & \mathrm{Y}^S_{I,2} \cdots\cdots \mathrm{Y}^S_{I,T} \end{bmatrix}$$

$$X^S = \begin{bmatrix} X^S_{1,1} & \mathrm{X}^S_{1,2} \cdots\cdots \mathrm{X}^S_{1,T} \\ X^S_{2,1} & \mathrm{X}^S_{2,2} \cdots\cdots \mathrm{X}^S_{2,T} \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ X^S_{I,1} & \mathrm{X}^S_{I,2} \cdots\cdots \mathrm{X}^S_{I,T} \end{bmatrix} \tag{5.14}$$

Moreover, we report $I \times T$ measures of a regulatory excess cost index for any firm i at time t as the inverse of the spatial productivity, $E^R_{i,t} = 1/TFP^S_{i,t}$, which given the assumption of exogenous input prices, provides an index of the excess of a firm's costs relative to those that would be achieved if it achieved the productivity benchmark. The set of the $I \times T$ measures of the regulatory excess cost index can be illustrated in the following matrix:

$$E^R = \begin{bmatrix} E^R_{1,1} & \mathrm{E}^R_{1,2} \cdots\cdots\cdots \mathrm{E}^R_{1,T} \\ E^R_{2,1} & \mathrm{E}^R_{2,2} \cdots\cdots\cdots \mathrm{E}^R_{2,T} \\ \cdots\cdots & \cdots\cdots\cdots\cdots\cdots \\ E^R_{I,1} & \mathrm{E}^R_{I,2} \cdots\cdots\cdots \mathrm{E}^R_{I,T} \end{bmatrix} \tag{5.15}$$

We now turn our discussion to the construction of the spatial total price performance index, $\left(TPP^S_{i,t}\right)$ and then the regulatory total price performance index, $\left(TPP^R_{i,t}\right)$. In (5.1) we defined the spatial total price performance of any firm i relative to the base firm as a ratio of output prices to input prices relative to the base firm. Since we defined the spatial TFP index as the productivity index of any firm relative to the best productive firm, we similarly define the spatial TPP index as the price performance index of any firm relative to the price performance of the most productive firm. To accomplish this, we firstly express turnover of a firm *i* relative to the base firm as $R_{i,t}^{S} = R_{i,t}/R_{b,t}$. The spatially consistent aggregate output price index, $(P_{i,t}^{S})$ is then calculated as $P_{i,t}^{S} = R_{i,t}^{S}/Y_{i,t}^{S}$. Similarly, we express nominal economic costs of a firm *i* relative to the base firm as $C_{i,t}^{S} = C_{i,t}/C_{b,t}$. The spatially consistent aggregate input price index, $(W_{i,t}^{S})$ is then calculated as $W_{i,t}^{S} = C_{i,t}^{S}/X_{i,t}^{S}$. Finally, a spatially consistent TPP index of any firm i relative to the base firm at any given time t, $(TPP_{i,t}^{S})$ can be obtained as:

$$TPP_{i,t}^{S} = \frac{\dfrac{R_{i,t}^{S}}{Y_{i,t}^{S}}}{\dfrac{C_{i,t}^{S}}{X_{i,t}^{S}}} = \frac{P_{i,t}^{S}}{W_{i,t}^{S}} \tag{5.16}$$

By rearranging (5.4) an estimate of $TPP_{i,t}^{R}$ can be obtained as a function of firm specific economic profitability, $\Pi_{i,t}$, and the spatially consistent regulatory excess cost index, $E_{i,t}^{R} = 1/TFP_{i,t}^{S}$:

$$TPP_{i,t}^{R} = \Pi_{i,t} E_{i,t}^{R} \tag{5.17}$$

Recall that $TPP_{i,t}^{R}$ measures the proportional deviation of output prices from those that are consistent with zero economic profits if the firm eliminated its excess regulatory costs. As discussed above, if $TPP_{i,t}^{R} < 1$, then we can characterize price regulation as "powerful" in the sense of requiring full catch up to the base firm and further productivity improvement to regain economic profitability. From (5.17) we can see that such "powerful" regulation has the empirically observable requirement that $\Pi_{i,t} < 1/E_{i,t}^{R}$. Similarly, our discussion above revealed that regulatory prices are "catch up promoting" if $TPP_{i,t}^{R} < E_{i,t}^{R}$, which from (5.17) requires that $\Pi_{i,t} < 1$. E.g. firms are required to at least partially eliminate their regulatory excess costs if they wish to regain economic profitability. Thus, it should be clear that the relatively straight forward comparison of $\Pi_{i,t}$, $E_{i,t}^{R}$, and $TPP_{i,t}^{R}$ can provide extremely relevant information with regard to the relative power of regulatory price caps.

Before turning to our empirical application, we must simultaneously highlight both the strength, as well as a potential pitfall, of our index number based methodology. Given that any model of company performance under regulation is only valuable if it can be empirically implemented, our methodology has the distinct advantage of potentially allowing for theoretically consistent cross sectional comparisons of relative productivity, profitability, and price performance, in samples with as few as 2 observations. Thus, when compared to econometric and DEA based approaches to performance measurement, this is a distinct advantage. However, when compared to other methodologies such as SFA and DEA our index number methodology does not allow us to as readily take into account differences in operating characteristics that may affect relative measures of productivity or price performance. Nevertheless, given that profitability is not influenced by these characteristics, and if differences in operating characteristics are relatively small, the methodology should be robust enough to accurately characterize trends in regulatory performance over time. Moreover, as we will see below, even when we control for substantial cross sectional and inter temporal variation in the quality of water and sewerage services in England and Wales, our underlying conclusions with regard to the implied power of regulatory price caps is not affected, even though our estimates of underlying productivity catch up are substantially different.

5.4 Data and the Impact of Quality Adjustment

Our model includes separate outputs for water and sewerage services, and the three inputs, capital, labor and other inputs. The data covered are for the period 1991-2008 for a balanced panel of 10 Water and Sewerage companies (WaSCs). Water connected properties and sewerage connected properties are the proxies for water and sewerage output and are drawn from the companies' regulatory returns to Ofwat. Water and sewage output prices were calculated as the ratio of the appropriate turnover in nominal terms, as available in Ofwat's regulatory returns, to measured output, thereby allowing construction of binary Fisher Output indices. These binary output indices then formed the basis of constructing fully spatially consistent output indices with the EKS method. Finally, spatially consistent aggregate quality-unadjusted output price indices were constructed as the ratio of relative aggregate turnover in nominal terms to this spatial aggregate quality-unadjusted output index, as discussed above.

Our physical capital stock measure is based on the inflation adjusted Modern Equivalent Asset (MEA) estimates of the replacement cost of physical assets contained in the companies' regulatory accounts. However, as periodic revaluations of these replacement cost values could create arbitrary changes in our measure of physical capital, we cannot directly employ these accounting based measures. Instead, we accept the year ending 2006 MEA valuations as our base value, and use net investment in real terms to update this series for earlier and later years. Real net investment is therefore taken as the sum of disposals, additions, investments and depreciation, as deflated by the Construction Output Price Index (COPI). Following Ofwat's approach, we averaged the resulting year ending and year beginning estimates to provide a more accurate estimate of the average physical capital stock available to the companies in a given year.

We subsequently employee a user-cost of capital approach, to calculate total capital costs as the sum of the opportunity cost of invested capital and capital depreciation relative to the MEA asset values, and construct the price of physical capital as the user cost of capital divided by the above MEA based measure of physical capital stocks. The opportunity cost of capital is defined as the product of the weighted average cost of capital (WACC) before tax and the companies' average Regulatory Capital Value (RCV). The RCV is the financial measure of capital stock accepted by Ofwat for regulatory purposes. The WACC calculation is broadly consistent with Ofwat's regulatory assumptions and is estimated with the risk free return assumed to be the average annual yields of medium-term UK inflation indexed gilts. The risk premium for company equity and corporate debt was assumed to be 2% following Ofwat's approach at past price reviews. We also allowed for differences in company gearing ratios and effective corporate tax rates, which were calculated as the sum of aggregate current and deferred tax divided by the aggregate current cost profit before taxation. Finally, following the approach in Ofwat's regulatory current cost accounts, capital depreciation was the sum of current cost depreciation and infrastructure renewals charge.

The average number of full-time equivalent (FTE) employees is available from the companies' statutory accounts. Firm specific labour prices were calculated as the ratio of total labour costs to the average number of full-time equivalent employees. Other costs in nominal terms were defined as the difference between operating costs and total labour costs.[2]Given the absence of data allowing a more refined break out of other costs, we employ the UK price index for materials and fuel purchased in purification and distribution of water, as the price index for other costs, and simply deflate nominal other costs by this measure to obtain a proxy for real usage of other inputs. Given these input quantity and price measures, we are able to calculate the spatially consistent indices of relative input usage discussed above. As total nominal economic costs are obtained as the sum of total capital costs, labour costs and other costs in nominal terms, division of this sum by the spatially consistent input index, allows construction of spatially consistent input price indices. Finally, economic profits are calculated as the difference between turnover and calculated economic costs.

We now have the necessary set of output and input quantity and price measures, as well as the necessary profit, cost, and turnover measures to proceed with out model. However, as differences in operating characteristics may result in legitimate differences in required inputs to produce a given output, variation in measured spatial productivity may result partially from these differences. We would argue that while most such characteristics, such as density of population supplied, or differences in sources of water supply, etc., have an impact on relative performance, these differences are largely stable over time, and will also have only a small impact on explaining differences between the required inputs of WaSCs. In other words, if we are primarily focussed on measuring changes in relative performance over time, then the stability of these differences in characteristics as well as their relatively small impact on input requirements, will not significantly influence trends in relative productivity performance. Nevertheless, as we wish to test the impact of operating characteristics on our model results, and because past research has demonstrated that quality improvements do significantly impact temporal productivity estimates, we also adapt our model to allow for the cross sectional and intertemporal variation in the sewage and drinking water quality.

As is well documented in past studies, the water and sewerage companies have been obliged to carry substantial capital investment projects in order to improve water and sewerage quality and environmental standards. Thus, it is important to measure

---

[2] While it would be particularly desirable to disaggregate other input usage data further, and in particular to allow for separate energy and chemical usage inputs, the data available at company level from Ofwat's regulatory return does not allow a further meaningful decomposition of other input usage.

the impact of quality in our profitability, productivity and price performance measures. We therefore calculated quality-adjusted measures of output for water and sewerage services, as the product of water output and a drinking water quality index and sewerage output and sewage treatment quality index respectively.

Following Saal and Parker (2001) the drinking water quality index is calculated as the ratio of the average percentage of each WaSC's water supply zones that are fully compliant with key water quality parameters, relative to the average compliance percentage for England and Wales in 1991. Water supply zones are areas designated by the water companies by reference to a source of supply in which not more than 50,000 people reside. The data were drawn from the DWI's annual reports for drinking water quality for the years ending 1991-2007[3]. Due to changes in some of the drinking water quality standards and the new regulations, we employed six water quality parameters[4] that are also employed by Ofwat to reflect how well treatment works and distribution systems are operated and maintained (Ofwat, 2006).

The sewage treatment quality index is defined as a weighted index of the percentage of connected population for which sewage receives primary treatment and the percentage of population for which sewage receives at least secondary treatment. It also implicitly includes the percentage of connected population for which sewage is not treated with a zero weight. This data choice reflects both the availability of consistent data capturing quality trends for the entire 1991-2008 period, and does clearly capture substantial increases in sewage treatment levels, particularly in the earlier part of the sample period. The sewage treatment data were taken from *Waterfacts* for the period 1990-91 to 1995-96 and the companies' regulatory returns for the years 1996-97 to 2007-08.

It is clearly necessary to employ a weighted index of these measures as both the quality and costs of higher treatment levels exceed those associated with non treatment or primary treatment alone. We therefore endeavoured to construct a cost based weighting system, although the necessary data to accomplish this was relatively

---

[3] The DWI provides quality data based on calendar years, while all other information employed in this paper is based on fiscal years ending March 31st. We note this inconsistency in the data, but emphasize that the reported years overlap each other for 9 months. Thus, the year end to year end estimates of quality change obtained from the DWI data provide consistent estimates of quality change by the water companies, at a fixed point 9 months into each fiscal year.

[4] The six water quality parameters, which form the Operational Performance Index (OPI) are iron, manganese, aluminium, turbidity, faecal coliforms and trihalomethanes. The resulting drinking water quality index suggests an increase in quality of 10.3 percent between 1991 and 2008 after aggregating the data for all WaSCs.

limited. However, we were able to calculate relative cost measures based on the ratio of sewerage treatment costs to volumes of sewerage treatment, using two alternative cost estimates available from company regulatory returns. One of these alternative estimates was based on total sewerage treatment functional expenditure and direct costs for all treatment works, while the other was based on total sewage treatment costs for large treatment works only. These estimates suggest that higher levels of treatment are 1.68 to 2.40 times more costly than primary treatment only. Given this estimate range, we chose to weight the percentage of population receiving secondary treatment of sewage or more twice as much as the percentage receiving primary treatment only. While admittedly, somewhat ad hoc, we emphasize there is some empirical evidence to support these weights. We note that it is straightforward to demonstrate that the resulting weighted quality index is nested between an index based solely on the percentage of population receiving at least primary sewage treatment, which would underestimate gains in sewage treatment quality, and one based solely on the percentage of population receiving at least secondary sewage treatment, which would overestimate gains in sewage treatment quality.[5]

Once the quality adjusted water and sewerage outputs are constructed, quality adjusted indices are straightforward to produce, by simply repeating the procedures identified above to first produce spatially consistent quality adjusted output indices ($Y_{i,t}^{S,Q}$). A spatial aggregate quality-adjusted aggregated output price index is then constructed as $P_{i,t}^{S,Q} = R_{i,t}^{S} / Y_{i,t}^{S,Q}$. We can also derive a spatial implicit quality index ($Q_{i,t}^{S}$) which measures the implied difference in quality relative to the base firm as $Q_{i,t}^{S} = Y_{i,t}^{S,Q} / Y_{i,t}^{S}$. Therefore, quality adjusted spatial outputs and prices can also be respectively expressed as $Y_{i,t}^{S,Q} = Q_{i,t}^{S} Y_{i,t}^{S}$ and $P_{i,t}^{S,Q} = P_{i,t}^{S} / Q_{i,t}^{S}$, which illustrate that the impact on spatial output quantities will be perfectly balanced by changes in spatial output prices. This also implies that measured economic profitability ($\Pi_{i,t}$) is not influenced by quality adjustment. In contrast, the impact of quality adjustment

---

[5] To highlight this, we note that while our weighted index implies an increase in sewage treatment quality of 19.3% for all England and Wales between 1991 and 2008, an index based only on population receiving at least primary treatment would indicate a quality improvement of 13.7% while one based only on the percentage of population receiving at least secondary treatment of sewage would indicate a 25.4% quality improvement. However, our approach not only provides a mid range estimate between these two more extreme measures, but also better reflects the process of improving sewage treatment quality that occurred through both treating previously untreated sewage, and increasing the level of sewage treatment.

implies that quality adjusted spatial TFP can be expressed as $TFP_{i,t}^{S,Q} = Q_{i,t}^{S} TFP_{i,t}^{S}$, or equivalently that the quality adjusted excess cost index can be expressed as $E_{i,t}^{R,Q} = E_{i,t}^{R} / Q_{i,t}^{S}$. Similarly, quality adjusted regulatory price performance can be expressed as $TPP_{i,t}^{R,Q} = TPP_{i,t}^{R} / Q_{i,t}^{S}$ and economic profitability can be decomposed as:

$$\Pi_{i,t} = TPP_{i,t}^{R,Q} TFP_{i,t}^{S,Q} \tag{5.18}$$

Thus, for example, if we assume that $Q_{i,t}^{S} < 1$, which implies that firm i has lower measured quality than the base firm, then the quality adjusted model will result in $TFP_{i,t}^{S,Q} < TFP_{i,t}^{S}$, but $TPP_{i,t}^{R,Q} > TPP_{i,t}^{R}$. This demonstrates that without quality adjustment, $TFP_{i,t}^{S}$ does not reveal the full extent of the firm's excess costs due to low productivity, while $TPP_{i,t}^{R}$ also results in a perfectly proportional understatement of the excess of allowed revenues to benchmark costs. However, these relationships also suggest that only if $Q_{i,t}^{S}$ deviates significantly from 1, will there be significant differences between the results and policy implications of the quality-adjusted and quality-unadjusted models. Equations (5.4) and (5.18) can be visualised in Figure 5.1. As adjustments for quality affect the spatial productivity and regulatory price performance measures leaving the actual economic profitability unchanged, actual economic profitability can be decomposed into a quality adjusted spatial productivity and quality-adjusted regulatory price performance index. This can be further decomposed as a function of the quality unadjusted spatial productivity index and a spatial implicit quality index and the quality unadjusted regulatory price performance index, and the inverse of a spatial implicit quality index.

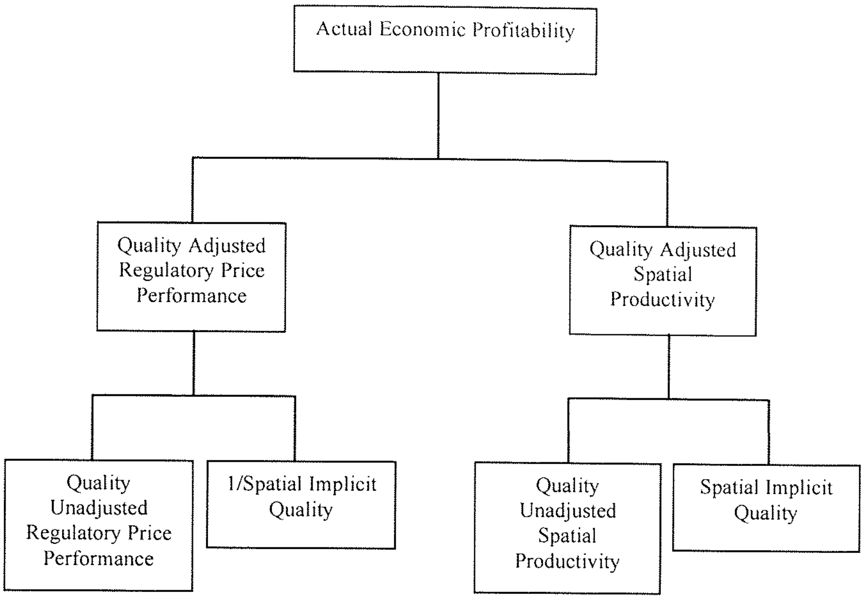


Figure 5-1 Actual Economic Profitability Decomposition

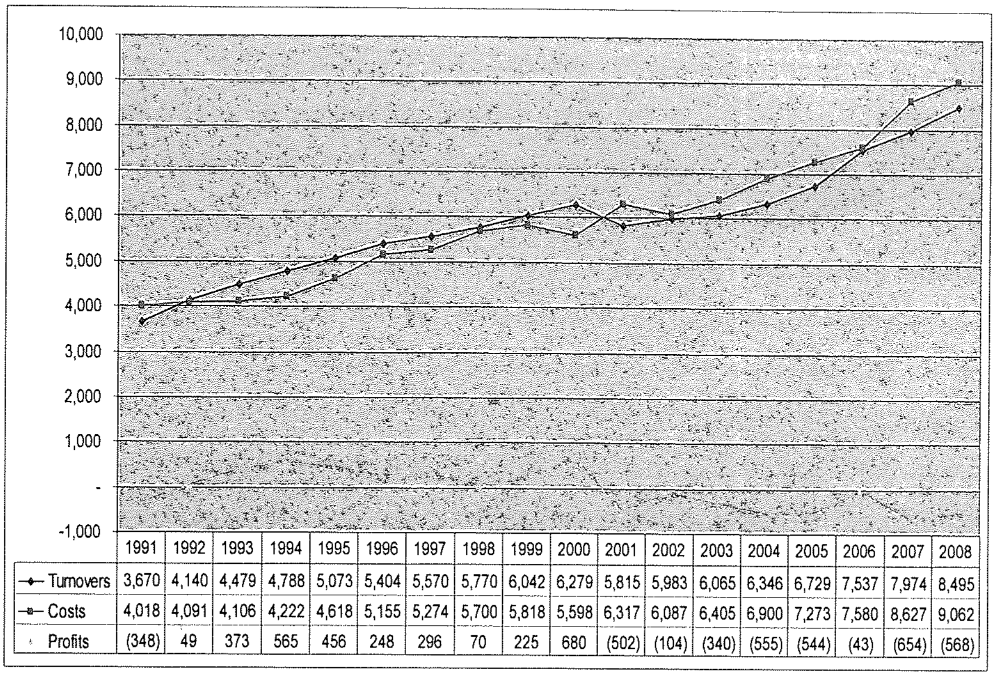
5.5 *Empirical Results*

Before turning to our model results, we first consider trends in aggregate WaSC turnover, costs and profits since privatization, as reported in Figure 5.2. As we should expect, these trends closely follow the regulatory cycle. Thus, a substantial economic loss in 1991 was rapidly eliminated and the industry became increasingly profitable until 1994, when Ofwat not only declared that it would exercise its right to review the relatively lax 10 year price caps set at privatisation after five years, but that it would also effectively rescind the price increases of firms which used their full price cap allowance in 1995 in the new five year price review that would come into effect in 1996. Even though the first price review in 1994/95 tightened regulatory price increases, economic profits remained positive despite falling from 565 to 70 million pounds between 1994 and 1998. Moreover, perhaps reflecting increased incentives to contain costs, there is a noticeable shift in cost trend between 1998 and 2000, and this led the aggregate WaSCS to achieve their highest nominal economic profitability in 2000 with profits of 680 million pounds.

The 1999 price review, which set prices for 2001-2005, marked a shift to considerably tighter regulation by Ofwat. Thus, the 10 year trend of above inflation price increases that had been justified as necessary to fund the industry's capital investment needs, was followed by a substantial reduction in regulatory price caps in

2001, which Ofwat justified as necessary in order to pass cost savings to consumers. This shift in regulation is evidenced by the fall in aggregate turnover from 6,279 to 5,815 million pounds between 2000 and 2001, which is the only example of a nominal decline in aggregate WaSC revenue during the entire 1991-2008 period. When coupled, with the substantial increase in aggregate economic costs in 2001, which can be substantially explained by RPI inflation effects which particularly effect the estimated normal rate of return on invested capital between 1999 and 2002, this resulted in aggregate WaSC losses to 502 million pounds in 2001.

These economics losses do not only reflect a momentary blip in turnover or estimated economic costs in 2001, but rather suggest the institution of a consistently tighter regulatory regime. Thus, despite nominal turnover increasing in every year after 2001, aggregate economic cost increases outstripped allowed revenue increases until 2005 when aggregate economic losses had fallen to 544 million pounds. Moreover, although the implementation of the 2004 price review in 2006 appears to have allowed for a momentarily closer link between regulated revenues and costs, thereby reducing aggregate economic losses to 43 million pounds, subsequent revenue increases have been by far outstripped by increases in economic costs, and by 2008 economic losses again increased to 568 million pounds. Thus, even a straightforward analysis of aggregate WaSC economic profits suggests a shift from a regulatory policy that tolerated above normal returns for the entire period before 2000, to one which set prices resulting in below normal returns after 2000.



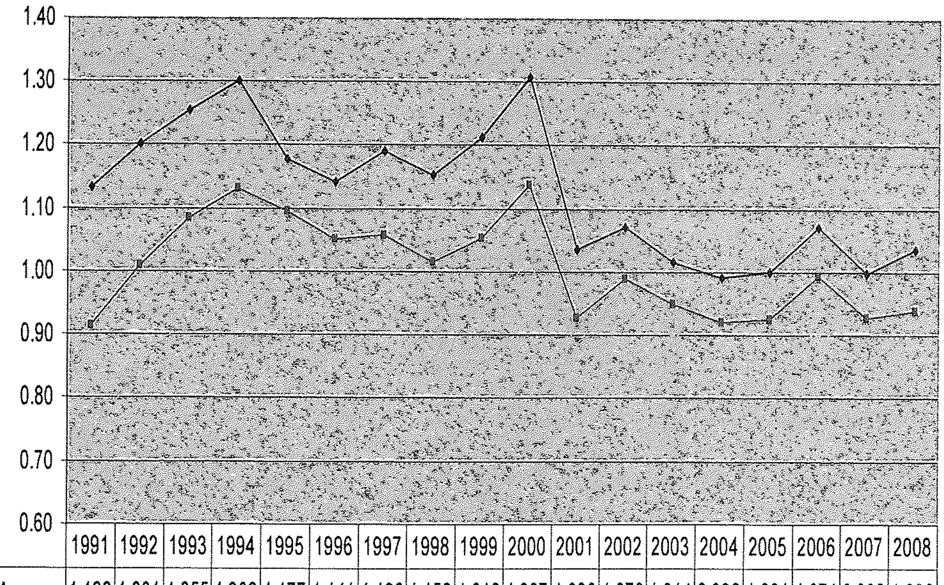
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Turnovers	3,670	4,140	4,479	4,788	5,073	5,404	5,570	5,770	6,042	6,279	5,815	5,983	6,065	6,346	6,729	7,537	7,974	8,495
Costs	4,018	4,091	4,106	4,222	4,618	5,155	5,274	5,700	5,818	5,598	6,317	6,087	6,405	6,900	7,273	7,580	8,627	9,062
Profits	(348)	49	373	565	456	248	296	70	225	680	(502)	(104)	(340)	(555)	(544)	(43)	(654)	(568)

Figure 5-2 Aggregate WaSC Turnover, Costs and Profits: Millions of Pounds

Given these general trends, we begin the presentation of our model results with Figure 5.3, which depicts the geometric average, as well as the range of WaSC profitability over the sample period. We also remind the reader that reported economic profitability reflects actual firm profitability based on firm specific economic costs and revenues, and is in no way influenced by spatial comparisons. These firm specific indices largely confirm the conclusions drawn from Figure 5.2. Thus, profitability peaked in 1994 when the average company made a profit of 12.9%, the most profitable firm made a profit of 30.2%, and the least profitable firm made an economic profit of 0.04%. Regulatory tightening does appear to have shifted this range downward by 1998, but without substantially tightening the range of observed profitability given that the maximum, average, and minimum profitability respectively fell to 1.152, 1.016, and 0.913. However, by 2000 economic profitability reached its highest observed levels as the average company made a profit of 13.7% above the normal rate of return, the highest observed economic profit was 30.7%, and the lowest economic profit was 1.5%. While these high profits in 2000 at least partially reflect the observed reduction in total economic costs between 1998 and 2000, we would emphasize the continued wide range in observed profitability before 2000, as well as the lack of a significant number of firms that made economic losses after 1991.

Moreover, this high profitability is indicative of what in hindsight were inappropriate regulated prices, as prices appear to have had little relationship to the actual economic costs of firms, let alone benchmark economic costs.

Figure 5.3 highlights the dramatic shift in regulatory practice implemented in the 1999 price review. Thus, in 2001, not only did average economic profitability fall to a loss of 7.4%, but the range of observed economic profitability tightened substantially as the highest observed profit was 3.6%, and the most extreme loss was 11.3%. This substantially reduced range in estimated economic profitability, which is sustained in every year after 2000, suggests that Ofwat more closely aligned regulated revenues with actual firm costs after the 1999 price review, and particularly in 2006 which was the first year of the current price review period. Moreover, the consistent economic losses realized by many firms during this period also suggest that Ofwat had begun to deliberately set revenues below actual economic costs, so as to better incentivize firms to reduce their excessive regulatory costs.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Max	1.132	1.201	1.255	1.302	1.177	1.141	1.190	1.152	1.212	1.307	1.036	1.070	1.014	0.992	1.001	1.071	0.998	1.036
■ Average	0.913	1.009	1.085	1.129	1.094	1.051	1.057	1.016	1.052	1.137	0.926	0.990	0.950	0.921	0.924	0.993	0.927	0.938
▲ Min	0.751	0.840	0.938	1.004	0.977	0.944	0.958	0.913	0.944	1.015	0.887	0.966	0.906	0.863	0.861	0.907	0.879	0.877

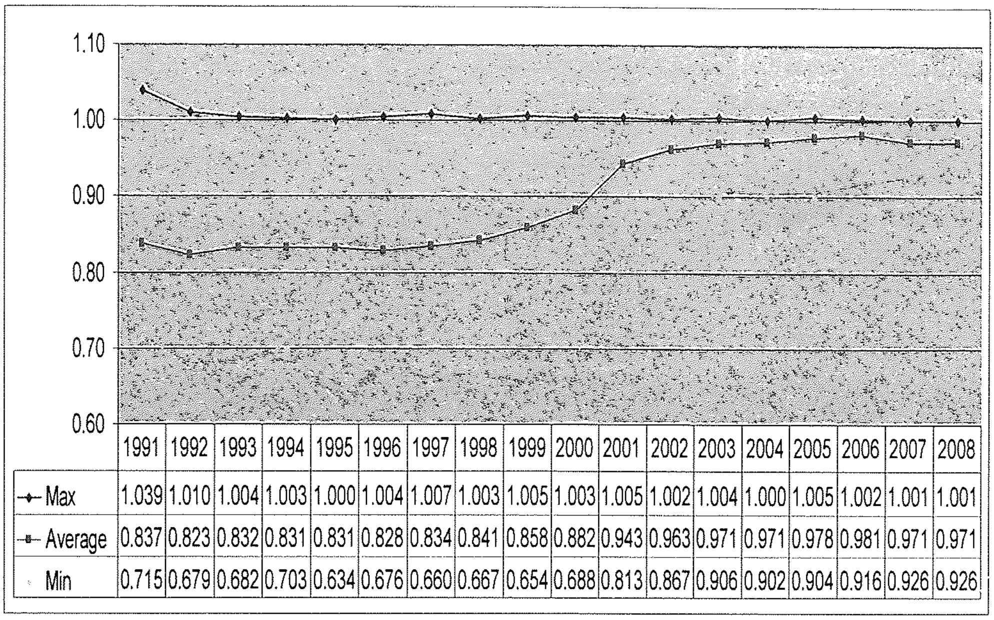
Figure 5-3 Economic Profitability: Firm Specific Estimates

Before considering our spatial estimates of quality unadjusted ($E^{R}_{i,t}$.)and adjusted ($E^{R,Q}_{i,t}$) regulatory excess costs[6], which are respectively reported in Figure 5.5 and Figure 5.6, we first consider the average and range of the spatial implicit quality index ($Q^{S}_{i,t}$) over the sample period so that we can highlight the significant role of quality in our results.[7] Figure 5.4 demonstrates that over the entire sample period, the once wide divergence in the spatial implicit quality index was eliminated, as the companies improved their drinking and sewerage treatment quality conditions relative to the most productive company. Thus, in 1991, the average and worst performing company's implicit quality index were respectively only 83.7% and 71.5% of the base firm's measured quality, while the highest observed quality exceeded the most productive firm's measured quality by 3.9%. In contrast, in 2008, the average and worst company's quality index were respectively 97.1% and 92.6% relative to the most productive company. Moreover, it is worthwhile to note that despite significant investment in drinking water and sewage quality improvement throughout the sample period, little to no convergence in the average and minimum relative quality index occurred before 1998, while most of this convergence occurred between 1998 and 2003. This is likely to reflect what were in fact considerable lags between the provision of revenues necessary to fund quality improving capital investments, and the actual date when the resulting quality improving assets became operational.

---

[6] We have not identified firms for confidentially reasons. The same firm is consistently found to have the highest spatial productivity estimates for both quality unadjusted and quality adjusted models in all years, and is therefore modelled as the benchmark most productive firm in each year of our study Moreover, we note that his same firm was found to have the highest spatial productivity estimates in each year of the study regardless of whether we applied the spatially consistent Fisher indices provided in the main text, similar spatially consistent Tornqvist indices, or the multilateral translog index for WaSCs based on the Tornqvist index developed by Caves et al (1982a). Furthermore, there is little substantive difference between the results regardless of which method is employed.

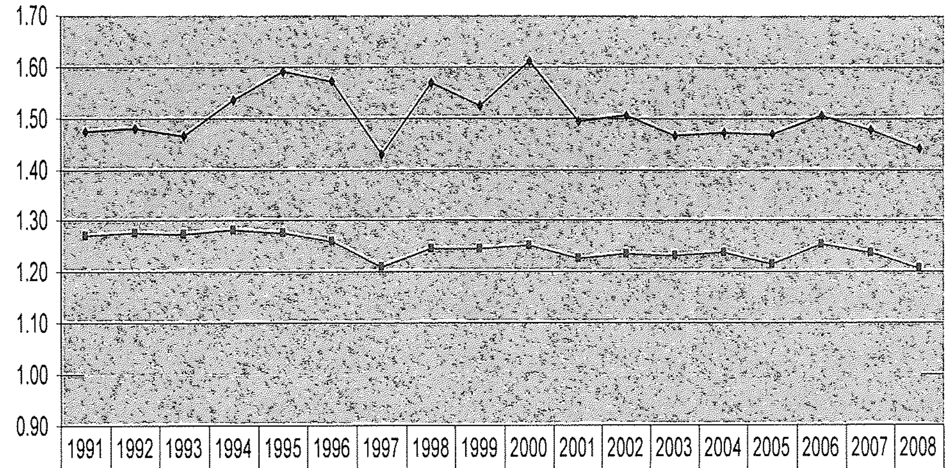
[7] The maximum of the spatial implicit quality index is marginally above 1 in almost yeas of the sample. This reflects the fact that while the base firm is chosen based on its superior quality unadjusted and quality adjusted spatial productivity estimates, its spatial implicit quality index is marginally inferior to at least one firm in the sample.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Max	1.039	1.010	1.004	1.003	1.000	1.004	1.007	1.003	1.005	1.003	1.005	1.002	1.004	1.000	1.005	1.002	1.001	1.001
Average	0.837	0.823	0.832	0.831	0.831	0.828	0.834	0.841	0.858	0.882	0.943	0.963	0.971	0.971	0.978	0.981	0.971	0.971
Min	0.715	0.679	0.682	0.703	0.634	0.676	0.660	0.667	0.654	0.688	0.813	0.867	0.906	0.902	0.904	0.916	0.926	0.926

Figure 5-4 Spatial Implicit Quality Index: Firm Specific Estimates

As, $E^{R,Q}_{i,t} = E^{R}_{i,t} / Q^{S}_{i,t}$, and since $Q^{S}_{i,t} < 1$ for almost all the observations in our sample, the geometric average of quality unadjusted excess costs($E^{R}_{i,t}$) reported in Figure 5.5 is always lower than the geometric average of quality adjusted excess costs ($E^{R,Q}_{i,t}$) reported in Figure 5.6, and the proportional difference is equal to the geometric average of $Q^{S}_{i,t}$ reported in Figure 5.4. Moreover, the strong convergence in $Q^{S}_{i,t}$ documented in Figure 5.4, explains the considerably lower convergence of unadjusted excess costs over the sample period when compared to the convergence of quality adjusted excess costs. Thus, between 1991 and 2008 average quality unadjusted excess costs only declined from 1.27 to 1.205, thereby suggesting that on average productivity catch up by laggard firms contributed only a 5.39% reduction in WaSC costs. In contrast, average quality adjusted excess costs decreased from 1.518 to 1.242, thereby suggesting a much more considerable 18.18% reduction in average costs attributable to productivity catch up by laggard firms. The latter estimate, which is broadly consistent with estimates of cost savings attributable to eliminating efficiency made by Ofwat, demonstrates the empirical necessity of controlling for quality over the 1991-2008 period. However, the lack of significant quality differences after 2003, when the average of $Q^{S}_{i,t}$ always exceeds 0.97, suggest that results in this latter period will not be significantly affected by the quality adjustment method employed in this study. Stated differently, this implies that quality adjustment is necessary if we wish to consider long term trends in the industry, but has very little influence on estimates of excess costs in recent years, which is important if we consider it is precisely these latter estimates that are most relevant for the forthcoming 2009 price review.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Max	1.473	1.480	1.466	1.536	1.592	1.573	1.429	1.569	1.524	1.611	1.495	1.504	1.466	1.473	1.469	1.505	1.477	1.440
Average	1.270	1.277	1.274	1.282	1.277	1.259	1.209	1.247	1.246	1.250	1.226	1.234	1.231	1.238	1.216	1.254	1.238	1.205
Min	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

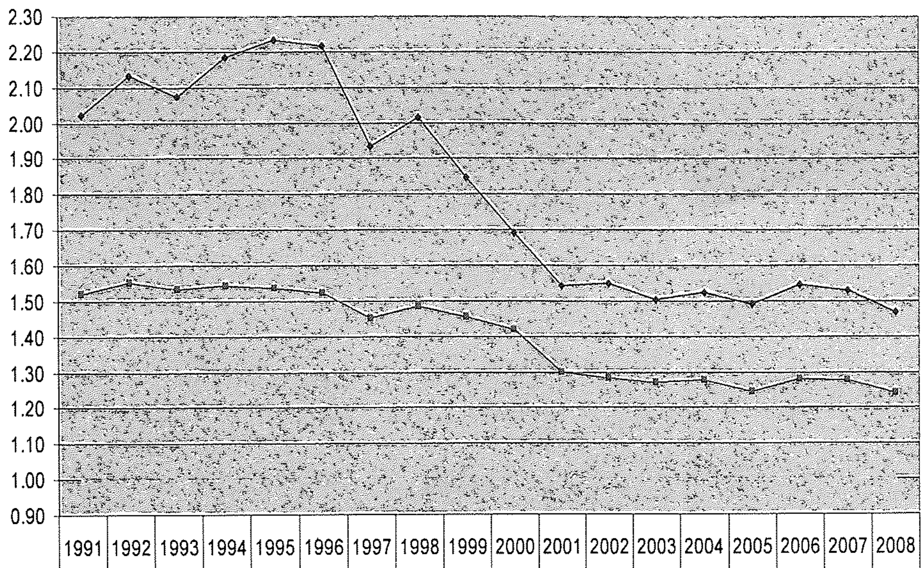
Figure 5-5 Regulatory Excess Costs Quality Unadjusted: Firm Specific Estimates

Given this general discussion of the excess cost estimates, we now discuss their implication when set in context of the regulatory history of the English and Welsh water industry. Despite the wide divergence between estimated quality unadjusted and adjusted excess costs in the early years of the sample, there is nonetheless a remarkable convergence with regard to evidence demonstrating the failure of the regulatory system to eliminate of excess costs during this period. Thus, average quality unadjusted excess costs increased from 27.0% to 27.7% between 1991 and 1995 while average quality adjusted excess costs increased from 51.8% to 53.6%. Even more strikingly, the worst laggard firms saw their quality unadjusted excess costs increase from 47.3% to 59.2% over the same period, while their estimated quality adjusted excess costs increased from 102.1% to 123.3%. Thus, there is clear evidence that during the WaSCs first five years under price cap regulation, little to no improvement in relative productivity/cost performance occurred.

As there is a general consensus that during the 1990s Ofwat's price capping policies was strongest between 1994 and 1997 it is interesting to note that both the quality unadjusted and adjusted excess cost indices fall to a temporary low in 1997, when average quality unadjusted excess costs fell to 20.9% and average quality adjusted excess costs fell to 45.0% of benchmark costs. However, this reduction in excess costs was not sustained in the quality unadjusted model, and its decline was temporarily halted in the quality adjusted model. As a result, if we focus on changes over the formal five year price cap period covering 1996 to 2000, there is a considerable difference in the implications of the unadjusted and quality adjusted results. Thus, the quality unadjusted results reported in Figure 5.5, provide limited evidence of sustained convergence in average excess costs, which fall from 27.7% to 25.0%, while the quality adjusted results suggest a substantial fall in average excess costs from 53.6% to 41.7%. Moreover, while quality unadjusted results suggest no sustained improvement in the excess costs of the worst laggard firm, which actually saw its excess costs increase from 59.2% to 61.1%, the quality adjusted results suggest a considerable improvement in laggard firm performance, given that the maximum quality adjusted excess cost estimate fell from 123.3% to 69.2%. We would suggest that this difference is indicative of the need to control for substantial differences in quality, as well as convergence in quality, if one wishes to properly measure spatial differences in productivity/excess costs. Nevertheless, given that even the quality adjusted excess cost index does not show sustained improvement before 1998, these results may also suggest that the tightening of regulation also acted to reduce the lag between provision of revenues for quality enhancement programmes and their delivery.

Our results do suggest consistent trends with regard to regulatory excess costs for the five year period covered by the 1999 price review, even if the quality unadjusted results, show a much more dampened reduction in excess costs. Thus, between 2000 and 2005, on average the quality unadjusted excess cost index fell from 1.250 to 1.216, while the quality adjusted index fell from 1.417 to 1.243. Both measures also show relatively large average and laggard firm excess cost reductions in 2001, although it must be noted that the magnitude of the quality adjusted excess cost reduction is influenced by the largest observed annual quality increase for both the average and the lowest observed spatial implicit quality index. It is also notable that both indexes also suggest a considerable improvement in regulatory excess costs in 2005. The excess cost results therefore suggest that firms' cost reducing efforts were concentrated in 2001 in a clear response to the large reduction in maximum allowed prices in the first year of the price period, and in the last year of the period, which may suggest they were working to reduce costs to improve their position for the 2006-10 price determination. Moreover, particularly if we focus on the worst laggard firms as represented by the maximum observed excess cost estimates, which declined from 1.611 to 1.469 in the unadjusted model and from 1.692 to 1.490 in the quality adjusted model, there is fairly clear evidence that the tightening of price caps in the 1999 review led to sustained improvements in the relative productivity performance of laggard firms.

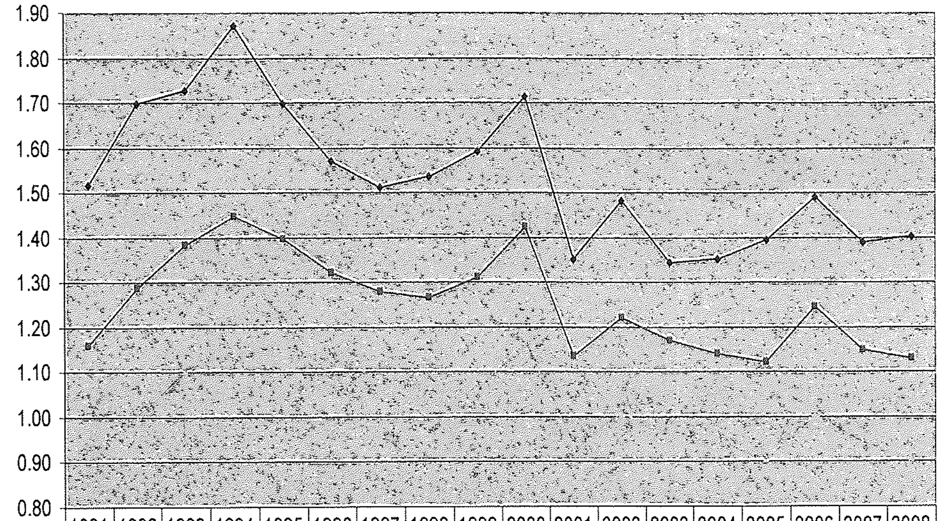
The immediate impact of the 2004 price review, which allowed for an initial increase in operating costs in 2006, has already been observed in Figure 5.2 and Figure 5.3, where there is a clear shift from substantial economic losses in 2005 to a near perfect alignment between revenues and estimated economic costs in 2006. This may or may not be justified on the grounds that Ofwat has a duty to maintain the financial viability of firms in addition to its duty to promote efficiency. Nevertheless, our excess cost results suggest that the realignment of regulated revenues with actual costs led to an immediate increase in excess costs relative to benchmark costs in 2006, an increase which is consistent with the reduction in regulatory incentives to reduce costs in 2006. However, as price caps in years subsequent to 2006 reverted to allowing below inflation price increases, it would appear that the industry not only reverted to economic losses, but also began to improve excess costs relative to the base firm. Nevertheless, while geometric average quality adjusted excess costs fell from 27.8% to 24.2% between 2006 and 2008, as firms again worked to improve productivity relative to benchmark levels, the negative impact of the momentary loosening of price caps in 2006 is demonstrated by the fact that average quality adjusted excess costs in 2008 were only 0.1% lower than they were in 2005.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Max	2.021	2.131	2.073	2.183	2.233	2.217	1.934	2.015	1.846	1.692	1.541	1.549	1.504	1.523	1.490	1.544	1.527	1.468
■ Average	1.518	1.551	1.532	1.542	1.536	1.521	1.450	1.483	1.453	1.417	1.300	1.282	1.269	1.275	1.243	1.278	1.275	1.242
Min	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5-6 Regulatory Excess Costs Quality Adjusted: Firm Specific Estimates

Given this discussion of regulatory excess costs, which measure the excess of actual costs relative to benchmark costs, Figures 5.7 and 5.8 now focus our attention on regulatory TPP, which measures the excess of regulated revenues to benchmark costs, thereby allowing a direct estimate of the tightness of regulatory price caps. From 1991 to 1994, both the unadjusted and quality adjusted results quantify what amounted to a significant loosening in regulatory price caps, as the average excess of regulated revenues to benchmark costs respectively increased from 16% to 44.8% and from 38.7% to 74.2%. In contrast, after 1995 the average values of $TPP_{i,t}^{R}$ and $TPP_{i,t}^{R,Q}$ both suggest a considerable tightening of price caps that persisted until 1998 when they respectively indicate that regulated revenues exceeded benchmark costs by 26.6% and 50.6%. However, during the last two years of the 1996-2000 price cap period, average regulatory TPP again increased, thereby suggesting that price caps had effectively become looser again.

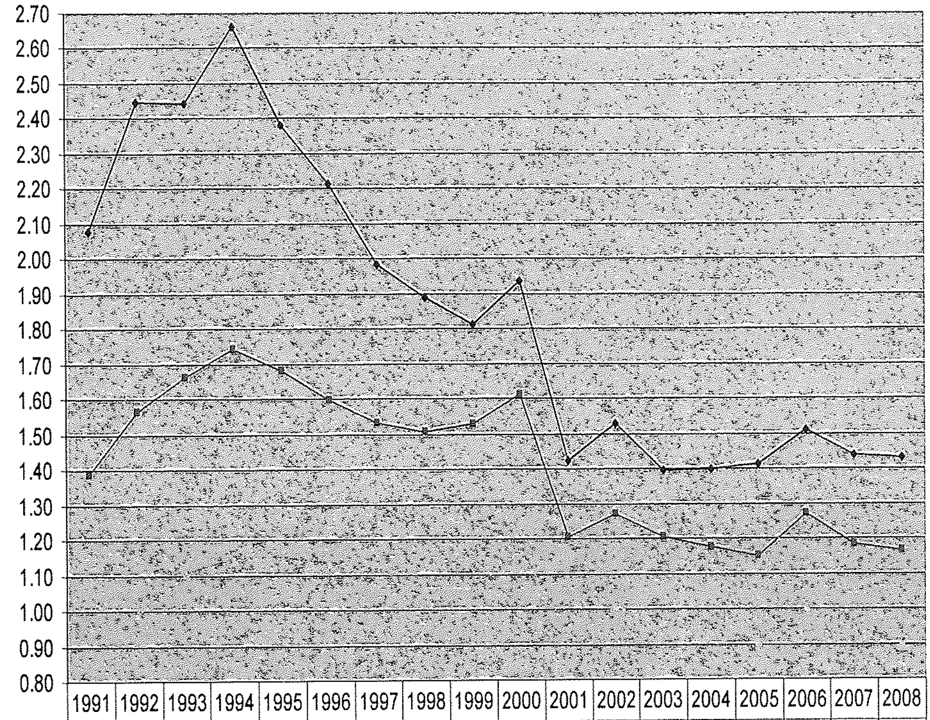


	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Max	1.515	1.698	1.726	1.871	1.697	1.569	1.512	1.535	1.591	1.714	1.352	1.483	1.343	1.350	1.394	1.491	1.389	1.403
■ Average	1.160	1.288	1.382	1.448	1.397	1.323	1.278	1.266	1.312	1.422	1.135	1.222	1.170	1.140	1.123	1.245	1.148	1.131
· Min	0.922	0.993	1.104	1.146	1.149	1.133	1.086	1.063	1.064	1.168	0.938	1.008	0.968	0.938	0.907	1.004	0.942	0.905

Figure 5-7 Regulatory TPP Quality Unadjusted: Firm Specific Estimates

The increased and sustained regulatory demands of the 1999 price review are clearly illustrated in the dramatic fall in estimated regulatory TPP between 2000 and 2001. Thus, in a single year, the average excess of regulated revenues over benchmark costs respectively decreased from 42.2% to 13.5% and from 61.2% to 20.4% for the unadjusted and quality adjusted models. Moreover, the wide dispersion in regulatory TPP, which suggests a more accommodative policy for laggard firms up to 2000 also came to a sudden end, as the range of allowed excess revenues relative to benchmark costs tightened, and in particular, regulatory TPP for the worst performing firms was reduced more than for other firms. If we focus on $TPP_{i,t}^{R,Q}$, as illustrated in Figure 5.8, then this tightening in the range of regulatory rigour is illustrated by the decline of 51.3% in the maximum value of $TPP_{i,t}^{R,Q}$ while the minimum value only declined by 23.0%.

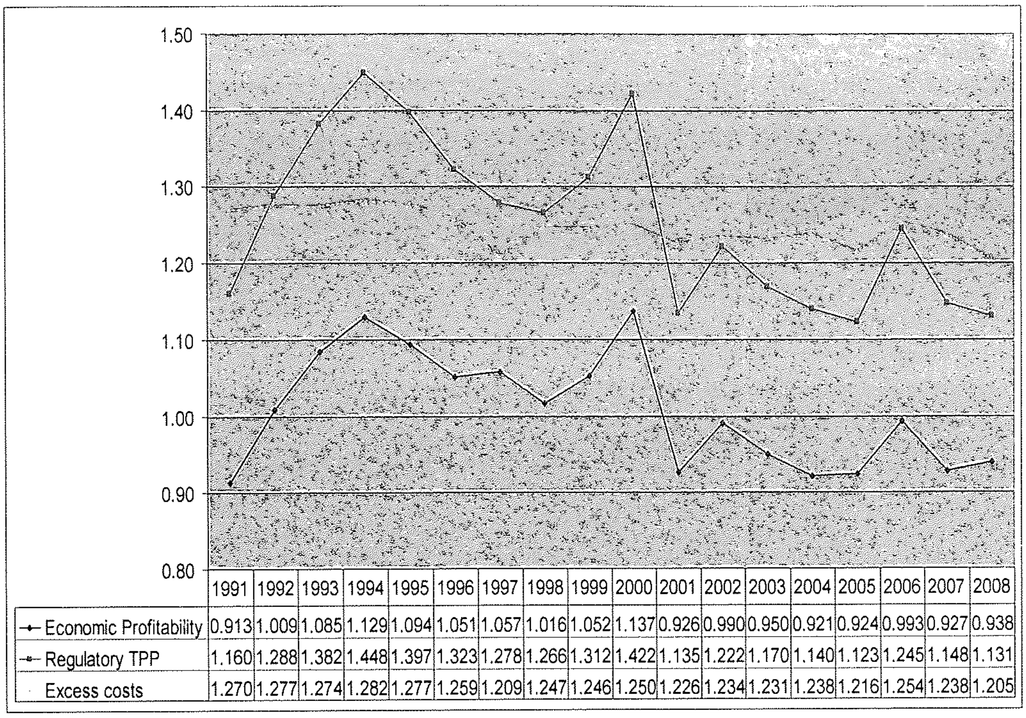
While our regulatory TPP estimates largely suggest continuity in regulatory policy for the post 2000 period, they do provide evidence for a small loosening of price caps in 2006 followed by a return to tougher price caps. Thus between 2005 and 2006 average $TPP_{i,t}^{R}$ and $TPP_{i,t}^{R,Q}$ respectively increased from 1.123 to 1.245 and from 1.149 to 1.269, thereby suggesting an average increase of 12% in the excess of allowed revenues to benchmark costs. However, by 2008 average $TPP^{R}_{i,t}$ and $TPP^{R,Q}_{i,t}$ had respectively fallen to 1.131 and 1.165, thereby demonstrating the return to price caps that were of broadly equivalent tightness to those that had been in place in 2005.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Max	2.079	2.445	2.441	2.661	2.381	2.211	1.982	1.890	1.813	1.934	1.421	1.527	1.393	1.396	1.413	1.511	1.436	1.430
Average	1.387	1.565	1.661	1.742	1.681	1.598	1.533	1.506	1.529	1.612	1.204	1.269	1.205	1.173	1.149	1.269	1.182	1.165
Min	0.956	0.993	1.104	1.146	1.149	1.133	1.114	1.063	1.064	1.168	0.938	1.008	0.968	0.938	0.907	1.004	0.942	0.905

Figure 5-8 Regulatory TPP Quality Adjusted: Firm Specific Estimates

In order to clearly illustrate our underlying model of regulatory price caps, Figures 5.9 and 5.10 respectively report the decomposition of average economic profitability into regulatory TPP and regulatory excess costs indices for the unadjusted and quality adjusted models. As both models suggest the same conclusions with regard to changes in regulatory policy over the sample period, we have chosen to focus on the quality adjusted models in the interest of brevity.

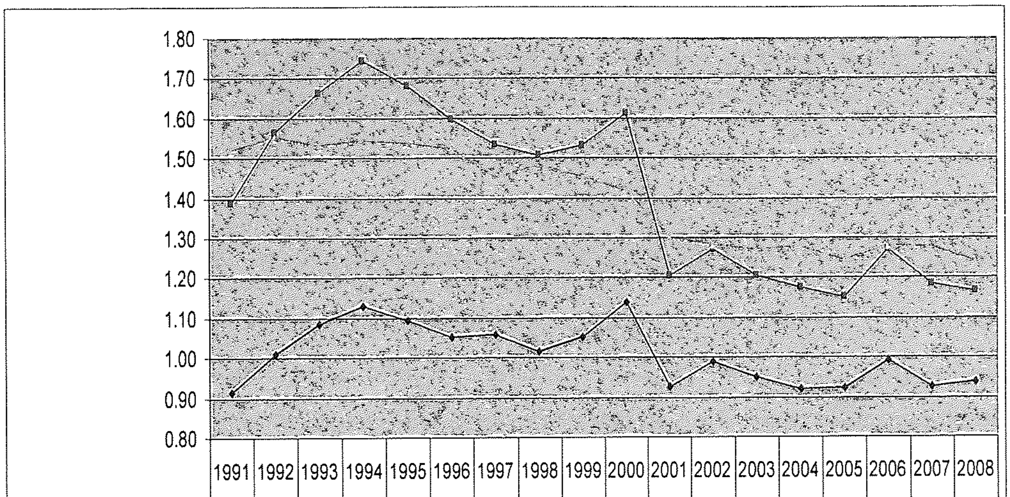


	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Economic Profitability	0.913	1.009	1.085	1.129	1.094	1.051	1.057	1.016	1.052	1.137	0.926	0.990	0.950	0.921	0.924	0.993	0.927	0.938
Regulatory TPP	1.160	1.288	1.382	1.448	1.397	1.323	1.278	1.266	1.312	1.422	1.135	1.222	1.170	1.140	1.123	1.245	1.148	1.131
Excess costs	1.270	1.277	1.274	1.282	1.277	1.259	1.209	1.247	1.246	1.250	1.226	1.234	1.231	1.238	1.216	1.254	1.238	1.205

Figure 5-9 Quality Unadjusted Economic Profitability, Regulatory TPP and Excess Costs: Geometric Average of Firm Specific Estimates

As average $TPP_{i,t}^{R,Q}$ substantially exceeds one in all sample years there is no evidence that the WaSCs have ever been subject to a "powerful" price cap regime requiring immediate full catch up to benchmark costs to regain economic profitability. Moreover, the trend in average $TPP_{i,t}^{R,Q}$ suggests that price caps became progressively looser until 1994 when on average regulated revenues exceeded benchmark costs by 74.2 percent. As the quality adjusted excess cost index ($E_{i,t}^{R,Q}$) suggests that, on average, actual costs only exceeded benchmark costs by 54.2 percent in 1994, the lack of progress is reducing $E_{i,t}^{R,Q}$ before 1995, is fully consistent with the weak incentives created by regulatory price caps that allowed for *increased* economic profitability even in the absence of any effort to improve productivity. Subsequent declines in the average value of $TPP_{i,t}^{R,Q}$ to 1.506 in 1998 demonstrate a substantial reduction in allowed revenues that is indicative of tighter regulation, as firms would at least be required to improve productivity in order to maintain their existing level of economic profitability in the future. However, we emphasize again that even in 1998 average $TPP_{i,t}^{R,Q}$ still exceeded average $E_{i,t}^{R,Q}$, which was 1.483. Moreover, by 2000 average

$TPP_{i,t}^{R,Q}$ had been allowed to increase to 1.612 while at the same time $E_{i,t}^{R,Q}$ declined to 1.417. Thus, despite some improvements in incentives after 1994, 1991-2000 can still be characterized as a period of "weak" regulation because $TPP_{i,t}^{R,Q}$ was generally allowed to exceed $E_{i,t}^{R,Q}$, thereby allowing the retention of above normal returns even in the absence of any effort to achieve benchmark productivity levels.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Economic Profitability	0.913	1.009	1.085	1.129	1.094	1.051	1.057	1.016	1.052	1.137	0.926	0.990	0.950	0.921	0.924	0.993	0.927	0.938
Regulatory TPP	1.387	1.565	1.661	1.742	1.681	1.598	1.533	1.506	1.529	1.612	1.204	1.269	1.205	1.173	1.149	1.269	1.182	1.165
Excess costs	1.518	1.551	1.532	1.542	1.536	1.521	1.450	1.483	1.453	1.417	1.300	1.282	1.269	1.275	1.243	1.278	1.275	1.242

Figure 5-10 Quality Adjusted Economic Profitability, Regulatory TPP and Excess Costs: Geometric Average of Firm Specific Estimates

As our $TPP_{i,t}^{R,Q}$ measure clearly indicates that existing price caps in 2000 were highly inappropriate and allowed firms to achieve average economic profitability of 13.7%, the dramatic tightening of price caps that took place in 2001 was more than justified. Nevertheless, as the average and almost all firm specific estimates of $TPP_{i,t}^{R,Q}$ remained in excess of one even after 2001, there is no evidence that Ofwat moved to a "powerful" price cap regime as defined in Section 5.2. Instead, as clearly illustrated in Figure 5.10, after 2001, Ofwat effectively moved from a policy of setting price caps which allowed regulated revenues to exceed regulatory excess costs ($TPP_{i,t}^{R,Q} > E_{i,t}^{R,Q}$) to a policy of setting regulated revenues below regulatory excess costs ($TPP_{i,t}^{R,Q} < E_{i,t}^{R,Q}$). Therefore, Ofwat's dramatic regulatory tightening in 2001 amounted to a move from "weak" price caps that allowed economic profits even without productivity catch up to "catch up promoting" price caps that required the elimination of at least some excess costs in order to regain economic profitability.

While we would emphasize the overall persistence of this new policy for the entire post 2000 period, the near perfect alignment on average between $TPP_{i,t}^{R,Q}$ (1.269) and $E_{i,t}^{R,Q}$ (1.278) in 2006, followed by resumption of a price caps resulting in $TPP_{i,t}^{R,Q} < E_{i,t}^{R,Q}$ in subsequent years suggests that Ofwat refined its price capping model in the 2004 price review. Specifically, our results suggest that Ofwat's price caps effectively allowed companies their actual economic costs in 2006, but required catch up productivity improvements in subsequent years. This suggests the effective use of a hybrid regulatory model, where companies' costs in the base year were accepted (rate of return), but productivity improvements were expected in later years (benchmarking). However, as the average of $E_{i,t}^{R,Q}$ increased markedly from 1.243 to 1.275 between 2005 and 2006, we would suggest that this hybrid system was inappropriate because the temporary dampening of productivity enhancing incentives led to firms falling further behind the benchmark firm in 2006. Thus, we would argue that a sustained policy of strictly "catch up promoting" price caps may have resulted in regulatory excess costs falling significantly below their 2005 levels by 2008. Instead, Ofwat's apparent temporary dampening of incentives in 2006, would appear to have effectively eliminated 3 years of continued reduction in regulatory excess costs, as illustrated by the average $E_{i,t}^{R,Q}$ of 1.242 in 2008, which was virtually unchanged from its level in 2005.

5.6 Conclusions

This chapter analyzed the impact of regulation on the financial performance of water and sewerage companies in England and Wales over the period 1991-2008. We developed a cross sectional index number technique to decompose actual economic profitability into a spatial productivity and a regulatory price performance index and also measure the spatial implicit impact of quality. The inverse of the spatial productivity index is equivalent to a regulatory excess costs index, which denotes the excess of a firm's actual costs relative to benchmark costs, whereas the regulatory price performance index measures the excess of revenues above benchmark costs. We then demonstrated that increases (decreases) in regulatory price performance are indicative of the loosening (tightening) of price cap regulation. We also showed that the relationship between actual economic profitability, regulatory excess costs and regulatory price performance indices can be used to categorize regulatory price caps as "weak", "powerful" or "catch-up promoting".

The results indicate that throughout the entire 1991-2008 period price caps were never "powerful", in the sense that they required less productive firms to immediately and fully catch-up to the most productive firm to regain economic profitability. As regulatory TPP increased markedly until 1994, we are able to quantify the extent to which price caps became laxer in the early post privatisation period, and how this offered firms the potential to increase their economic profitability without making any effort to reduce their regulatory excess costs. In contrast, between 1994 and 1998, a substantial reduction in regulatory TPP occurred, thereby quantifying the extent of regulatory tightening after the 1994 price review, as falling regulatory TPP implies that laggard firms must reduce their regulatory excess costs, or would otherwise face a reduction in economic profitability. However, our results suggest a renewed increase in regulatory price performance between 1998 and 2000, suggesting that regulatory incentives once again weakened during this period, and economic profitability reached its peak in 2000. In sum, while our results do suggest substantial regulatory tightening after 1994, we would emphasize that the period 1991-2000 can be characterised as a period of "weak" regulation since allowed regulatory revenues almost always exceeded regulatory excess costs, thereby demonstrating that price caps during this period allowed firms to maintain economic profitability regardless of whether they made any progress in catching up to benchmark productivity levels.

Our methodology performs particularly well in demonstrating and quantifying the dramatic tightening of Ofwat's regulatory policies in the 1999 price review. Thus, a sharp tightening in regulation in 2001 is quantified as a substantial fall in the ratio of allowed regulatory revenues relative to benchmark costs, as measured by regulatory TPP. We also clearly demonstrate that Ofwat's dramatic regulatory tightening in 2001 amounted to a move from "weak" price caps that allowed economic profits even without productivity catch up to "catch up promoting" price caps that required the elimination of at least some excess costs in order to regain economic profitability. Furthermore, while our regulatory TPP index clearly demonstrates a momentary but substantial reduction in regulatory incentives in 2006, which was the first year of the current price review, it also demonstrates a return to tighter regulation in subsequent years. Thus, our results suggest that since 2001 Ofwat has implemented "catch up promoting" price caps since average regulated revenues were always below average regulatory excess costs indicating that the firms were required to eliminate at least some excess costs in order to regain economic profitability. We would also emphasize that as our results also clearly demonstrate a much closer alignment between allowed revenues and benchmark costs after 2001, Ofwat's approach during this period was not only appropriate, but should also be continued in the 2009 price review.

We finally emphasize that our methodological approach is generally applicable. This is because it allows regulators to assess relative performance in cases where the number of observations is extremely limited, thereby directly providing firm specific evidence of potential productivity catch up as measured by deviation from benchmark productivity levels, as well as evidence of the deviation of regulated revenues from those that would be consistent with benchmark costs. It also facilitates a backward-looking approach that allows conclusions to be drawn with regard to the effectiveness of price cap regulation. More specifically, using our methodology, regulators and policy makers can determine if past regulatory decisions have not only promoted productive efficiency by providing appropriate efficiency incentives to firms, but also whether they have led to increased allocative efficiency by aligning consumer prices more closely with efficient costs.

CHAPTER 6 INDEX NUMBER BASED APPROACH INCLUDING SPATIAL, TEMPORAL AND OVERALL PRODUCTIVITY MEASURES

6.1 Introduction

In Chapter 5 we developed a cross sectional (spatial) index number technique to allow for the cross-sectional (spatial) measurement of productivity, regulatory price performance, and profitability; and showed the subsequent comparison of how these cross sectional measures have changed over time. In Chapter 6, we measure economic profitability and decompose it into total factor productivity (TFP) and total price performance (TPP), thereby extending a methodological framework originally suggested by Hill (2004) to allow for price indexes that span both multiple firms at a given time (multilateral spatial indexes) and a single firm over multiple periods (temporal indexes). This methodology overcomes the fact that multilateral spatial indexes, which allow consistent comparisons across multiple firms at any given time, are not necessarily consistent with temporal unit-specific indexes, which allow consistent comparison of a given firm across times. Our reconciliation of separate spatial and unit-specific profitability, TFP, and TPP indexes into a single index spanning both firms and time has a significant benefit in application. This is because it allows not only for indexes of unit-specific profitability, TFP, and TPP change, as in Saal & Parker (2001), Water and Street (1998), Han and Hughes (1999) and Salerian (2003), but also allows spatially consistent measurement of changes in these performance measures relative to other firms.

Our methodology is therefore particularly applicable to comparative performance measurement under regulation, where consideration of both temporal and spatial differences in profitability, TFP, and TPP are necessary for setting appropriate regulated prices. Moreover, as alternative methodologies, such as DEA and SFA, require a relatively large number of observations to specify an efficient frontier, our index number based approach has the further potential advantage of allowing meaningful comparative performance measurement even if the number of available observations is extremely limited.

As we demonstrate below, our analysis illustrates several theoretically related methods to measure and decompose financial performance across companies and over time. Firstly, we provide measures of temporal (unit-specific) profitability, productivity and price performance across time for each firm. Secondly, we allow profitability, productivity and price performance comparisons across companies at any given year (multilateral spatial comparisons) calculated by using a multilateral Fisher index. Thirdly, by reconciling together the temporal and spatial profitability, productivity and price performance into relative profitability, productivity and price performance measures, we provide a single index that consistently measures performance change between both firms and over time. Finally, the reconciliation of the spatial, temporal and relative profitability, productivity and price performance measures allows us to decompose the unit-specific index based number profitability growth as a function of the profitability, productivity, price performance growth achieved by benchmark firms, and the catch-up to the benchmark firm achieved by less productive firms. This not only extends the approach of Saal & Parker (2001), Water and Street (1998), Han and Hughes (1999) and Salerian (2003), by allowing a more comprehensive decomposition of a firm's performance changes, but is highly relevant in regulatory and other applications, where comparative performance measurement is appropriate.

Moreover, since the UK water and sewerage industry is characterised by high capital investment programs to improve drinking water quality and environmental standards and past research has demonstrated that quality improvements do significantly impact temporal and spatial productivity and price performance measures (see Saal & Parker, 2001 and Maziotis, Saal and Thanassoulis, 2009), we therefore test the impact of quality on our profitability, productivity and price performance measures. As adjustments for quality affect the productivity and price performance measures leaving the measured economic profitability unchanged, the unit-specific profitability growth can be expressed as a function of the unit-specific quality adjusted productivity and quality-adjusted price performance change. This can be further decomposed as a function of the quality adjusted catch-up in productivity, and the quality adjusted productivity growth of the benchmark firm, and the quality adjusted catch-up in price performance, and the quality adjusted price performance growth of the benchmark firm. The inclusion of quality in our analysis allows us to finally decompose unit-specific economic profitability change as a function of the quality unadjusted catch-up in productivity, the catch-up in quality regarding productivity, and the quality-unadjusted productivity and quality performance over time of the benchmark firm, and the quality unadjusted catch-up in price performance, the catch-up in quality regarding price performance, and the price performance and quality growth of the benchmark firm. We illustrate our analytical decomposition of profit change with an empirical application to the regulated English and Welsh water and sewerage industry during the period 1991-2008.

This chapter unfolds as follows. Section 6.2 discusses the potential application of index number techniques for measuring profitability, productivity and price performance in a binary context. Section 6.3, then considers the methodology necessary to empirically apply this approach in a multilateral setting, whereas section 6.4 discusses the data that were used in this study. The following section provides an application of this methodology followed by a discussion of empirical results. The last section offers some conclusions.

6.2 Profitability, Productivity and Price Performance: A Theoretical Illustration With Bilateral Indices.

A firm's economic performance is commonly measured by its profitability (π). However, changes in profitability can be decomposed into changes in productivity and price performance. Total factor productivity (TFP) captures changes in performance attributable to increased physical production of outputs relative to inputs. In contrast, total price performance (TPP) captures the impact of changes in output prices relative to input prices. Comparing changes in TFP and TPP therefore allows determination of whether profit change is primarily explained by improvements in productivity or is simply attributable to an increase in output prices relative to input prices that has improved the firm's price mark up relative to actual costs.

Saal & Parker (2001) demonstrates an index number approach to decompose a firm's economic profitability change into TFP change and TPP change. For any given firm, this methodology allows identification of the relative contributions of productivity and price performance to observed profit change and the paper illustrates how changes in regulatory policy influenced both the productivity and price performance of regulated water and sewerage companies (WaSCs) in England and Wales (E&W). Nevertheless, while this methodological approach has the strong advantage of allowing the decomposition of profit change even if data is only available for a single firm, it only allows comparison of cross firm differences in the rate of change of TFP, TPP and profitability. Therefore, the lack of any link between firms' indices makes it impossible to measure differences in the level of TFP, TPP and profitability across firms. The implication of this limitation is highlighted if one notes that Saal & Parker (2001) considers an industry subject to price cap regulation in which prices are set using a comparative yardstick regime that measures firm performance levels relative to other regulated firms, but it does not in fact provide a methodology that allows for measurement of such performance differences. This chapter therefore proposes an extension of Saal & Parker (2001) that allows for measurement of a firm's TFP, TPP and profit performance relative to its peers and across time.

Before proceeding, we note that Grifell-Tatje & Lovell (1999) demonstrates a profit decomposition approach, dependent on frontier estimation techniques such as DEA or SFA that decomposes a firm's profitability change while accounting for efficiency catch up relative to the estimated frontier technology. However, while we find no fault with this methodology per se, we note two potential limitations. Firstly, as this approach relies on frontier estimation techniques to obtain measures of relative performance, its application is limited by the requirement of having a sufficient number of degrees of freedom to estimate a meaningful DEA or parametric frontier. In contrast, the empirical index number methodology we propose in Section 6.3 can be applied to decompose profitability growth regardless of the number of inputs and outputs specified, even in cases were the number of observations are extremely limited. Secondly, while the approach of Grifell-Tatje & Lovell (1999) allows for the impact of differences in relative performance on the production side, it has not, to our knowledge, yet been extended to allow for differences between firms in price performance. We feel such distinctions are important, particularly in the regulatory context.

In this section we first illustrate our index number based approach using an example based on bilateral comparisons between two observations. We first illustrate unit specific, spatial and relative indices of economic profitability and their decomposition. We also employ these binary indices to illustrate how unit specific profitability change can be decomposed as a function of the profitability growth of a base firm and profitability catch-up relative to that firm over time. After this illustration, Section 6.3 will tackle the thornier issue of applying these concepts in an empirical multilateral setting.

6.2.1 Unit specific Profitability, Productivity and Price Performance Indices

We first define the unit specific decomposition of profitability following the approach of Saal & Parker (2001) as originally illustrated in Waters & Tretheway (1999). This approach links profits, productivity and price performance between two time periods, year t and the base year 1 for firm i. It therefore only measures differences in the temporal dimension for the given firm.

We define economic profits of firm i at the base year 1, $\Pi_{i,1}$, as a ratio of total revenues, $R_{i,1}$ and total costs in year 1, $C_{i,1}$. Total revenues of a firm i at period 1, $R_{i,1}$, are defined as $R_{i,1} = P_{i,1} \times Y_{i,1}$, where $P_{i,1}$ and $Y_{i,1}$ respectively represent the output price index and the aggregate output index at period 1. Similarly, $C_{i,1} = W_{i,1} \times X_{i,1}$ .We can thus define and decompose a unit-specific (temporal) index of economic profitability for firm i at period t relative to the base period 1, $\pi_{i,t}^{US}$, as follows:

$$\pi_{i,t}^{US} = \frac{\Pi_{i,t}}{\Pi_{i,1}} = \frac{\dfrac{R_{i,t}}{C_{i,t}}}{\dfrac{R_{i,1}}{C_{i,1}}} = \frac{\dfrac{P_{i,t}Y_{i,t}}{W_{i,t}X_{i,t}}}{\dfrac{P_{i,1}Y_{i,1}}{W_{i,1}X_{i,1}}} = \frac{TFP_{i,t}}{TFP_{i,1}} \times \frac{TPP_{i,t}}{TPP_{i,1}} = \frac{\dfrac{Y_{i,t}}{Y_{i,1}}}{\dfrac{X_{i,t}}{X_{i,1}}} \times \frac{\dfrac{P_{i,t}}{P_{i,1}}}{\dfrac{W_{i,t}}{W_{i,1}}} = \frac{Y_{i,t}^{US}}{X_{i,t}^{US}} \times \frac{P_{i,t}^{US}}{W_{i,t}^{US}} = TFP_{i,t}^{US} \times TPP_{i,t}^{US} \qquad (6.1)$$

Thus, the unit-specific economic profitability index, $\pi_{i,t}^{US}$ can be expressed as a function of an index of unit-specific total factor productivity in period t relative to the base year 1, $TFP_{i,t}^{US}$ and an index of unit-specific total price performance between period t and 1, $TPP_{i,t}^{US}$. As $TFP_{i,t}^{US} = Y_{i,t}^{US} / X_{i,t}^{US}$ and $TPP_{i,t}^{US} = P_{i,t}^{US} / W_{i,t}^{US}$ these indices can be further decomposed as functions of the unit-specific output ($Y_{i,t}^{US} = Y_{i,t} / Y_{i,1}$), input ($X_{i,t}^{US} = X_{i,t} / X_{i,1}$), output price ($P_{i,t}^{US} = P_{i,t} / P_{i,1}$) and input price ($W_{i,t}^{US} = W_{i,t} / W_{i,1}$) indices. This decomposition highlights that observed changes in unit-specific profitability over time can be explained by changes in productivity, changes in price performance, or changes in both. Such unit specific measures provide useful information with regard to both changes in unit specific performance as well as its sources.

6.2.2 Spatial Profitability, Productivity, and Price Performance Indices

We next consider the relationship between profits, productivity and price performance for firm i relative to a base firm b at time t, which we call a spatial index, thereby adopting the terminology employed in the price index literature (Hill, 2004). As a result of its definition, these indices only directly measure differences in performance in the spatial dimension (between firms) at any given time.

We define the economic profits of the base firm b at time t, $\Pi_{b,t}$, as a ratio of its total revenues, $R_{b,t}$ and total costs, $C_{b,t}$, at time t. Thus, the total revenues of the base firm b at period t are defined as $R_{b,t} = P_{b,t} \times Y_{b,t}$, where $P_{b,t}$ and $Y_{b,t}$ present the output price index and the aggregate output index respectively of the base firm b at period t. Its total costs at year t, $C_{b,t}$, are defined as $C_{b,t} = W_{b,t} \times X_{b,t}$, where $W_{b,t}$ and $X_{b,t}$ denotes the input price index and the aggregate input index respectively of the base firm at year t. Similarly, we can define economic profits of any firm i at period t, $\Pi_{i,t}$ as a ratio of its total revenues, $R_{i,t}$ and its total costs, $C_{i,t}$. We can thus define and decompose a spatial economic profitability index for any firm i relative to the base firm b at period t, $\pi^S_{b,t}$ as follows:

$$\pi^S_{i,t} = \frac{\Pi_{i,t}}{\Pi_{b,t}} = \frac{\frac{R_{i,t}}{C_{i,t}}}{\frac{R_{b,t}}{C_{b,t}}} = \frac{\frac{P_{i,t}Y_{i,t}}{W_{i,t}X_{i,t}}}{\frac{P_{b,t}Y_{b,t}}{W_{b,t}X_{b,t}}} = \frac{TFP_{i,t}}{TFP_{b,t}} \times \frac{TPP_{i,t}}{TPP_{b,t}} = \frac{\frac{Y_{i,t}}{Y_{b,t}}}{\frac{X_{i,t}}{X_{b,t}}} \times \frac{\frac{P_{i,t}}{P_{b,t}}}{\frac{W_{i,t}}{W_{b,t}}} = \frac{Y^S_{i,t}}{X^S_{i,t}} \times \frac{P^S_{i,t}}{W^S_{i,t}} = TFP^S_{i,t} \times TPP^S_{i,t} \quad (6.2)$$

Thus, at time t, a spatial economic profitability index, $\pi^S_{i,t}$ can be expressed as a function of an index of spatial total factor productivity for firm i relative to the base firm b, $TFP^S_{i,t}$ and a spatial index of total price performance between firm i and the base firm b, $TPP^S_{i,t}$. As $TFP^S_{i,t} = Y^S_{i,t}/X^S_{i,t}$ and $TPP^S_{i,t} = P^S_{i,t}/W^S_{i,t}$ these indices can be further decomposed as functions of the spatial output ($Y^S_{i,t} = Y_{i,t}/Y_{b,t}$), input ($X^S_{i,t} = X_{i,t}/X_{b,t}$), output price ($P^S_{i,t} = P_{i,t}/P_{b,t}$) and input price ($W^S_{i,t} = W_{i,t}/W_{b,t}$) indices. This decomposition of spatial profitability highlights that, at any given time, observed differences in profitability between firms can be explained by differences in productivity, differences in price performance, or differences in both.

By definition spatial indices estimate firm *i*'s performance relative to any potential base firm *b*, and therefore should have potential applications in regulatory settings on this basis alone. However, spatial measures also contain information on relative performance across firms, which unit-specific indices do not. Spatial performance indices can therefore also be employed to measure catch up in relative performance. Thus, if we have access to data for the base year 1 and any other year t, we can define and decompose an index of economic profitability catch up for any firm i at time t and relative to the base firm b at period t, $\pi^C_{i,t}$,as follows:

$$\pi^C_{i,t} = \frac{\pi^S_{i,t}}{\pi^S_{i,1}} = \frac{TFP^S_{i,t}}{TFP^S_{i,1}} \times \frac{TPP^S_{i,t}}{TFP^S_{i,1}} = \frac{\dfrac{Y^S_{i,t}}{Y^S_{i,1}}}{\dfrac{X^S_{i,t}}{X^S_{i,1}}} \times \frac{\dfrac{P^S_{i,t}}{P^S_{i,1}}}{\dfrac{W^S_{i,t}}{W^S_{i,1}}} = \frac{Y^C_{i,t}}{X^C_{i,t}} \times \frac{P^C_{i,t}}{W^C_{i,t}} = TFP^C_{i,t} \times TPP^C_{i,t} \tag{6.3}$$

Thus, for firm i at time t, an index of economic profitability catch up, $\pi^C_{i,t}$ can be expressed as a function of an index of total factor productivity catch up for firm i relative to the base firm *b*, $TFP^C_{i,t}$ and an index of total price performance catch up relative to firm *b*, $TPP^C_{i,t}$. As $TFP^C_{i,t} = Y^C_{i,t} / X^C_{i,t}$ and $TPP^C_{i,t} = P^C_{i,t} / W^C_{i,t}$ these indices can be further decomposed as functions of catch up indices for outputs ($Y^C_{i,t} = Y^S_{i,t} / Y^S_{i,1}$), inputs ($X^C_{i,t} = X^S_{i,t} / X^S_{i,1}$), output prices ($P^C_{i,t} = P^S_{i,t} / P^S_{i,1}$) and input prices ($W^C_{i,t} = W^S_{i,t} / W^S_{i,1}$). This decomposition of profitability catch up highlights that a firm's catch up in profitability can be explained not only by improving its productivity performance relative to the base firm, but also by improving its price performance relative to the base firm. Thus, evidence of improved relative profitability cannot be taken as definitive evidence of improved productivity performance.

6.2.3 Relative Profitability, Productivity and Price Performance Indices

We finally define the relationship between profits, productivity and price performance for any firm i at any time t relative to a base firm b at the base time 1. As by construction these indices are measured relative to a constant base for all t and all i, they therefore capture differences in both the spatial and the temporal dimensions for any given firm at any given time.

As above, we define the economic profits of the base firm b at year 1, $\Pi_{b,1}$, as a ratio of its total revenues, $R_{b,1}$ and total costs, $C_{b,1}$, at year 1. Thus, the total revenues of the base firm b at period 1 are defined as $R_{b,1} = P_{b,1} \times Y_{b,1}$, where $P_{b,1}$ and $Y_{b,1}$ present the output price index and the aggregate output index respectively at period 1. Its total costs at year 1, $C_{b,1}$, are defined as $C_{b,1} = W_{b,1} \times X_{b,1}$, where $W_{b,1}$ and $X_{b,1}$ denotes the input price index and the aggregate input index respectively of the base firm at year 1. We can thus define and decompose a relative index of economic profitability change at time t for firm i relative to the base firm b at time 1, $\pi^R_{i,t}$, as follows:

$$\pi^R_{i,t} = \frac{\Pi_{i,t}}{\Pi_{b,1}} = \frac{\dfrac{R_{i,t}}{C_{i,t}}}{\dfrac{R_{b,1}}{C_{b,1}}} = \frac{\dfrac{P_{i,t}Y_{i,t}}{W_{i,t}X_{i,t}}}{\dfrac{P_{b,1}Y_{b,1}}{W_{b,1}X_{b,1}}} = \frac{TFP_{i,t}}{TFP_{b,1}} \times \frac{TPP_{i,t}}{TPP_{b,1}} = \frac{\dfrac{Y_{i,t}}{Y_{b,1}}}{\dfrac{X_{i,t}}{X_{b,1}}} \times \frac{\dfrac{P_{i,t}}{P_{b,1}}}{\dfrac{W_{i,t}}{W_{b,1}}} = \frac{Y^R_{i,t}}{X^R_{i,t}} \times \frac{P^R_{i,t}}{W^R_{i,t}} = TFP^R_{i,t} \times TPP^R_{i,t} \quad (6.4)$$

Thus, for firm i at time t, the relative economic profitability index, $\pi^R_{i,t}$ can be expressed as a function of an index of relative total factor productivity for firm i at time t relative to the base firm b at time 1, $TFP^R_{i,t}$,and an index of total price performance for firm i at time t relative to the base firm b at time 1,, $TPP^R_{i,t}$. As $TFP^R_{i,t} = Y^R_{i,t} / X^R_{i,t}$ and $TPP^R_{i,t} = P^R_{i,t} / W^R_{i,t}$ these indices can be further decomposed as functions of the relative output ($Y^R_{i,t} = Y_{i,t} / Y_{b,1}$), input ($X^R_{i,t} = X_{i,t} / X_{b,1}$), output price ($P^R_{i,t} = P_{i,t} / P_{b,1}$) and input price ($W^R_{i,t} = W_{i,t} / W_{b,1}$) indices.

Given the binary definition of $\pi^P_{i,t}$ and its components ($TFP^R_{i,t}$, $TPP^R_{i,t}$, $Y^R_{i,t}$, $X^R_{i,t}$, $P^R_{i,t}$ and $W^R_{i,t}$) these relative performance estimates are theoretically equivalent to the separate binary performance estimates provided by the unit-specific and spatial performance measures. Thus, as $\pi^{US}_{i,t} = \pi^R_{i,t} / \pi^R_{i,1}$, $TFP^{US}_{i,t} = TFP^R_{i,t} / TFP^R_{i,1}$, $TPP^{US}_{i,t} = TPP^R_{i,t} / TPP^R_{i,1}$, $Y^{US}_{i,t} = Y^R_{i,t} / Y^R_{i,1}$, $X^{US}_{i,t} = X^R_{i,t} / X^R_{i,1}$, $P^{US}_{i,t} = P^R_{i,t} / P^R_{i,1}$ and

$W_{i,t}^{US} = W_{i,t}^{R} / W_{i,1}^{R}$ it is straightforward to demonstrate that $\pi_{i,t}^{US}$ can be estimated and fully decomposed as a function of relative performance measure estimates.

$$\pi_{i,t}^{US} = \frac{\pi_{i,t}^{R}}{\pi_{i,1}^{R}} = \frac{\dfrac{Y_{i,t}^{R}}{Y_{i,1}^{R}}}{\dfrac{X_{i,t}^{R}}{X_{i,1}^{R}}} \times \frac{\dfrac{P_{i,t}^{R}}{P_{i,1}^{R}}}{\dfrac{W_{i,t}^{R}}{W_{i,1}^{R}}} = \frac{TFP_{i,t}^{R}}{TFP_{i,1}^{R}} \times \frac{TPP_{i,t}^{R}}{TPP_{i,1}^{R}} \tag{6.5}$$

Similarly, as $\pi_{i,t}^{S} = \pi_{i,t}^{R} / \pi_{b,t}^{R}$, $TFP_{i,t}^{S} = TFP_{i,t}^{R} / TFP_{b,t}^{R}$, $TPP_{i,t}^{S} = TPP_{i,t}^{R} / TPP_{b,t}^{R}$, $Y_{i,t}^{S} = Y_{i,t}^{R} / Y_{b,t}^{R}$, $X_{i,t}^{S} = X_{i,t}^{R} / X_{b,t}^{R}$, $P_{i,t}^{S} = P_{i,t}^{R} / P_{b,t}^{R}$ and $W_{i,t}^{S} = W_{i,t}^{R} / W_{b,t}^{R}$:

$$\pi_{i,t}^{S} = \frac{\pi_{i,t}^{R}}{\pi_{b,t}^{R}} = \frac{\dfrac{Y_{i,t}^{R}}{Y_{b,t}^{R}}}{\dfrac{X_{i,t}^{R}}{X_{b,t}^{R}}} \times \frac{\dfrac{P_{i,t}^{R}}{P_{b,t}^{R}}}{\dfrac{W_{i,t}^{R}}{W_{b,t}^{R}}} = \frac{TFP_{i,t}^{R}}{TFP_{b,t}^{R}} \times \frac{TPP_{i,t}^{R}}{TPP_{b,t}^{R}} \tag{6.6}$$

Estimates of $\pi_{i,t}^{C}$ can then be constructed with the underlying relative profitability indices, and can in fact be constructed as the ratio of either unit specific or spatial indices as defined in (6.5) and (6.6). This also clearly demonstrates that the catch up index is, at its core, simply a ratio of unit specific profitability growth rates.

$$\pi_{i,t}^{C} = \frac{\pi_{i,t}^{S}}{\pi_{i,1}^{S}} = \frac{\dfrac{\pi_{i,t}^{R}}{\pi_{b,t}^{R}}}{\dfrac{\pi_{i,1}^{R}}{\pi_{b,1}^{R}}} = \frac{\dfrac{\pi_{i,t}^{R}}{\pi_{i,1}^{R}}}{\dfrac{\pi_{b,t}^{R}}{\pi_{b,1}^{R}}} = \frac{\pi_{i,t}^{US}}{\pi_{b,t}^{US}} \tag{6.7}$$

Moreover, by rearranging (6.7) and decomposing the profitability index we can write:

$$\pi_{i,t}^{US} = \pi_{i,t}^{C} \times \pi_{b,t}^{US} = \left(TFP_{i,t}^{C} \times TFP_{b,t}^{US}\right) \times \left(TPP_{i,t}^{C} \times TPP_{b,t}^{US}\right) \tag{6.8}$$

Thus, given the availability of relative performance indices, the temporal economic profitability of a firm i over time, $\pi_{i,t}^{US}$ can be decomposed as a function of the profitability growth of the base firm b, $\pi_{b,t}^{US}$ and the profitability catch-up of the firm i relative to the base firm between year 1 and t, $\pi_{i,t}^{C}$, e.g. profit performance of any firm can be decomposed into a measure capturing the profit change of a reference firm, and the given firm's performance change relative to that reference firm. If $\pi_{i,t}^{C} > 1$, then firm i improved its economic profitability relative to the base firm over time, whereas $\pi_{i,t}^{C} < 1$ implies that relative profitability of firm i has declined relative to that of the base firm. Moreover, as (6.8) also demonstrates, $\pi_{i,t}^{US}$ can be further decomposed to measure not only the relative contributions of unit specific measures of price performance and productivity to profitability, but also to measure these unit specific changes relative to change in TFP and TPP for the base firm. Thus, for example if $TFP_{i,t}^{C} > 1$, then firm i improved its productivity relative to the base firm from year 1 to t, whereas a value lower than 1 indicates that relative productivity of firm i has declined relative to that of the base firm. Equation (6.8) therefore highlights the strong potential to apply this index based approach to regulatory settings where it is desirable to not only measure firm performance, but also to judge that performance relative to a base firm, normally defined as a "best practice" or "benchmark" firm. The decomposition of the unit specific profitability change in equation (6.8) can be visualized in Figure 6.1. Temporal economic profitability change can be expressed as a function of the profitability growth of the benchmark firm and the profitability catch-up relative to the benchmark firm. Moreover, unit specific economic profitability change can be further decomposition into a unit specific productivity and price performance change. The former can be expressed as a function of a function of the quality unadjusted productivity growth of the benchmark firm and the productivity catch-up relative to the benchmark firm, whereas the latter can be expressed as a function of the quality unadjusted price performance growth of the benchmark growth and the price performance catch-up relative to the benchmark firm. Our next section therefore discusses a methodological approach that allows the actual application of the bilateral concepts detailed above in an empirical multilateral setting.

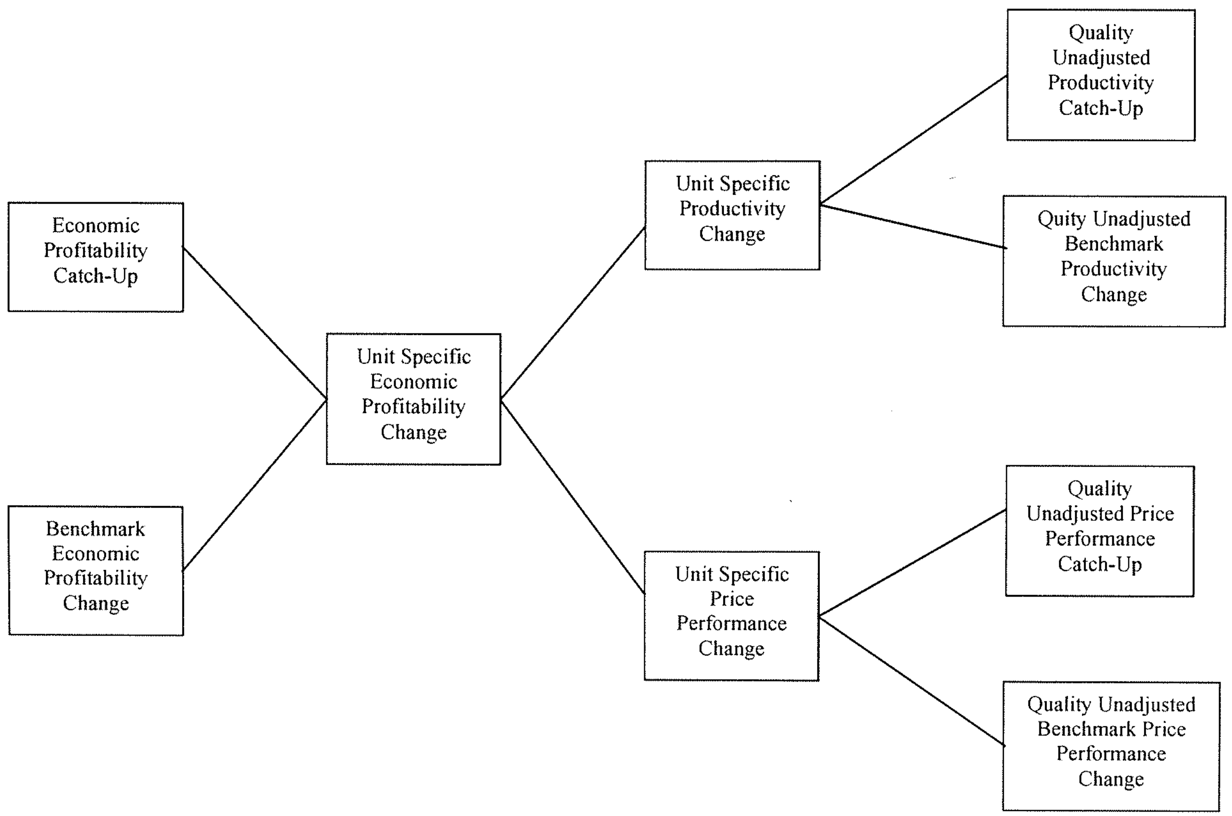


Figure 6-1 Decomposition of Unit Specific Economic Profitability Change

6.3 *Multilateral Productivity, Price Performance and Profitability Computations In Practice*

6.3.1 Chained Unit-specific Productivity, Price Performance and Profitability Over Time

In this section we calculate chained unit-specific profitability, productivity and price performance growth following Saal and Parker's approach (2001). We thus measure these performance measures for any firm between two time periods by using a temporal Fisher index number approach.

Temporal Fisher output and input indexes between two time periods 1 and t, where 1 is the base period in the case of m outputs and n inputs for a firm i are respectively, $Y_{i,t}$ and $X_{i,t}$, :

$$Y_{i,t} = \left[\frac{\sum_{m=1}^{M} P_1^m Y_t^m}{\sum_{m=1}^{M} P_1^m Y_1^m} \times \frac{\sum_{m=1}^{M} P_t^m Y_t^m}{\sum_{m=1}^{M} P_t^m Y_1^m}\right]^{\frac{1}{2}} \qquad X_{i,t} = \left[\frac{\sum_{n=1}^{N} W_1^n X_t^n}{\sum_{n=1}^{N} W_1^n X_1^n} \times \frac{\sum_{n=1}^{N} W_t^n X_t^n}{\sum_{n=1}^{N} W_t^n X_1^n}\right]^{\frac{1}{2}} \qquad (6.9)$$

where Y_t^m and Y_1^m denote the quantities for the *mth* output for periods t and 1 respectively, whereas X_t^n and X_1^n present the quantities for the *nth* inputs for periods t and 1 respectively. Moreover, P_t^m and P_1^m are the prices for *mth* output, while W_t^n and W_1^n denote the input prices. The Fisher output and input indexes of a firm i between two time periods, 1 and t, can also be expressed as the geometric means of Laspeyers and Paasche output and input indexes. A temporal Fisher productivity index, $TFP_{i,t}$ is then constructed as a ratio of Fisher output index relative to Fisher input index, which takes the value 1 in the year 1 (base period):

$$TFP_{i,t} = \frac{Y_{i,t}}{X_{i,t}} \qquad (6.10)$$

A temporal Fisher productivity index can be used in the unchained form denoted above or in a chained form where weights are more closely matched to pair-wise comparisons of observations (Diewert & Lawrence, 2006). The unit-specific output and input indices are thus chained indices, $Y^{CH}_{i,t}$ and $X^{CH}_{i,t}$ between observations 1 and t which are given by:

$$Y^{CH}_{i,t} = 1 \times Y_{i,1,2} \times Y_{i,2,3} \times \ldots \times Y_{i,t-1,t} \qquad X^{CH}_{i,t} = 1 \times X_{i,1,2} \times X_{i,2,3} \times \ldots \times X_{i,t-1,t} \tag{6.11}$$

The unit-specific productivity of a firm i over time can be similarly calculated as a chained index, although it can be equivalently calculated as a ratio of the chained unit-specific output and input indices over time, $Y^{CH}_{i,t}$ and $X^{CH}_{i,t}$:

$$TFP^{CH}_{i,t} = \frac{Y^{CH}_{i,t}}{X^{CH}_{i,t}} \tag{6.12}$$

The set of $I \times T$ unit-specific chained productivity, output and input indices over time can then be summarized in the following matrices:

$$TFP^{CH} = \begin{bmatrix} TFP^{CH}_{1,1} & \mathrm{TFP}^{\mathrm{CH}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{CH}}_{1,\mathrm{T}} \\ TFP^{CH}_{2,1} & \mathrm{TFP}^{\mathrm{CH}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{CH}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ TFP^{CH}_{I,1} & \mathrm{TFP}^{\mathrm{CH}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{CH}}_{\mathrm{I},\mathrm{T}} \end{bmatrix} \qquad Y^{CH} = \begin{bmatrix} Y^{CH}_{1,1} & \mathrm{Y}^{\mathrm{CH}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{CH}}_{1,\mathrm{T}} \\ Y^{CH}_{2,1} & \mathrm{Y}^{\mathrm{CH}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{CH}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ Y^{CH}_{I,1} & \mathrm{Y}^{\mathrm{CH}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{CH}}_{\mathrm{I},\mathrm{T}} \end{bmatrix}$$

$$X^{CH} = \begin{bmatrix} X^{CH}_{1,1} & \mathrm{X}^{\mathrm{CH}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{CH}}_{1,\mathrm{T}} \\ X^{CH}_{2,1} & \mathrm{X}^{\mathrm{CH}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{CH}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ X^{CH}_{I,1} & \mathrm{X}^{\mathrm{CH}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{CH}}_{\mathrm{I},\mathrm{T}} \end{bmatrix} \tag{6.13}$$

Given these chained unit-specific indexes, we can proceed to derive related TPP and Profitability indices as in Saal and Parker (2001). To derive TPP index we firstly express unit-specific turnover at period t relative to the base year 1 as $R^{US}_{i,t} = R_{i,t} / R_{i,1}$. The chained unit-specific aggregate output price index, $\left(P^{CH}_{i,t}\right)$ is then calculated as $P^{CH}_{i,t} = R^{US}_{i,t} / Y^{CH}_{i,t}$. Similarly, we express unit-specific nominal economic costs at period t relative to the base year 1 as $C^{US}_{i,t} = C_{i,t} / C_{i,1}$. The chained unit-specific aggregate input price index, $\left(W_{i,t}^{CH}\right)$ is then calculated as $W_{i,t}^{CH} = C_{i,t}^{US} / X_{i,t}^{CH}$. Finally, a chained unit-specific TPP index for any firm i over time, $\left(TPP_{i,t}^{CH}\right)$ can be obtained as:

$$TPP_{i,t}^{CH} = \frac{\dfrac{R_{i,t}^{US}}{Y_{i,t}^{CH}}}{\dfrac{C_{i,t}^{US}}{X_{i,t}^{CH}}} = \frac{P_{i,t}^{CH}}{W_{i,t}^{CH}} \tag{6.14}$$

Therefore, a chained unit-specific economic profitability index at period t relative to the base year 1 , $\pi_{i,t}^{CH}$ is calculated as the product of a chained index of unit-specific total factor productivity over time, $TFP_{i,t}^{CH}$ and a chained unit-specific index of total price performance over time, $TPP_{i,t}^{CH}$.

6.3.2 Spatial Productivity, Price Performance and Profitability

In the previous section, we used a chained Fisher index to measure profitability, productivity and price performance of any firm between period 1 and period t. In this section, we derive a multilateral Fisher index to measure profitability, productivity and price performance across companies at any given year (multilateral spatial comparisons). When the price and quantities across different companies are compared, it is important that such comparisons are undertaken for every pair of companies being considered (multilateral comparisons). However, in order to achieve consistency between all the pairs of comparisons we need to derive multilateral indexes that fulfill the property of transitivity. Internal consistency (transitivity) implies that a direct comparison between two firms gives the same result when comparing indirectly these two firms through a third firm.

Bilateral Fisher output and input indexes between two firms i and j in the case of m outputs and n inputs are respectively, $Y_{i,j}$ and $X_{i,j}$:

$$Y_{i,j} = [\frac{\sum_{m=1}^{M} P_j^m Y_i^m}{\sum_{m=1}^{M} P_j^m Y_j^m} \times \frac{\sum_{m=1}^{M} P_i^m Y_i^m}{\sum_{m=1}^{M} P_i^m Y_j^m}]^{\frac{1}{2}} \qquad X_{i,j} = [\frac{\sum_{n=1}^{N} W_j^n X_i^n}{\sum_{n=1}^{N} W_j^n X_j^n} \times \frac{\sum_{n=1}^{N} W_i^n X_i^n}{\sum_{n=1}^{N} W_i^n X_j^n}]^{\frac{1}{2}} \qquad (6.15)$$

where Y_i^m and Y_j^m denote the quantities for the *mth* output for firms i and j respectively, whereas X_i^n and X_j^n present the quantities for the *nth* inputs for firms i and j respectively. Moreover, P_i^m and P_j^m are the prices for *mth* output, while W_i^n and W_j^n denote the input prices. The Fisher output and input indexes measure firm *i*'s output and input as a proportion of firm j and are the geometric means of Laspeyers and Paasche output and input indexes. For instance, Laspeyers output and input indexes use company j's prices to weight quantity changes, whereas Paasche output and input indexes use firm *i*'s prices to weight quantity changes. The bilateral Fisher productivity index is then constructed as a ratio of the Fisher output index relative to Fisher input index:

$$TFP_{i,j} = \frac{Y_{i,j}}{X_{i,j}} \qquad (6.16)$$

The above formula is a binary comparison that can be applied directly when we are only interested in making comparisons between two firms. However, when we are interested in making meaningful comparisons between more than two firms, the multilateral nature of spatial comparisons creates some difficulties, which arise from the fact that more than two firms are compared at the same time. Firstly, the number of comparisons may be quite large depending on the number of companies that we have in our sample so the calculation of productivity index can be quite difficult. Secondly, we need consistent comparisons between all firms such that the relative comparisons between any two firms are consistent with other comparisons (transitivity).

Following standard practice, the process of calculating a transitive Fisher output ($Y_{i,j}$) and input ($X_{i,j}$) indices begins by calculating all the possible binary comparisons, $i, j = 1, ..., I$ where I is the total number of companies, and results in the following $I \times I$ matrices of binary comparisons:

$$\begin{bmatrix} \mathrm{Y}_{1,1} \; Y_{1,2} \ldots\ldots Y_{1,I} \\ Y_{2,1} \; Y_{2,2} \ldots\ldots Y_{2,I} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ Y_{I,1} \; Y_{I,2} \ldots\ldots Y_{I,I} \end{bmatrix} \qquad \begin{bmatrix} \mathrm{X}_{1,1} \; X_{1,2} \ldots\ldots X_{1,I} \\ X_{2,1} \; X_{2,2} \ldots\ldots X_{2,I} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ X_{I,1} \; X_{I,2} \ldots\ldots X_{I,I} \end{bmatrix} \tag{6.17}$$

These binary Fisher indices can be converted into multilateral consistent transitive indices by applying the EKS method developed by Elteto-Koves (1964) and Szulc (1964) to derive transitive Fisher indices (see Caves, Christensen and Diewert (1982a), Diewert and Lawrence (2006) and Ball et al (2001) for a discussion on multilateral transitive indices). We therefore derive transitive Fisher output and input indices using the EKS method, which is equivalent to taking the geometric mean of the I possible direct and indirect (through any possible 3rd firm k) binary Fisher comparisons of firms i and j. The resulting Fisher output and input indices, Y_{ij}^{S} and X_{ij}^{S} therefore fulfill the transitivity property:

$$Y_{ij}^{S} = \prod_{k=1}^{I} \left[Y_{ik} \times Y_{kj} \right]^{\frac{1}{I}} \qquad X_{ij}^{S} = \prod_{k=1}^{I} \left[X_{ik} \times X_{kj} \right]^{\frac{1}{I}} \tag{6.18}$$

Adopting the terminology of the price index literature (Hill, 2004) we refer to these multilateral output and input indices as spatial indices, as they provide spatially consistent measures across all firms.

The spatial total factor productivity Fisher index for a firm i relative to firm j, $TFP_{i,j}^{S}$, can then be constructed as the ratio of the spatial Fisher output index relative to spatial Fisher input index:

$$TFP_{ij}^{S} = \frac{Y_{ij}^{S}}{X_{ij}^{S}} \tag{6.19}$$

However, one can also derive fully equivalent transitive Fisher productivity indices using the EKS method by directly taking the geometric mean of all I possible direct and indirect (through any possible 3rd firm k) binary Fisher productivity comparisons of firms i and j:

$$TFP_{ij}^{S} = \prod_{k=1}^{I} \left[TFP_{ik} \times TFP_{kj}\right]^{\frac{1}{I}} \tag{6.20}$$

The resulting index fulfills the transitivity property since it is derived using the EKS method, so any direct comparison between two firms i and j is the same with an indirect comparison between these two firms with a third firm k:

$$TFP_{i,j}^{S} = TFP_{i,k}^{S} \times TFP_{k,j}^{S} \qquad \forall i, j \tag{6.21}$$

While we can generate the $I \times I$ possible transitive spatial output, input and productivity indexes between all firms, transitivity also implies that all meaningful information with regard to relative productivity is available in a subset of only I of these indices. Thus, if we arbitrarily choose one firm as a base firm and set $j = b$, then each spatial measure, is a measure of firm i relative to the chosen base firm and we can also simplify notation such that $TFP_{i,b}^{S} = TFP_{i}^{S}, Y_{i,b}^{S} = Y_{i}^{S}, X_{i,b}^{S} = X_{i}^{S}$. Therefore, productivity relative to the base firm's productivity can be expressed as:

$$TFP_{i}^{S} = \frac{Y_{i}^{S}}{X_{i}^{S}} \tag{6.22}$$

However, this simplification comes at no loss of generality as another spatial productivity measure between any given firms can simply be calculated as $TFP_{i,j}^{S} = TFP_{i}^{S} / TFP_{j}^{S}$. Similarly, $Y_{i,j}^{S} = Y_{i}^{S} / Y_{j}^{S}$ and $X_{i,j}^{S} = X_{i}^{S} / X_{j}^{S}$.

If spatial comparisons are available for each of T time periods indexed by t, and we assume the same base firm in all years, we can define the spatial productivity of firm i relative to firm b at time t as:

$$TFP_{i,t}^{S} = \frac{Y_{i,t}^{S}}{X_{i,t}^{S}} \tag{6.23}$$

These $I \times T$ measures then form the elements of a complete set of spatial comparisons indicating the productivity, output and input of firm i relative to the base firm at time t, and can be succinctly illustrated in the matrices:

$$TFP^{S} = \begin{bmatrix} TFP_{1,1}^{S} \;\; \mathrm{TFP}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{TFP}_{1,\mathrm{T}}^{\mathrm{S}} \\ TFP_{2,1}^{S} \;\; \mathrm{TFP}_{2,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{TFP}_{2,\mathrm{T}}^{\mathrm{S}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ TFP_{I,1}^{S} \;\; \mathrm{TFP}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{TFP}_{1,\mathrm{T}}^{\mathrm{S}} \end{bmatrix} \qquad Y^{S} = \begin{bmatrix} Y_{1,1}^{S} \;\; \mathrm{Y}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{Y}_{1,\mathrm{T}}^{\mathrm{S}} \\ Y_{2,1}^{S} \;\; \mathrm{Y}_{2,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{Y}_{2,\mathrm{T}}^{\mathrm{S}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_{I,1}^{S} \;\; \mathrm{Y}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{Y}_{1,\mathrm{T}}^{\mathrm{S}} \end{bmatrix}$$

$$X^{S} = \begin{bmatrix} X_{1,1}^{S} \;\; \mathrm{X}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{X}_{1,\mathrm{T}}^{\mathrm{S}} \\ X_{2,1}^{S} \;\; \mathrm{X}_{2,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{X}_{2,\mathrm{T}}^{\mathrm{S}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_{I,1}^{S} \;\; \mathrm{X}_{1,2}^{\mathrm{S}} \ldots\ldots\ldots\ldots \mathrm{X}_{1,\mathrm{T}}^{\mathrm{S}} \end{bmatrix} \tag{6.24}$$

We now turn our discussion to the construction of the spatial total price performance index, $\left(TPP_{i,t}^{S}\right)$. Firstly, we express turnover of a firm i relative to the base firm as $R_{i,t}^{S} = R_{i,t} / R_{b,t}$. The spatially consistent aggregate output price index, $\left(P_{i,t}^{S}\right)$ is then calculated as $P_{i,t}^{S} = R_{i,t}^{S} / Y_{i,t}^{S}$. Similarly, we express nominal economic costs of a firm i relative to the base firm as $C_{i,t}^{S} = C_{i,t} / C_{b,t}$. The spatially consistent aggregate input price index, $\left(W_{i,t}^{S}\right)$ is then calculated as $W_{i,t}^{S} = C_{i,t}^{S} / X_{i,t}^{S}$. Finally, a spatially consistent TPP index of any firm i relative to the base firm at any given time t, $\left(TPP_{i,t}^{S}\right)$ can be obtained as:

$$TPP_{i,t}^{S} = \frac{\dfrac{R_{i,t}^{S}}{Y_{i,t}^{S}}}{\dfrac{C_{i,t}^{S}}{X_{i,t}^{S}}} = \frac{P_{i,t}^{S}}{W_{i,t}^{S}} \tag{6.25}$$

Therefore, a spatial economic profitability index at time t, $\pi_{i,t}^{S}$ is calculated as the product of an index of spatial total factor productivity for firm *i* relative to the base firm *b*, $TFP_{i,t}^{S}$ and a spatial index of total price performance between firm *i* and the base firm *b*, $TPP_{i,t}^{S}$. Finally, we also compute matrices of $I \times T$ measures that include the spatial TPP, output and input prices and economic profitability comparisons across companies at any given year.

6.3.3 Relative Productivity, Price Performance and Profitability Change Over Time

In order to simultaneously measure and decompose the profitability growth of any firm in the sample across time and relative to other firms, in practice it is necessary to reconcile the spatial profitability measures defined above with the underlying unit-specific chained profitability of each firm. This is because while section 6.2 has theoretically demonstrated that relative productivity measures can be expressed as a function of unit-specific and spatial productivity measures, this is not as straightforward in a multilateral empirical application. Thus, as demonstrated by Hill (2004) we cannot, in practice, derive multilateral measures of the productive change of any firm *i* relative to the base firm, which can satisfy both spatial and temporal consistency.[8]

We have therefore chosen to pursue measures of relative productivity change over time that guarantee spatial consistency, as this approach is most consistent in the regulatory application we demonstrate below. Thus regulators in comparative or yard stick regulatory regimes typically employ cross section techniques to measure differences in productivity or efficiency across firms (relative comparative performance) and therefore use what are, in fact, spatial performance measures to inform their decision with regard to appropriate regulated prices. Thus, as our applied relative performance measures retain spatial consistency by construction, the relative performance indices will yield comparative performance measures that are consistent with regulatory practice in any given year. However, because our relative measures will also allow intertemporal analysis across firms, they have the advantage of

---

[8] Spatially consistency implies that each year's relative productivity measures do not depend on the other years in the comparison and temporal consistency implies that each firm's productivity estimates do not depend on the number of observations in the time series

allowing a more detailed analysis of firm performance change over time, which is not possible with a spatial index alone. .

Given these arguments, we follow Hill's approach (2004). Therefore, firm i's relative productivity change over time ($TFP^{R}_{i,t}$) is determined as the geometric average of the I alternative potential estimates of relative productivity, as derived by employing the chained time trends and spatial productivities of all the I firms in the sample:

$$TFP^{R}_{i,t} = \left[\prod_{j=1}^{I} \left[(TFP^{CH}_{j,t} \times TFP^{S}_{j,1}) \times \frac{TFP^{S}_{i,t}}{TFP^{S}_{j,t}} \right] \right]^{\frac{1}{I}} \qquad (6.26)$$

Thus, when i = j, $TFP^{R}_{i,t}$ can be simply expressed as the product of the firm's own chained productivity index and its spatial productivity measure in year 1: $TFP^{R}_{i,t} = TFP^{CH}_{i,t} TFP^{S}_{i,1}$. In contrast, for the alternative I-1 estimates when, $i \neq j$. $TFP^{R}_{i,t}$ can also be expressed as a function of any other firm j's relative productivity index calculated as $TFP^{R}_{j,t} = TFP^{CH}_{j,t} TFP^{S}_{j,1}$, and the spatial productivity of firm i relative to firm j, which given the definition of our spatial productivity measures, can be expressed as $\frac{TFP^{S}_{i,t}}{TFP^{S}_{j,t}}$. Thus, rather than relying on a single one of these potential estimates, the definition of $TFP^{R}_{i,t}$ in (6.26) employs all available spatial and chained productivity estimates to provide an arguably superior geometric average estimate of $TFP^{R}_{i,t}$.

We can similarly derive measures of the relative output and input indices over time, $Y^{R}_{i,t}$ and $X^{R}_{i,t}$. The resulting measures of the relative productivity, output and input change of any firm over time are given by the following matrices, where $i = 1,2,...,I$, where I is the total number of firms in the sample and $t = 1,2,...,T$, where T is the total number of years in our panel:

$$TFP^R = \begin{bmatrix} TFP^R_{1,1} & \mathrm{TFP}^{\mathrm{R}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{R}}_{1,\mathrm{T}} \\ TFP^R_{2,1} & \mathrm{TFP}^{\mathrm{R}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{R}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ TFP^R_{I,1} & \mathrm{TFP}^{\mathrm{R}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{TFP}^{\mathrm{R}}_{\mathrm{I},\mathrm{T}} \end{bmatrix} \qquad Y^R = \begin{bmatrix} Y^R_{1,1} & \mathrm{Y}^{\mathrm{R}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{R}}_{1,\mathrm{T}} \\ Y^R_{2,1} & \mathrm{Y}^{\mathrm{R}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{R}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ Y^R_{I,1} & \mathrm{Y}^{\mathrm{R}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{Y}^{\mathrm{R}}_{\mathrm{I},\mathrm{T}} \end{bmatrix}$$

$$X^R = \begin{bmatrix} X^R_{1,1} & \mathrm{X}^{\mathrm{R}}_{1,2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{R}}_{1,\mathrm{T}} \\ X^R_{2,1} & \mathrm{X}^{\mathrm{R}}_{2,2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{R}}_{2,\mathrm{T}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ X^R_{I,1} & \mathrm{X}^{\mathrm{R}}_{\mathrm{I},2} \ldots\ldots\ldots\ldots \mathrm{X}^{\mathrm{R}}_{\mathrm{I},\mathrm{T}} \end{bmatrix} \tag{6.27}$$

Following our approach in (6.4) these relative measures are indices of any firm i measured relative to the base firm in the base year. Construction of consistent price, and TPP indices can therefore be accomplished by firstly expressing turnover of firm i relative to the base firm at the base year 1 as $R^R_{i,t} = R_{i,t}/R_{b,1}$. The relative aggregate output price index over time, $(P^R_{i,t})$ is then calculated as $P^R_{i,t} = R^R_{i,t}/Y^R_{i,t}$. Similarly, we express nominal economic costs of a firm i relative to the base firm at the base year 1 as $C^R_{i,t} = C_{i,t}/C_{b,1}$. The relative aggregate input price index over time, $(W^R_{i,t})$ is then calculated as $W^R_{i,t} = C^R_{i,t}/X^R_{i,t}$. Finally, a relative TPP index of any firm i relative to the base firm at the base year 1, $(TPP^R_{i,t})$ can be obtained as:

$$TPP^R_{i,t} = \frac{\dfrac{R^R_{i,t}}{Y^R_{i,t}}}{\dfrac{C^R_{i,t}}{X^R_{i,t}}} = \frac{P^R_{i,t}}{W^R_{i,t}} \tag{6.28}$$

As a result, a relative economic profitability index, $\pi^R_{i,t}$ can be calculated as the product of an index of relative total factor productivity for firm i relative to the base firm b at base year 1, $TFP^R_{i,t}$ and a relative index of total price performance between firm i and the base firm b at the base year 1, $TPP^R_{i,t}$.

In order to achieve our ultimate goal of decomposing unit specific profit growth in the multilateral context, as demonstrated in (6.8) in the bilateral context, we must finally derive unit specific indices which are consistent with the relative indices developed in (6.26) to (6.28). We therefore calculate a consistent measure of unit-specific productivity over time, which can be obtained as $TFP_{i,t}^{US} = \frac{TFP_{i,t}^{R}}{TFP_{i,1}^{R}}$. Similarly, consistent measures of unit-specific output and input growth are respectively $Y_{i,t}^{US} = \frac{Y_{i,t}^{R}}{\mathrm{Y}_{\mathrm{i,1}}^{\mathrm{R}}}$ and $X_{i,t}^{US} = \frac{X_{i,t}^{R}}{\mathrm{X}_{\mathrm{i,1}}^{\mathrm{R}}}$. In an analogous manner, consistent measures of unit-specific TPP output price, input price and economic profitability indexes are respectively, $TPP_{i,t}^{US} = \frac{TPP_{i,t}^{R}}{TPP_{i,1}^{R}}$, $P_{i,t}^{US} = \frac{P_{i,t}^{R}}{\mathrm{P}_{\mathrm{i,1}}^{\mathrm{R}}}$, $W_{i,t}^{US} = \frac{W_{i,t}^{R}}{\mathrm{W}_{\mathrm{i,1}}^{\mathrm{R}}}$ and $\pi_{i,t}^{US} = TFP_{i,t}^{US} TPP_{i,t}^{US}$.

Given our modelling decision to maintain spatial consistency at the cost of temporal consistency, and the subsequent employment of the geometric average of the I alternative potential relative indicators as appropriate unit specific relative productivity, output and input indices, we must note that the unit-specific chained temporal indexes will, by construction, not be perfectly consistent with the unit specific temporal indexes constructed from the multilateral relative indices. Nevertheless, it can be readily mathematically demonstrated that the geometric average of the I chained unit specific temporal indices and those derived from the relative indices detailed in equations (6.26) to (6.28) are equal. Thus, for example, if we take the geometric average across all firms I in the sample, then $\left[\prod_{i=1}^{I}\left(TFP_{i,t}^{CH}\right)\right]^{\frac{1}{I}} = \left[\prod_{i=1}^{I}\left(TFP_{i,t}^{US}\right)\right]^{\frac{1}{I}}$, and $\left[\prod_{i=1}^{I}\left(TPP_{i,t}^{CH}\right)\right]^{\frac{1}{I}} = \left[\prod_{i=1}^{I}\left(TPP_{i,t}^{US}\right)\right]^{\frac{1}{I}}$. This implies that while our approach to deriving the relative indicators necessary to decompose unit-specific trends in firm performance can result in minor deviations from the temporal trends implied by the unit-specific chained indices, we can nonetheless be fully confident that on average, the unit specific estimates are consistent with the underlying chain-based estimates of temporal change in firm performance. We therefore, focus on these average estimates and their decomposition in our results below.

This section has specified a methodology to allow the empirical application of unit-specific, spatial and relative economic profitability indices and their decomposition into unit-specific, spatial and relative productivity and price performance indices in a multilateral setting. We firstly, calculated chained productivity, price performance and profitability indices for each firm over time. Then, we derived spatial productivity, price performance and profitability indices across firms for each year. Then by reconciling together temporal chained and spatial indices, we were able to derive relative productivity, price performance and profitability comparisons across firms and over time that guarantee spatial consistency. Moreover, we have demonstrated that these estimates are not only spatially consistent, but are also, on average, consistent with alternative unit-specific chained indices of temporal performance change. Thus, this section has demonstrated an appropriate methodology to allow for decompositions of profitability indices in a multilateral setting, thereby extending the approach illustrated in equations (6.1), (6.2) and (6.3) in the binary context. Consequently, we are able to consistently decompose unit specific profitability change as a function of the profitability growth of a base firm and profitability catch-up relative to that firm over time, which can be further decomposed as a function of the productivity and price performance of a base firm and productivity and price performance catch-up relative to that firm over time, in a multilateral setting, as illustrated in equation (6.8) in the binary context. Finally, our index number methodology does not allow us to as readily take into account differences in operating characteristics that may affect relative measures of productivity or price performance. Nevertheless, given that profitability is not influenced by these characteristics, and if differences in operating characteristics are relatively small, the methodology should be robust enough to accurately characterize trends in regulatory performance over time.

6.4 Data and the Impact of Quality Adjustment

Our model includes separate outputs for water and sewerage services, and the three inputs, capital, labor and other inputs. The data covered are for the period 1991-2008 for a balanced panel of 10 Water and Sewerage companies (WaSCs). Water connected properties and sewerage connected properties are the proxies for water and sewerage output and are drawn from the companies' regulatory returns to Ofwat, which are used to construct the output indices. These binary output indices then formed the basis of constructing fully spatially consistent output indices with the EKS method. Finally, spatially consistent aggregate quality-unadjusted output price indices were constructed as the ratio of relative aggregate turnover in nominal terms to this spatial aggregate quality-unadjusted output index, as discussed above.

Our physical capital stock measure is based on the inflation adjusted Modern Equivalent Asset (MEA) estimates of the replacement cost of physical assets contained in the companies' regulatory accounts. However, as periodic revaluations of these replacement cost values could create arbitrary changes in our measure of physical capital, we cannot directly employ these accounting based measures. Instead, we accept the year ending 2006 MEA valuations as our base value, and use net investment in real terms to update this series for earlier and later years. Real net investment is therefore taken as the sum of disposals, additions, investments and depreciation, as deflated by the Construction Output Price Index (COPI). Following Saal and Parker's (2001) approach, we averaged the resulting year ending and year beginning estimates to provide a more accurate estimate of the average physical capital stock available to the companies in a given year.

We subsequently employed a user-cost of capital approach, to calculate total capital costs as the sum of the opportunity cost of invested capital and capital depreciation relative to the MEA asset values, and construct the price of physical capital as the user cost of capital divided by the above MEA based measure of physical capital stocks. The opportunity cost of capital is defined as the product of the weighted average cost of capital (WACC) before tax and the companies' average Regulatory Capital Value (RCV). The RCV is the financial measure of capital stock accepted by Ofwat for regulatory purposes. The WACC calculation is broadly consistent with Ofwat's regulatory assumptions and is estimated with the risk free return assumed to be the average annual yields of medium-term UK inflation indexed gilts. The risk premium for company equity and corporate debt was assumed to be 2% following Ofwat's approach at past price reviews. We also allowed for differences in company gearing ratios and effective corporate tax rates, which were calculated as the sum of aggregate current and deferred tax divided by the aggregate current cost profit before taxation. Finally, following the approach in Ofwat's regulatory current cost accounts, capital depreciation was the sum of current cost depreciation and infrastructure renewals charge.

The average number of full-time equivalent (FTE) employees is available from the companies' statutory accounts. Firm specific labour prices were calculated as the ratio of total labour costs to the average number of full-time equivalent employees. Other costs in nominal terms were defined as the difference between operating costs and total labour costs.[9] Given the absence of data allowing a more refined break out of other costs, we employ the UK price index for materials and fuel purchased in purification and distribution of water, as the price index for other costs, and simply deflate nominal other costs by this measure to obtain a proxy for real usage of other inputs. Given these input quantity and price measures, we are able to calculate indices of unit-specific, spatial and relative input usage discussed above. As total nominal economic costs are obtained as the sum of total capital costs, labour costs and other costs in nominal terms, division of this sum by the unit-specific, spatial and relative input index, allows the construction of unit-specific, spatial and relative input price indices. Finally, economic profits are calculated as the difference between turnover and calculated economic costs.

We now have the necessary set of output and input quantity and price measures, as well as the necessary profit, cost, and turnover measures to proceed with our model. As is well documented in past studies (see Saal & Parker 2000, 2001, Saal, Parker and Weyman-Jones, 2007, Maziotis, Saal and Thanassoulis 2009), the English and Welsh water and sewerage companies have been obliged to carry substantial capital investment projects in order to improve water and sewerage quality and environmental standards. Saal and Parker (2001) and Maziotis, Saal and Thanassoulis (2009) demonstrated that quality improvements do significantly impact temporal and spatial productivity and price performance estimates. Thus, we feel it is important to measure the impact of quality in our unit-specific, spatial and relative profitability, productivity and price performance measures, thereby allowing for the cross sectional and intertemporal variation in the sewage and drinking water quality. We therefore calculated quality-adjusted measures of output for water and sewerage services, as the product of water output and a drinking water quality index and sewerage output and a sewage treatment quality index, respectively.

---

[9] While it would be particularly desirable to disaggregate other input usage data further and in particular to allow for separate energy and chemical usage inputs, the data available at company level from Ofwat's regulatory return does not allow a further meaningful decomposition of other input usage.

Following Saal and Parker (2001) the drinking water quality index is calculated as the ratio of the average percentage of each WaSC's water supply zones that are fully compliant with key water quality parameters, relative to the average compliance percentage for England and Wales in 1991. Water supply zones are areas designated by the water companies by reference to a source of supply in which not more than 50,000 people reside. The data were drawn from the DWI's annual reports for drinking water quality for the calendar years ending 1991-2007[10]. The drinking water quality can be defined either based on the sixteen water quality parameters or nine water quality parameters identified as being important for aesthetic, health reasons and cost reasons or based on based on the six water quality parameters identified as being indicative of how well treatment works and distribution systems are operated and maintained. Due to changes in some of the drinking water quality standards and the new regulations, the DWI report for 2005 no longer included the two quality indices that compared companies' compliance for the sixteen or nine water quality parameters with the average for England and Wales. So we decided to report results for the drinking water quality based on the six water quality parameters[11] that Ofwat also employs in his assessment and reflect how well treatment works and distribution systems are operated and maintained (Ofwat, 2006).

The sewage treatment quality index is defined as a weighted index of the percentage of connected population for which sewage receives primary treatment and the percentage of population for which sewage receives at least secondary treatment. It also implicitly includes the percentage of connected population for which sewage is not treated with a zero weight. This data choice reflects both the availability of consistent data capturing quality trends for the entire 1991-2008 period, and does clearly capture substantial increases in sewage treatment levels, particularly in the earlier part of the sample period. The sewage treatment data were taken from *Waterfacts* for the first years 1990-91 to 1995-96 and the companies' regulatory

---

[10] The DWI provides quality data based on calendar years, while all other information employed in this paper is based on fiscal years ending March 31st. We note this inconsistency in the data, but emphasize that the reported years overlap each other for 9 months. Thus, the year end to year end estimates of quality change obtained from the DWI data provide consistent estimates of quality change by the water companies, at a fixed point 9 months into each fiscal year.

[11] The six water quality parameters, which form the Operational Performance Index (OPI) are iron, manganese, aluminium, turbidity, faecal coliforms and trihalomethanes. The resulting drinking water quality index suggests an increase in quality of 10.3 percent between 1991 and 2008 after aggregating the data for all WaSCs.

returns for the fiscal years 1996-97 to 2007-08. Moreover, we henceforward refer to data based on the ending year of the fiscal years.

It is clearly necessary to employ a weighted index of these measures as both the quality and costs of higher treatment levels exceed those associated with non treatment or primary treatment alone. We therefore endeavoured to construct a cost based weighting system, although the necessary data to accomplish this was relatively limited. However, we were able to calculate relative cost measures based on the ratio of sewerage treatment costs to volumes of sewerage treatment, using two alternative cost estimates available from company regulatory returns. One of these alternative estimates was based on total sewerage treatment functional expenditure and direct costs for all treatment works, while the other was based on total sewage treatment costs for large treatment works only. These estimates suggest that higher levels of treatment are 1.68 to 2.40 times more costly than primary treatment only. Given this estimate range, we chose to weight the percentage of population receiving secondary treatment of sewage or more twice as much as the percentage receiving primary treatment only. While admittedly, somewhat ad hoc, we emphasize there is some empirical evidence to support these weights. We note that it is straightforward to demonstrate that the resulting weighted quality index is nested between an index based solely on the percentage of population receiving at least primary sewage treatment, which would underestimate gains in sewage treatment quality, and one based solely on the percentage of population receiving at least secondary sewage treatment, which would overestimate gains in sewage treatment quality.[12]

Once the quality adjusted water and sewerage outputs are constructed, quality adjusted indices are straightforward to produce, by simply repeating the procedures identified above to first produce spatially consistent quality adjusted output indices ($Y_{i,t}^{S,Q}$). A spatial aggregate quality-adjusted aggregated output price index is then constructed as $P_{i,t}^{S,Q} = R_{i,t}^{S} / Y_{i,t}^{S,Q}$. We can also derive a spatial implicit quality index ($Q_{i,t}^{S}$) which measures the implied difference in quality relative to the base firm as

---

[12] To highlight this, we note that while our weighted index implies an increase in sewage treatment quality of 19.3% for all England and Wales between 1991 and 2008, an index based only on population receiving at least primary treatment would indicate a quality improvement of 13.7% while one based only on the percentage of population receiving at least secondary treatment of sewage would indicate a 25.4% quality improvement. However, our approach not only provides a mid range estimate between these two more extreme measures, but also better reflects the process of improving sewage treatment quality that occurred through both treating previously untreated sewage, and increasing the level of sewage treatment.

$Q_{i,t}^{S} = Y_{i,t}^{S,Q} / Y_{i,t}^{S}$. Therefore, quality adjusted spatial outputs and output prices can also be respectively expressed as $Y_{i,t}^{S,Q} = Q_{i,t}^{S} Y_{i,t}^{S}$ and $P_{i,t}^{S,Q} = P_{i,t}^{S} / Q_{i,t}^{S}$, which illustrate that the impact on spatial output quantities will be perfectly balanced by changes in spatial output prices. This also implies that measured spatial economic profitability ($\pi_{i,t}^{S}$) is not influenced by quality adjustment. In contrast, the impact of quality adjustment implies that quality adjusted spatial TFP can be expressed as $TFP_{i,t}^{S,Q} = Q_{i,t}^{S} TFP_{i,t}^{S}$ and similarly, quality adjusted spatial price performance can be expressed as $TPP_{i,t}^{S,Q} = TPP_{i,t}^{S} / Q_{i,t}^{S}$ and spatial economic profitability can be decomposed as, $\pi_{i,t}^{S} = TFP_{i,t}^{S,Q} TPP_{i,t}^{S,Q}$.

In an analogous manner, we can derive measures of relative quality adjusted output indices over time, $Y_{i,t}^{R,Q}$ and relative implicit quality index over time ($Q_{i,t}^{R}$) which measures the implied difference in quality over time relative to the base firm at the base period as $Q_{i,t}^{R} = Y_{i,t}^{R,Q} / Y_{i,t}^{R}$. Therefore, measures of quality adjusted relative outputs and output prices can also be expressed as $Y_{i,t}^{R,Q} = Q_{i,t}^{R} Y_{i,t}^{R}$ and $P_{i,t}^{R,Q} = P_{i,t}^{R} / Q_{i,t}^{R}$. Thus, quality adjusted relative TFP and TPP over time can be expressed as $TFP_{i,t}^{R,Q} = Q_{i,t}^{R} TFP_{i,t}^{R}$ and $TPP_{i,t}^{R,Q} = TPP_{i,t}^{R} / Q_{i,t}^{R}$, whereas the relative economic profitability over time as $\pi_{i,t}^{R} = TFP_{i,t}^{R,Q} TPP_{i,t}^{R,Q}$. This also implies that measured relative economic profitability ($\pi_{i,t}^{R}$) is not influenced by quality adjustment. Also, we can produce measures of unit-specific quality adjusted output indices over time, $Y_{i,t}^{US,Q}$ and implicit quality index over time ($Q_{i,t}^{US}$) which measures the implied difference in unit-specific quality over time as $Q_{i,t}^{US} = Y_{i,t}^{US,Q} / Y_{i,t}^{US}$. Therefore, estimates of temporal quality adjusted outputs and output prices can also be expressed as $Y_{i,t}^{US,Q} = Q_{i,t}^{US} Y_{i,t}^{US}$ and $P_{i,t}^{US,Q} = P_{i,t}^{US} / Q_{i,t}^{US}$. Thus, the quality adjusted unit-specific TFP and TPP over time can be expressed as $TFP_{i,t}^{US,Q} = Q_{i,t}^{US} TFP_{i,t}^{US}$ and $TPP_{i,t}^{US,Q} = TPP_{i,t}^{US} / Q_{i,t}^{US}$, while the unit-specific economic profitability over time as $\pi_{i,t}^{US} = TFP_{i,t}^{US,Q} TPP_{i,t}^{US,Q}$.

As stated above, our adjustment of output prices and quantities for quality implies that any changes in the quality adjusted TPP index over time are balanced by an equivalent proportional change in the quality adjusted TFP index over time, thereby keeping the measurement of economic profitability unaffected by quality adjustment. We wish to emphasize that this is a reasonable assumption. Firstly, taking account of quality should not effect our underlying definition of economic profitability as turnover divided by economic costs. Secondly, it reflects the mathematical necessity that if turnover is constant, allowing for increases in output resulting from quality improvements, must result in a perfectly proportional reduction in output prices. Therefore, by adjusting TFP and TPP measures for quality keeping economic profitability unchanged, we are able to offer an alternative decomposition of unit-specific profitability growth, which, will more properly attribute quality improvements to productivity improvement, rather than to over estimated improvements in price performance that would result from a quality-unadjusted measure. Moreover, as we will illustrate, this allows a further decomposition of equation (6.8) into the catch-up in quality regarding productivity and price performance achieved by less productivity firms and the quality growth in productivity and price performance of the base firm in a multilateral context.

Given the derivation of the spatial implicit output quality index ($Q^S_{i,t}$) which measures the implied difference in quality relative to the base firm, we are able to construct measures of the catch-up in quality, $Q^C_{i,t}$, as a ratio of the spatial implicit quality index for any firm i to the base firm between year 1 and t, $Q^C_{i,t} = \frac{Q^S_{i,t}}{Q^S_{i,1}}$. Moreover, given the availability of $Q^S_{i,t}$, $Q^{US}_{i,t}$ and $Q^R_{i,t}$ the catch up in quality can be expressed in a similar manner to what was demonstrated in equation (6.7):

$$Q^C_{i,t} = \frac{Q^S_{i,t}}{Q^S_{i,1}} = \frac{\dfrac{Q^R_{i,t}}{Q^R_{b,t}}}{\dfrac{Q^R_{i,1}}{Q^R_{b,1}}} = \frac{\dfrac{Q^R_{i,t}}{Q^R_{i,1}}}{\dfrac{Q^R_{b,t}}{Q^R_{b,1}}} = \frac{Q^{US}_{i,t}}{Q^{US}_{b,t}} \qquad (6.29)$$

Rearranging (6.29), we can express the unit-specific quality index of any firm i over time as a function of the catch-up in quality to the base firm and the quality improvement of the base firm, $Q^{US}_{i,t} = Q^C_{i,t} Q^{US}_{b,t}$.

Given our discussion of our approach to quality adjustment, the decomposition of firm specific economic profitability change detailed in (6.8) can now be extended, in the multilateral context, as follows:

$$\begin{aligned}
\pi_{i,t}^{US} &= \pi_{i,t}^{C}\pi_{b,t}^{US} = (TFP_{i,t}^{C}TFP_{b,t}^{US})(TPP_{i,t}^{C}TPP_{b,t}^{US}) \\
&= \left(TFP_{i,t}^{US,Q}\right)\left(TPP_{i,t}^{US,Q}\right) = \left(TFP_{i,t}^{C,Q}TFP_{b,t}^{US,Q}\right)\left(TPP_{i,t}^{C,Q}TPP_{b,t}^{US,Q}\right) \qquad (6.8') \\
&= \left(TFP_{i,t}^{C}Q_{i,t}^{C}\right)\left(TFP_{b,t}^{US}Q_{b,t}^{US}\right)\left(\frac{TPP_{i,t}^{C}}{Q_{i,t}^{C}}\right)\left(\frac{TPP_{b,t}^{US}}{Q_{b,t}^{US}}\right)
\end{aligned}$$

Thus, as in (6.8), in the first line of (6.8'), unit-specific economic profitability change, $\pi_{i,t}^{US}$, can be decomposed as a function of the quality unadjusted catch-up in productivity, $TFP_{i,t}^{C}$ and the productivity growth of the benchmark firm, $TFP_{b,t}^{US}$ and the quality unadjusted catch-up in price performance, $TPP_{i,t}^{C}$ and the price performance growth of the benchmark firm, $TPP_{b,t}^{US}$ By including quality in TFP and TPP measures, in the second line of equation (6.8'), the unit-specific economic profitability over time can be expressed as a function of the unit-specific quality adjusted productivity, $TFP_{i,t}^{US,Q}$ and quality-adjusted price performance change, $TPP_{i,t}^{US,Q}$. This can be further decomposed as a function of the quality adjusted catch-up in productivity, $TFP_{i,t}^{C,Q}$ and the quality adjusted productivity growth of the benchmark firm, $TFP_{b,t}^{US,Q}$ and the quality adjusted catch-up in price performance, $TPP_{i,t}^{C,Q}$ and the quality adjusted price performance growth of the benchmark firm, $TPP_{b,t}^{US,Q}$. Finally, the third line of (6.8') demonstrates the impact of quality in TFP and TPP measures over time. Thus, unit-specific economic profitability change, $\pi_{i,t}^{US}$, can be decomposed as a function of the quality unadjusted catch-up in productivity, $TFP_{i,t}^{C}$, the catch-up in quality regarding productivity, $Q_{i,t}^{C}$ and the quality-unadjusted productivity and quality performance over time of the benchmark firm, $TFP_{b,t}^{US}$ and $Q_{b,t}^{US}$ and the quality unadjusted catch-up in price performance, $TPP_{i,t}^{C}$ the catch-up in quality regarding price performance, $1/Q^{C}_{i,t}$ and the price performance and quality growth of the benchmark firm, $TPP^{US}_{b,t}$ and $1/Q^{US}_{b,t}$. If $TFP^{C}_{i,t} > 1$ or $TPP^{C}_{i,t} > 1$, then firm i improved its productivity or price performance relative to the base firm from year 1 to t, whereas a value lower than 1 indicates that productivity or price performance of firm i has declined relative to that of the base firm. If $Q^{C}_{i,t} > 1$ or $1/Q^{C}_{i,t} < 1$, then the firm i improved its quality regarding productivity or price performance relative to the base firm from year 1 to year t, whereas a value lower than 1 indicates that relative quality regarding productivity or price performance of firm i has declined relative to that of the base firm. Finally, the decomposition of the unit specific economic profitability over time in equation (6.8') can be visualized in Figure 6.2. As adjustments for quality affect the productivity and price performance measures leaving the measured economic profitability unchanged, the unit-specific profitability growth can be expressed as a function of the unit-specific quality adjusted productivity and quality-adjusted price performance change. This can be further decomposed as a function of the quality adjusted catch-up in productivity, and the quality adjusted productivity growth of the benchmark firm, and the quality adjusted catch-up in price performance, and the quality adjusted price performance growth of the benchmark firm. The inclusion of quality in our analysis allows us to finally decompose unit-specific economic profitability change as a function of the quality unadjusted catch-up in productivity, the catch-up in quality regarding productivity, and the quality-unadjusted productivity and quality performance over time of the benchmark firm, and the quality unadjusted catch-up in price performance, the catch-up in quality regarding price performance, and the price performance and quality growth of the benchmark firm.

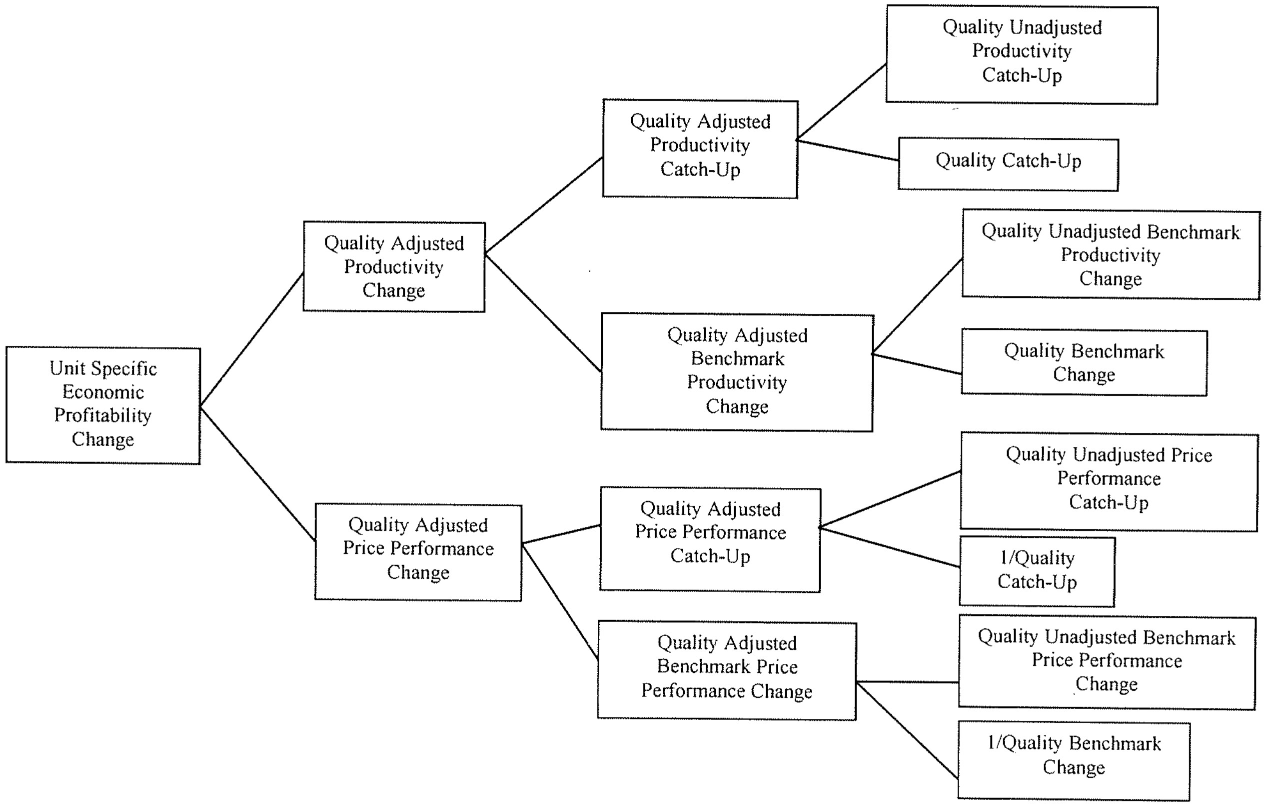


Figure 6-2 Decomposition of Unit Specific Economic Profitability Change after Adjustments for Quality

6.5 Results From Productivity, Price Performance and Profitability Computations

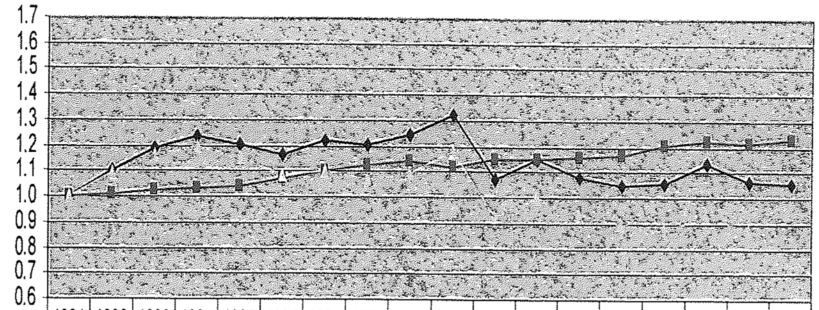
The above spatial and relative profitability, productivity and price performance measures were defined relative to the base firm in the sample. However, if the base firm is defined as the firm with the highest productivity in the sample, then each firm's productivity, prices and profits will be relative to this best practice or benchmark firm.[13] In this section we first report geometric average measures of unit-specific profitability, productivity and price performance in Figure 6.3. Subsequently, we demonstrate the further decomposition that is facilitated by our methodological approach by decomposing theses changes into an average catch-up component and the performance of the benchmark firm. Moreover, we first illustrate this for a quality unadjusted model in Figures 6.4, 6.5 and 6.6, and then illustrate the impact of quality on these measures in Figures 6.7 to 6.11.

Figure 6.3 illustrates the decomposition of unit-specific economic profitability change into unit-specific quality-unadjusted productivity and price performance change over the period 1991-2008, thereby replicating the work of other authors including Saal and Parker (2001), which provided measures of unit-specific economic profitability, productivity and price performance for WaSCs over 1985-1999 using a Tornqvist index. The results indicate that between 1991 and 2008, average economic profitability increased by 5.9%, which was attributed to an improvement in TFP of 22.9% and a reduction in TPP of 13.9%. On average there was a stable increase in TFP over time, while TPP followed an upward trend until 1994, which was interrupted in 1995, but was again followed by a substantial increase between 1999 and 2000. We note that during the years 1991-1994, average economic profitability increased due to increases in TPP which was substantially greater than TFP growth. As documented in previous studies, Ofwat's tightening of price caps in the 1994 price review decreased the growth in real output prices and therefore resulted in a downward trend for both TPP and economic profitability until 1998, while TFP

---

[13] We have not identified firms for confidentially reasons. The same firm is consistently found to have the highest spatial productivity estimates for both quality unadjusted and quality adjusted models in all years, and is therefore modelled as the benchmark most productive firm in each year of our study Moreover, we note that his same firm was found to have the highest spatial productivity estimates in each year of the study regardless of whether we applied the spatially consistent Fisher indices provided in the main text, similar spatially consistent Tornqvist indices, or the multilateral translog index for WaSCs based on the Tornqvist index developed by Caves et al (1982a). Furthermore, there is little substantive difference between the results regardless of which method is employed.

continued to rise steadily. Our finding therefore confirms Saal and Parker's (2001) study, which found that during 1991-1999, positive changes in economic profitability were mainly attributed to changes in TPP rather in TFP. However, Figure 6.3 extends their study by including results for unit-specific profitability, productivity and price performance changes until 2008.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Economic Profitability	1.000	1.102	1.186	1.234	1.205	1.170	1.217	1.206	1.244	1.326	1.069	1.150	1.082	1.050	1.054	1.138	1.064	1.059
■ TFP	1.000	1.006	1.021	1.029	1.039	1.072	1.100	1.126	1.140	1.120	1.150	1.148	1.157	1.168	1.209	1.220	1.215	1.229
▲ TPP	1.000	1.096	1.162	1.199	1.160	1.091	1.106	1.072	1.091	1.184	0.930	1.002	0.936	0.899	0.872	0.933	0.876	0.861

Figure 6-3 Decomposition of Average Unit Specific Profitability into Quality Unadjusted Average Unit Specific TFP and TPP

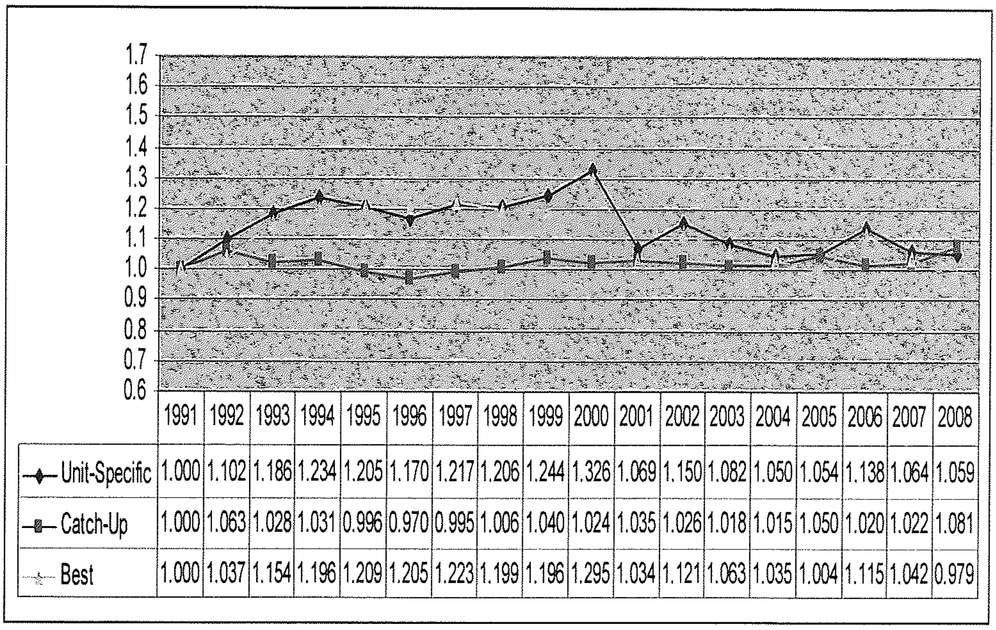
These extended results demonstrate that after 2000, reduced output prices caused TPP to dramatically decline, and its value remained consistently below 1 after 2000. This indicates that regulatory price changes implemented after 2000, caused the price performance of firms to fall substantially below its level in 1991. Moreover, average unit-specific TPP followed a downward trend except for 2006, when output prices were allowed to momentarily rise in the first year of the 2006-10 regulatory period. Unsurprisingly, given the dramatic fall in price performance after 2000, average economic profitability also substantially declined, even though TFP continued to follow a steady upward trend, which was only momentarily interrupted in 2007. Thus, in the post 2000 period, trends in temporal economic profitability continued to follow the trend of TPP, indicating that changes in price performance continue to be the main determinant of changes in economic profitability.

Nevertheless, while TPP fell below 1991 levels after 2000 average economic profitability did not, thereby implying that on average profitability in the industry remained moderately higher than in the immediate aftermath of privatisation. This is because of the significant and continuing gains in TFP between 1991 and 2000 that more than offset the dramatic tightening of regulated output prices in 2001. Thus, the immediate impact of the 1999 price review in 2001 is consistent with an interpretation emphasizing that Ofwat chose to pass considerable accumulated past productivity improvements to consumers, thereby worsening profitability, but still left the industry more profitable than in 1991. Moreover, the steady decline in average price performance, gains in TFP and relatively stable economic profitability that have characterized the 2001-2008 period, suggests that Ofwat is now more focused on passing productivity benefits to consumers, and maintaining stable profitability than in the earlier regulatory periods.

Our discussion of Figure 6.3 has clearly illustrated the decomposition of unit-specific economic profitability change into unit-specific quality-unadjusted productivity and price performance change and also demonstrates that this approach can capture the significant shift in regulatory practice after 2000. However, given that Ofwat operates a system of yardstick regulation which is designed to encourage catch up to benchmark firm performance, the methodology developed above, is particularly relevant. Thus, we should expect that the performance improvement of laggard firms should exceed that of benchmark firms. This is because the price caps set for benchmark firms should only require them to continue improving their performance through technical change, while price caps for non benchmark firms will also require them to catch up to the benchmark firm. Thus, the multilateral models develop above can be used to illustrate the contribution of benchmark performance and average catch-up to average firm performance. We therefore first turn to an illustration of the decomposition illustrated in (6.8) for the quality unadjusted in Figures 6.4-6.6, before illustrating the further quality adjusted decomposition detailed in (6.8’) in Figures 6.7-6.11.

Looking at Figure 6.4, we note that the lax price caps set at privatization as documented is past studies, allowed average economic profitability to increase significantly until 1994 by 23.4% and that this exceeded benchmark economic profitability growth which increased by 19.6%, therefore allowing an average catch-up to benchmark profitability of 3.1%. The tightening of price caps from 1994 resulted in a downward trend for average and benchmark economic profitability. Thus, during the years 1995-1998, the average firm did not improve its economic profitability relative to the benchmark but this was once again interrupted during

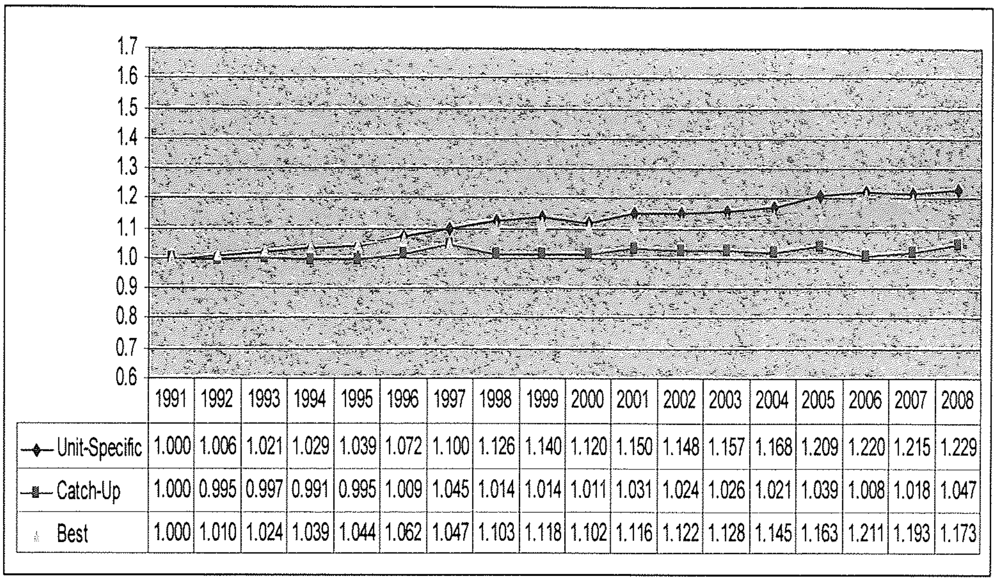
1998-2000, when average economic profitability increased more than benchmark profitability, allowing average catch-up of 2.4%. The substantial reduction in output prices due to the tightened 1999 price review resulted in a significant reduction in average and benchmark economic profitability for the subsequent years which showed an upward trend only in 2002 and in 2006. We note that benchmark firm realized significant decline in its economic profitability in 2001, and despite an improvement in 2002, further declines meant that its profitability in 2005 was only 0.04% of its level in 1991. Moreover, despite an uptick of benchmark profitability to 1.115 in 2006, by 2008 benchmark profitability was only 97.9% of its 1991 level. In contrast, while average economic profitability was also considerably lower after 2000, it has never declined below average 1991 levels. As a result, average firm showed high levels of catch-up in profitability relative to the benchmark after 2001. However, this is mainly explained by the relative decline in the economic profitability of the benchmark firm. Thus, over the 1991 to 2008 period the average company caught-up to benchmark economic profitability by 8.1%, but this was mainly attributable to a decline in benchmark profitability of 2.1%.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Unit-Specific	1.000	1.102	1.186	1.234	1.205	1.170	1.217	1.206	1.244	1.326	1.069	1.150	1.082	1.050	1.054	1.138	1.064	1.059
■ Catch-Up	1.000	1.063	1.028	1.031	0.996	0.970	0.995	1.006	1.040	1.024	1.035	1.026	1.018	1.015	1.050	1.020	1.022	1.081
▲ Best	1.000	1.037	1.154	1.196	1.209	1.205	1.223	1.199	1.196	1.295	1.034	1.121	1.063	1.035	1.004	1.115	1.042	0.979

Figure 6-4 Decomposition of Average Unit-Specific Profitability into Average Profitability Catch-Up and Profitability of the Benchmark Firm

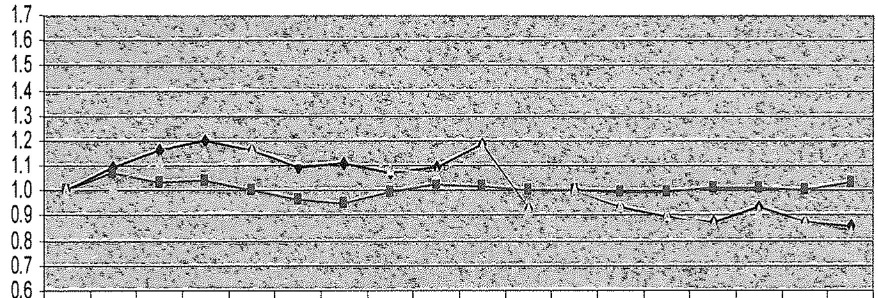
The decomposition of average unit-specific productivity growth into quality unadjusted productivity change of the benchmark firm and quality unadjusted average productivity catch-up relative to the benchmark firm is depicted in Figure 6.5. Until 1995 there were actually negative productivity catch-up as the productivity improvements for the average company amounted to 3.9%, while the benchmark company improved its productivity by 4.4%. This finding suggests that the lax price caps set at privatization encourage neither average or benchmark firms to achieve high productivity levels. This trend was interrupted after 1995 when both average and benchmark productivity performance significantly improved. We note that during the years 1996-2000 when price caps were first tightened, average companies should have had stronger incentives to catch-up to benchmark, while the benchmark company should also have been incentivized to continue to improve its productivity. By 2000, average cumulative productivity increased by 12% and this growth exceeded that of the benchmark firm, which achieved cumulative improvement of 10.2%, thereby indicating total catch-up in productivity of 1.1% between 1991 and 2000. Moreover, significant productivity gains for the average firm relative to the benchmark firm also continued after 2000. Thus, our results suggest that the implementation of even tighter price caps in 1999 further encouraged less productive firms to improve their performance relative to the benchmark, even though the benchmark firm continued to improve its performance. Thus, by 2004, the cumulative measures of productivity change since 1991 indicate that average company improved its productivity by 16.8% catching up to the benchmark productivity by 2.1%, while the benchmark firm improved its productivity by 14.5%. During the last price review period, average productivity growth again substantially exceeded the productivity growth of the benchmark firm, resulting in high levels of productivity catch-up between 2005 and 2008, although this is largely explained by substantial declines in benchmark productivity after 2006. Thus, in sum over the entire 1991-2008 regulatory period, average productivity improved by 22.9%, while benchmark productivity improved its productivity by 16.6% allowing an average productivity catch-up of 4.7%. Moreover, our results suggest that all of this catch-up can be attributed to the post 1995 period, after Ofwat first tightened price caps, and most of it can be attributed to the post 2000 period, following the even more stringent 1999 price review.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Unit-Specific	1.000	1.006	1.021	1.029	1.039	1.072	1.100	1.126	1.140	1.120	1.150	1.148	1.157	1.168	1.209	1.220	1.215	1.229
■ Catch-Up	1.000	0.995	0.997	0.991	0.995	1.009	1.045	1.014	1.014	1.011	1.031	1.024	1.026	1.021	1.039	1.008	1.018	1.047
▲ Best	1.000	1.010	1.024	1.039	1.044	1.062	1.047	1.103	1.118	1.102	1.116	1.122	1.128	1.145	1.163	1.211	1.193	1.173

Figure 6-5 Decomposition of Average Unit-Specific Quality Unadjusted TFP Change into Benchmark TFP Change and Average Catch-Up to the Benchmark Firm

The decomposition of average unit-specific economic price performance change into the quality unadjusted price performance change of the benchmark firm and quality unadjusted average price performance catch-up relative to that firm over time is displayed at Figure 6.6. The results indicated that until 1994 when price caps were relatively lax, both average and benchmark price performance significantly increased by 19.9% and 15.1% respectively. Average TPP growth exceeded benchmark TPP growth allowing an average catch-up in price performance of 4.1%. The tighter 1994 price review, led to a substantial downward trend in average and benchmark TPP until 1998. We note that during the years 1996-1998 benchmark TPP growth exceeded average TPP growth and therefore there were not any price performance catch-up gains on average. After 1998, average TPP increased more than benchmark TPP but by 2000, there was a broad convergence in average and benchmark TPP as the respectively demonstrated cumulative increases of 18.4% and 17.5% since 1991. However, the dramatic impact of the 1999 price review obliged the companies to reduce their output prices significantly and after 2000 there was a significant decline in average and benchmark TPP, except for the year 2006 when relatively looser price caps were introduced. We notice that during the years 2001-2004, there was little or no difference between average and benchmark TPP, while during the years 2005-2008 average TPP exceeded benchmark TPP showing the highest levels of price performance catch-up in 2006 and in 2008. By 2008, average TPP had been reduced by 13.9% relative to 1991 levels, while benchmark TPP had been reduced even more by 16.5%, thereby allowing an average catch-up in price performance of 3.2%. Thus, Figure 6.6 clearly illustrates that in the post 1999 price review period, the price performance of all firms is substantially lower than in the first 10 years after privatisation.

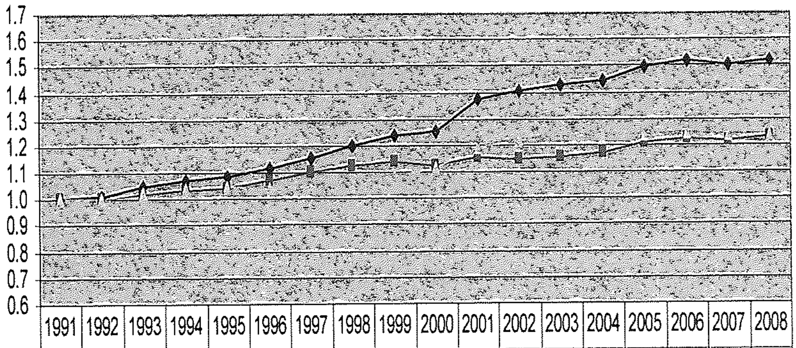


	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
—◆— Unit-Specific	1.000	1.096	1.162	1.199	1.160	1.091	1.106	1.072	1.091	1.184	0.930	1.002	0.936	0.899	0.872	0.933	0.876	0.861
—■— Catch-Up	1.000	1.068	1.031	1.041	1.002	0.962	0.952	0.992	1.026	1.014	1.003	1.002	0.992	0.994	1.010	1.013	1.004	1.032
✧ Best	1.000	1.027	1.127	1.151	1.158	1.134	1.168	1.087	1.070	1.175	0.926	1.000	0.943	0.904	0.863	0.921	0.873	0.835

Figure 6-6 Decomposition of Average Unit-Specific Quality Unadjusted TPP Change into Benchmark TPP Change and Average Catch-Up to the Benchmark Firm

As discussed in section 6.3, the inclusion of quality in our productivity and price performance measures, allows us to decompose unit-specific economic productivity as a function of quality adjusted TFP and TPP change, which can be further decomposed into a quality adjusted catch-up in TFP and TPP achieved by less productive firms relative to the benchmark firm and the quality adjusted TFP and TPP growth obtained by the benchmark firm. This decomposition illustrated at the second line of equation (6.8') and is visualised at Figures 6.7, 6.8, 6.9 and 6.10.

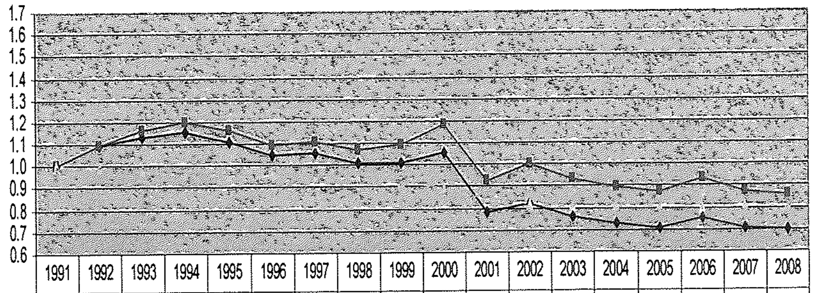
We begin with Figure 6.7 which depicts the decomposition of quality adjusted average TFP change into quality unadjusted average TFP change and quality change. High capital investment programs to improve quality conditions since privatization had a positive impact on quality adjusted output growth and consequently, quality adjusted TFP increased more than quality unadjusted TFP. Over the whole regulatory period average quality adjusted TFP improved by 51.7%, whereas average quality unadjusted TFP improved by only 22.9% implying that average estimated quality change amounted to 23.4%. Much of the measured quality improvement occurred during the years 1991-2002 and quality showed its highest level of improvement in the years 1999 and 2002. Thus, by 2002, average quality improved by 22% resulting in an increase in average quality adjusted TFP of 40.1% and exceeded average quality unadjusted TFP which improved by only 14.8%. After 2003, on average there were small improvements in quality and thus, small changes in the quality adjusted TFP growth rate, whereas in the last two years of our study average quality followed a slightly decline trend. Nevertheless productivity still continued to improve in this later period, suggesting that firms were able to achieve productivity improvements by reducing input usage.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
—◆— Quality Adjusted TFP	1.000	1.011	1.046	1.067	1.086	1.119	1.156	1.197	1.236	1.254	1.370	1.401	1.424	1.442	1.493	1.516	1.501	1.517
—■— Quality Unadjusted TFP	1.000	1.006	1.021	1.029	1.039	1.072	1.100	1.126	1.140	1.120	1.150	1.148	1.157	1.168	1.209	1.220	1.215	1.229
Quality	1.000	1.006	1.025	1.037	1.045	1.044	1.051	1.063	1.084	1.120	1.191	1.220	1.231	1.234	1.235	1.243	1.236	1.234

Figure 6-7 Decomposition of Average Unit Specific Quality Adjusted TFP Change into Average Unit-Specific TFP and Quality Change

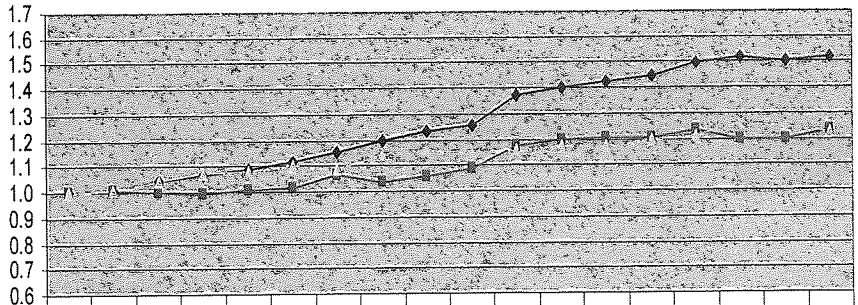
Figure 6.8 displays the decomposition of quality adjusted average unit-specific TPP change into quality unadjusted average TPP change and quality change. Since output prices are adjusted for quality as we discussed in section 6.4, on average the magnitude of change in quality adjusted TPP must exceed that of quality unadjusted TPP. We would therefore emphasize that the quality adjusted TPP index must also follow the general trend of the quality-unadjusted index, but it also must demonstrate a more significant decline in price performance, as it allows for the output enhancing impact of quality improvements. During the lax price cap period 1991-1994, increases in quality unadjusted TPP exceeded the quality adjusted TPP implying that increases in output prices were greater than the quality adjusted output prices. This upward trend was interrupted in 1995 followed by a downward trend until 1998, whereas during the years 1999-2000 average quality unadjusted TPP and quality adjusted TPP started to increase again. The tightened 1999 price review obliged the companies to reduce their output prices and the magnitude of the reduction in quality adjusted TPP was significantly greater than the quality unadjusted TPP on average. Between the years 2000 and 2001 there was a significant fall in average quality unadjusted TPP and quality adjusted TPP by 0.930/1.184 = 0.785 or 21.5% and 0.780/1.058 = 0.737 or 26.3% respectively. After 2001, there was a downward trend for average quality unadjusted and quality adjusted TPP except for the years 2002 and 2006, where new looser price caps were introduced. We note that after 1998, on average quality adjusted TPP took a value lower than 1 implying that after controlling for quality the reduction in quality adjusted output prices was greater than the quality unadjusted output prices and therefore, relative to 1991, by 2008 average quality adjusted TPP reduced by 30.2%, whereas average quality unadjusted TPP declined by 13.9%, implying that the impact of average quality in output prices and therefore in average TPP was approximately 19%. Thus, Figure 6.8 clearly suggests that, while quality improvements have contributed to the productivity performance of the WaSCs, they have also contributed negatively to their price performance.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Quality Adjusted TPP	1.000	1.090	1.134	1.156	1.109	1.045	1.052	1.008	1.007	1.058	0.780	0.821	0.760	0.729	0.706	0.750	0.709	0.698
■ Quality Unadjusted TPP	1.000	1.096	1.162	1.199	1.160	1.091	1.106	1.072	1.091	1.184	0.930	1.002	0.936	0.899	0.872	0.933	0.876	0.861
☆ 1/Quality	1.000	0.994	0.976	0.964	0.957	0.958	0.952	0.941	0.923	0.893	0.840	0.819	0.813	0.810	0.810	0.805	0.809	0.810

Figure 6-8 Decomposition of Average Units Specific Quality Adjusted TPP Change into Average Unit-Specific TPP and Quality Change

The decomposition of quality adjusted average unit-specific productivity growth into the quality adjusted productivity growth of the benchmark firm and average quality adjusted productivity catch-up is depicted in Figure 6.9. The figure clearly illustrates that until 1994 there were small or no catch up gains in quality-adjusted productivity by the average company since its productivity improved by 6.7%, while the benchmark company improved its productivity by 7.1%. In contrast, due to sharp increases in measure quality between 1996 and 2002, average quality adjusted TFP increased more rapidly than benchmark quality adjusted TFP, thereby allowing the average company to catch-up considerably, with catch up amounting to 19.5% of cumulative productivity growth for the average firm by 2002. Even after 2002 the average company achieved still significant levels of catch-up in quality adjusted productivity until 2005, which must be attributed to input usage reductions. Thus, relative to 1991 levels, by 2005, average quality adjusted productivity had increased by 49.3% and exceeded that of benchmark firm, which had improved by 21.2%, therefore indicating productivity catch-up of 23.2%. Moreover, the considerable increase in average profitability relative to the benchmark firm, observed in Figure 6.4 must be attributed to this catch up effect. Nevertheless, after 2005, when the relatively looser 2004 price review came into effect, high levels of productivity catch-up are no longer indicative of general productivity improvements, as average quality adjusted productivity levels were largely static after 2005. Instead, they reflect a substantial decline in the benchmark firm's productivity after 2006. Thus, our results may be interpreted as suggesting that after the 2004 price review, substantial productivity improvements were no longer occurring.

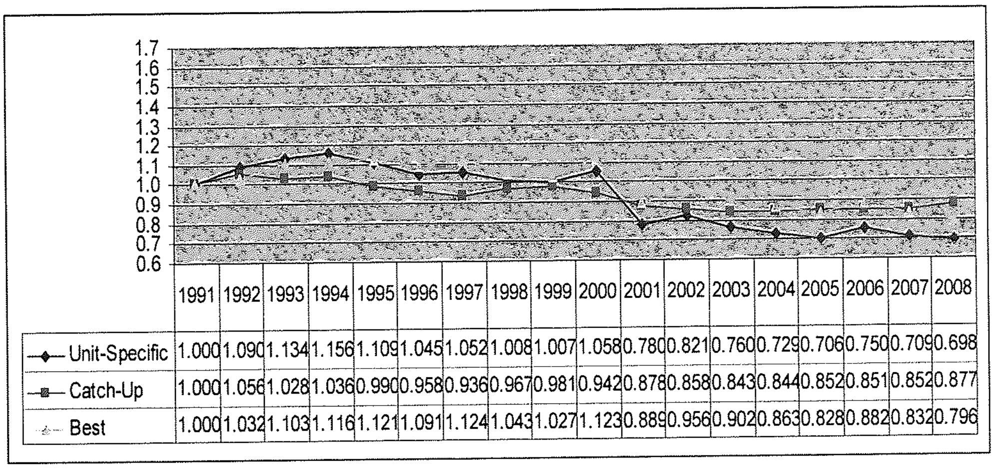


	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
—◆— Unit-Specific	1.000	1.011	1.046	1.067	1.086	1.119	1.156	1.197	1.236	1.254	1.370	1.401	1.424	1.442	1.493	1.516	1.501	1.517
—■— Catch-Up	1.000	1.006	1.000	0.996	1.007	1.013	1.063	1.041	1.061	1.088	1.179	1.195	1.208	1.202	1.232	1.199	1.199	1.233
▲ Best	1.000	1.005	1.046	1.071	1.079	1.104	1.088	1.150	1.164	1.153	1.162	1.172	1.179	1.199	1.211	1.265	1.252	1.230

Figure 6-9 Decomposition of Average Unit-Specific Quality Adjusted TFP Change into Benchmark TFP Change and Average Catch-Up to the Benchmark Firm

The decomposition of quality adjusted average unit-specific TPP growth into the quality adjusted TPP growth of the benchmark firm and average quality adjusted TPP catch-up to the benchmark firm over time is depicted in Figure 6.10. Until 1994, quality adjusted average TPP growth exceeded benchmark TPP growth allowing an average catch-up in price performance of 3.6%. The tightened 1994 price review resulted in a substantial downward trend in quality adjusted average and benchmark TPP during the years 1996-1998, which was interrupted in 1999. We note that after 1995 and until the end of the period of study, quality adjusted benchmark TPP always exceeded quality adjusted average TPP. Moreover between 1995 and 2000 there was also a steady erosion of average price performance relative to benchmark price performance, as reflected in the catch up index from 0.990 to 0.942. This suggests a considerable rebalancing of regulatory price decisions in favour of the benchmark firm, which was even more dramatically extended with the implementation of the 1999 price review in 2001. Thus, despite a massive reduction in benchmark price performance from 1.123 to 0.889 of 1991 levels between 2000 and 2001, average price performance fell even further, as the decline of average quality adjusted TPP from 1.058 to 0.780 resulted in the catch up index falling from 0.942 to 0.878. It is therefore appropriate to interpret these results as substantial positive evidence demonstrating that both the 1994 and 1999 price reviews resulted in considerable movement to a regulatory price cap system consistent with a yardstick regulation regime. We would moreover offer the suggestion, that this better alignment of regulated prices with the principles of yardstick regulation is likely to have contributed significantly to both the catch-up in quality adjusted productivity illustrated in Figure 6.9, and the catch up in economic profitability illustrated in Figure 6.4.

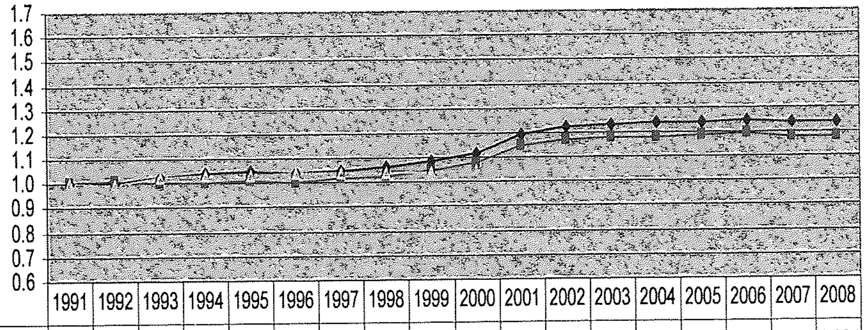
Further, considering the post 2001 period, reveals a steady downward trend in quality adjusted average and benchmark TPP except for the years 2002 and 2006. This overall finding supports a steady deterioration in price performance, which suggests that in practice, price caps have become even tighter since 2001. While , the catch up index reached a low of 0.843 in 2003 and has moderately increased to 0.877 in 2008, its trend in the post 2001 period largely suggests that the relatively superior price performance of the benchmark firm was maintained in the 2004 price review. Our results therefore suggest that when quality is taken into account in TPP measures, the broad convergence after 2000 between average and benchmark firm price performance which was observed in the quality unadjusted TPP results in Figure 6.6 is no longer present. Stated differently, when quality is taken into account, an average firm saw its price performance decline relative to the benchmark by 12.3% between 1991 and 2008 as benchmark quality adjusted benchmark TPP declined by only 20.4% while average TPP showed a higher reduction of 31.2%.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
—◆— Unit-Specific	1.000	1.090	1.134	1.156	1.109	1.045	1.052	1.008	1.007	1.058	0.780	0.821	0.760	0.729	0.706	0.750	0.709	0.698
—■— Catch-Up	1.000	1.056	1.028	1.036	0.990	0.958	0.936	0.967	0.981	0.942	0.878	0.858	0.843	0.844	0.852	0.851	0.852	0.877
—▲— Best	1.000	1.032	1.103	1.116	1.121	1.091	1.124	1.043	1.027	1.123	0.889	0.956	0.902	0.863	0.828	0.882	0.832	0.796

Figure 6-10 Decomposition of Average Unit-Specific Quality Adjusted TPP Change into Benchmark TPP Change and Average Catch-Up to the Benchmark Firm

Finally, Figure 6.11 shows the decomposition of average unit-specific quality change into average quality catch-up relative to the benchmark firm and the quality change of the benchmark firm, as illustrated in the third line of equation (6.8’). Until 1997, there were small or no gains in average quality relative to benchmark quality but after 1998 and most of the period of study average quality growth significantly exceeded benchmark quality growth, with particularly high levels of quality catch-up during between 1998 and 2002. By 2005, average quality improved by 23.5% while benchmark quality increased by 4.1% allowing average quality to catch-up to the benchmark by 18.6%. After 2005, average quality continue to increase at a lower rate, however, it showed a significant decline in 2007 and in 2008 which affected the quality adjusted TFP growth rates as we discussed in Figure 6.7, whereas benchmark quality followed a stable slow upward trend. We need to emphasize that the small quality growth of the benchmark firm did not imply that the benchmark did not achieve significant quality levels. In contrast, our results suggest that at privatization the quality standards of the benchmark firm had already been at a high level and by 2005 on average the less productive firms had significantly improved their quality relative to the benchmark and had finally reached the higher levels of quality of the benchmark firm. Given the considerable cost of these quality improvements, Figure 6.11 therefore only serves to further illustrate the importance of controlling for quality changes if we wish to properly gauge relative productivity, price, profitability, and catch up performance.



	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
◆ Unit-Specific	1.000	1.006	1.025	1.037	1.045	1.044	1.051	1.063	1.084	1.120	1.191	1.220	1.231	1.234	1.235	1.243	1.236	1.234
■ Catch-Up	1.000	1.011	1.003	1.005	1.012	1.004	1.017	1.026	1.047	1.076	1.143	1.167	1.177	1.178	1.186	1.190	1.178	1.177
△ Best	1.000	0.995	1.021	1.031	1.033	1.040	1.039	1.043	1.042	1.046	1.042	1.045	1.045	1.048	1.041	1.044	1.049	1.048

Figure 6-11 Decomposition of Average Unit-Specific Quality Change into Average Quality Change Catch-Up and Benchmark Quality Change

6.6 Summary and Conclusions

This chapter analyzed the impact of regulation on the financial performance of WaSCs in England and Wales over the period 1991-2008. We employed a panel index number technique to decompose profits into total factor productivity and price performance, and demonstrated several different but theoretically related methods to link productivity, price performance and profitability. Thus, we not only estimated and decomposed unit-specific (temporal) profitability of each firm over time, but also illustrated a multilateral spatial Fisher index derived by the EKS method, that allowed multilateral spatial measures between all the pairs of companies included in the analysis at any given year (multilateral spatial comparisons). We also linked together the spatial and temporal results in order to derive estimates of relative productivity, price performance and profitability measures over time. This allowed us to express the unit-specific profitability of any firm as a function of the profitability growth of the benchmark firm and actual catch-up to the benchmark firm. This was further decomposed into the productivity and price performance catch-up of each firm relative to the benchmark firm and the productivity and price performance of the benchmark firm. This decomposition allowed us to take into account the contribution of both profitability growth achieved by benchmark firms, as well as the contribution of profitability catch-up by less productive firms.

Since substantial improvements in quality have affected the productivity and price performance of the water industry, unit-specific profitability change was also expressed as a function of the unit-specific quality adjusted productivity and quality-adjusted price performance change. This was further decomposed as a function of the quality adjusted catch-up in productivity, and the quality adjusted productivity growth of the benchmark firm, and the quality adjusted catch-up in price performance, and the quality adjusted price performance growth of the benchmark firm. The inclusion of quality in our analysis allowed us to eventually decompose unit-specific economic profitability change as a function of the quality unadjusted catch-up in productivity, the catch-up in quality, and the quality-unadjusted productivity and quality performance over time of the benchmark firm, and the quality unadjusted catch-up in price performance, the catch-up in quality regarding price performance, and the price performance and quality growth of the benchmark firm, in a binary and multilateral context.

The results indicated that during the years 1991-2008, on average there was a stable increase in TFP, while TPP followed an upward trend until 1994, due to the lax price caps set at privatization, but was interrupted in 1995 due to the tightened 1994/95 price review and was followed by a substantial increase between 1999 and 2000. After 2000, average TPP and economic profitability substantially declined due to the tightened 1999/00 price review and followed a downward trend except for the years 2002 and 2006, while average productivity increased steadily. Thus, after 2001, average economic profitability followed the trend of TPP, indicating that changes in price performance were the main driver for changes in economic profitability. Moreover, after 2001, the steady decline in average price performance, gains in TFP and relatively stable economic profitability suggested that Ofwat was more focused on passing productivity benefits to consumers, and maintaining stable profitability than in the earlier regulatory periods.

Average economic profitability exceeded benchmark economic profitability during the years 1991-1994 and 1998-2008, showing high levels of catch-up relative to benchmark economic profitability after 2001 which was mainly explained by the relative decline in the economic profitability of the benchmark firm. Benchmark economic profitability significant declined in 2001 and despite an improvement in 2002 and in 2006, it declined below benchmark 1991 levels. In contrast, while average economic profitability was also considerably lower after 2000, it had never declined below average 1991 levels.

Focusing on the quality unadjusted productivity performance of the less productive and benchmark firms, we concluded that until 1995 average and benchmark firms did not have strong incentives to achieve high productivity levels This was interrupted after 1995, when price caps became tightened providing evidence that less productive firms had stronger incentives to catch-up to benchmark, while the benchmark company was also incentivized to continue to improve its productivity. Significant productivity gains for the average firm relative to the benchmark also continued after 2000. Thus, our results suggested that the implementation of even tighter price caps in 1999 further encouraged less productive firms to improve their performance relative to the benchmark, even though the benchmark firm continued to improve its performance. During the last price review period, average productivity growth again substantially exceeded the productivity growth of the benchmark firm, resulting in high levels of productivity catch-up between 2005 and 2008, although this was largely explained by substantial declines in benchmark productivity after 2006. Our results also suggested that all of this catch-up can be attributed to the post 1995 period, after Ofwat first tightened price caps, and most of it can be attributed to the post 2000 period, following the even more stringent 1999 price review.

Moreover, looking at the average and benchmark quality unadjusted price performance we concluded that average TPP exceeded benchmark TPP until 1995, however, by 2000, there was a convergence in average and benchmark TPP. During the years 2001-2004, there was little or no difference between average and benchmark TPP and during the years 2005-2008 average TPP exceeded benchmark TPP showing the highest levels of price performance catch-up in 2006 and in 2008. Our results suggested that in the post 1999 price review period, the price performance of all firms was substantially lower than in the first 10 years after privatisation.

Turning our discussion now to the quality adjusted results for productivity and price performance changes, we concluded that while quality improvements have contributed to the productivity performance of the WaSCs, they have also contributed negatively to their price performance. The quality adjusted TFP results indicated that although average productivity slightly exceeded benchmark productivity until 1995, the rate of quality adjusted productivity growth for the average and benchmark firms was significantly greater than the quality unadjusted TFP indicating that quality improvements did lead to higher productivity growths. After 1997 and until 2002, average quality adjusted TFP increased more rapidly than benchmark quality adjusted TFP, therefore allowing average company to catch-up to benchmark quality adjusted productivity. Even after 2002 the average company achieved still significant levels of catch-up in quality adjusted productivity until 2005, which must be attributed to input usage reductions. Furthermore, the considerable increase in average profitability relative to the benchmark firm must be attributed to this catch up effect. Nevertheless, after 2005, when the relatively looser 2004 price review came into effect, high levels of productivity catch-up were no longer indicative of general productivity improvements, as average quality adjusted productivity levels were largely static after 2005. Instead, they reflected a substantial decline in the benchmark firm's productivity after 2006. Thus, our results may be interpreted as suggesting that after the 2004 price review, substantial productivity improvements were no longer occurring. Furthermore, focusing on the results for the average and benchmark quality growth with respect to productivity we concluded that until 1997 there were small gains in average quality relative to benchmark quality but after 1998 average quality substantially exceeded benchmark quality showing high levels of catch-up during the years 2000-2005. By 2005 the less productive firms on average improved significantly their quality relative to the benchmark which already had high levels of quality since privatization.

Moreover, the quality adjusted TPP results suggested that until 1994, average TPP exceeded benchmark TPP but after 1998, there was a steady erosion of average price performance relative to benchmark price performance suggesting that there was a considerable rebalancing of regulatory price decisions in favour of the benchmark firm, which was even more dramatically extended with the implementation of the 1999 price review in 2001. The dramatic fall in both average and benchmark quality adjusted TPP suggested that that both the 1994 and 1999 price reviews resulted in considerable movement to a regulatory price cap system consistent with a yardstick regulation regime. We would moreover offer the suggestion that this better alignment of regulated prices with the principles of yardstick regulation is likely to have contributed significantly to both the catch-up in quality adjusted productivity and the catch up in economic profitability. Further, considering the post 2001 period revealed a steady downward trend in quality adjusted average and benchmark TPP except for the years 2002 and 2006. This overall finding supported a steady deterioration in price performance, which suggested that in practice, price caps have become even tighter since 2001. Also after 2001 average quality adjusted TPP fell more than benchmark quality adjusted TPP suggesting that the broad convergence after 2000 between average and benchmark firm price performance which was observed in the quality unadjusted TPP results was no longer present.

Overall, our index number based approach provided a backward-looking approach with respect to the impact of price cap regulation on the profitability, productivity and price performance of less productive and benchmark firms. It allowed us to calculate unit-specific profitability, TFP and TPP change and provide spatially consistent measurement of changes in these performance measures relative to other firms even if the number of available observations was extremely limited. We strongly believe that our methodology can be further used to aid regulators in setting X-factors under price cap regulation for regulated firms (forward-looking). Since X-factor requires the measurement of efficiency change (catch-up) and frontier shift (technical change), our approach provides evidence for catch-up (efficiency) in productivity by less productive firms based on the consistent spatial productivity measures across companies at any given year and also provides evidence for the productivity growth of the benchmark firm (technical change).

CHAPTER 7 DEA BASED PROFIT DECOMPOSITION

7.1 Introduction

In Chapter 5 we used a cross sectional technique to measure spatial productivity, regulatory price performance, and profitability, whereas in Chapter 6, we reconciled both cross sectional and temporal index numbers to measure relative, firm-specific and relative productivity change over time. In Chapter 7, we employ both index number techniques and Data Envelopment Analysis (DEA) methods to provide a link between productivity and financial performance of the Water and Sewerage Companies (WaSCs). There are also other determinants except for prices and productivity that can explain the changes in profits and can be useful for industry regulators and managers for performance evaluation and effectiveness of price cap scheme such as the activity, resource mix, product mix and scale effect.

There were several studies in the past that decomposed profit changes into three sources: a productivity change effect, an activity effect and a price change effect. Grifell-Tatje & Lovell (1999) provided a three-stage output oriented long-run profit decomposition to indentify the sources of profit change within the Spanish banking sector. The authors used Laspeyers and Paasche indicators to decompose economic profits into a quantity and price effect and linear programming methods to measure technical change, efficiency change, resource mix, product mix and scale effect. Also, De Witte & Saal (2009) employed Laspeyers and Paasche indicators and Free Disposal Hull (FDH) techniques to implement an input oriented instead three-stage profit decomposition for the Dutch regulated water industry. Moreover, Lim and Lovell (2006b) provided an output oriented short-run profit decomposition by taking into account the impact of quasi-fixed inputs and applied their decomposition to US Railroads for the period 1996-2003. In another study, Grifell-Tatje and Lovell (2008) provided another type of profit decomposition to measure productivity and price changes in US post offices. The authors decomposed profits into a quantity, margin and productivity effect by using Bennet indicators and then the productivity effect was further decomposed into a cost efficiency, technical change and scale effect. Finally, Sahoo & Tone (2009) employed both radial and non-radial DEA methods and both Laspeyers & Paasche and Bennet indicators, as weights, to value the contributions of various profit determinants on the Indian commercial banking sector.

However, none of the above studies include any exogenous factors in the profit decomposition analysis. Especially, since the UK water and sewerage industry is characterised by high capital investment programs to improve drinking water quality and environmental standards and past research has demonstrated that water and sewerage quality do significantly impact productivity and price performance measures across firms and over time (see Saal & Parker, 2001 and Maziotis, Saal and Thanassoulis, 2009), the inclusion of quality in a profit decomposition analysis is therefore important.

The purpose of this chapter is the evaluation of various profit drivers such as price changes, productivity changes and activity levels on the financial performance of the Water and Sewerage Companies (WaSCs) over time in the case when the number of observations is limited. In order to achieve this, we firstly, follow the approach of De Witte & Saal (2009) and decompose profits into a quantity and price effect using Bennet indicators to weigh the changes in quantities and prices and then we employ DEA techniques to take into account the impact of efficiency change, technical change and scale effect on profit changes. Secondly and more significantly, we extend Grifell and Lovell's (1999) approach by accounting for differences in output characteristics like water and sewerage quality in the profit decomposition analysis. Thirdly, as in previous studies (see Grifell and Lovell (1999)), our sequential DEA technique allows measurement of the productivity and the activity effect and their components where the number of observations is extremely limited. Finally, we provide a comparison of results from the profit decompositions without and after controlling for quality on Water and Sewerage Companies (WaSCs) in England and Wales over the period 1991-2008. The results demonstrate minor differences when we do not control for differences in output quality but in both cases, the policy implications for the UK water and sewerage industry are significant.

This chapter unfolds as follows. Section 7.2 discusses the concept of distance functions. It includes an analysis of the decomposition of profits into its components and the Data Envelopment Analysis (DEA) technique in order to estimate the components of the profit decomposition without and with adjustments for quality. The following section presents the data that are used in our study followed by a discussion of empirical results. The last section concludes.

7.2 *Methodology*

7.2.1 Distance Functions

We define the production technology at each period t as the set that includes all feasible output - inputs correspondences. The inputs are represented by a positive input quantity vector $X = (X_1, X_2, ..., X_N)$ where N denotes the total number of inputs that a company uses in order to produce a vector of non-negative outputs $Y = (Y_1, Y_2, ..., Y_M)$ where M denotes the total number of outputs. Let us assume that we have a positive vector of input prices $W = (W_1, W_2, ..., W_N)$ and a positive vector of output prices $P = (P_1, P_2, ..., P_M)$. The *production technology* or *production possibility* set for period t is then represented as:

$$S^t = \{(X^t, Y^t) : X^t \text{ can produce } Y^t \}, \qquad \text{where } t = 1,2,....,T \tag{7.1}$$

Let also the input set, $L^t(Y^t)$, represent the set of all input vectors that can produce a given output vector at period t, Y^t:

$$L^t(Y^t) = \{X^t : X^t \text{ can produce } Y^t\} = \{X^t : (X^t, Y^t) \in S^t\} \tag{7.2}$$

The input set is assumed to be closed and convex and satisfying strong disposability of inputs. Strong disposability of inputs means excess inputs can be disposed at no cost. The lower bound of an input set is the input isoquant given by:

$$I^t(Y^t) = \{X^t : X^t \in L^t(Y^t), \lambda X^t \notin L^t(Y^t), \lambda \prec 1\} \tag{7.3}$$

Shephard (1970) introduced the input distance function to provide a functional representation of production technology. The input distance function defined as a minimal proportional reduction of the input vector given an output vector at each period t is given by:

$$D_I^t(Y^t, X^t) = \max\{\mu : (X^t / \mu) \in L^t(Y^t)\} \tag{7.4}$$

For $X^t \in L^t(Y^t), D_I^t(Y^t,X^t) \geq 1$ and for $X^t \in I^t(Y^t), D_I^t(Y^t,X^t) = 1$.

Let us also define the output set, $O^t(X^t)$, which represents the set of all output vectors, Y^t, that can be produced using the input vector, X^t in period t:

$$O^t(X^t) = \{Y^t : X^t \text{ can produce } Y^t\} = \{Y^t : (X^t, Y^t) \in S^t\}, \quad \text{where } t = 1,2,....,T \qquad (7.5)$$

The output set is assumed to be closed and convex and satisfy strong disposability of outputs and inputs. The outer bound of an output set is its output isoquant:

$$I^t(X^t) = \{Y^t : Y^t \in O^t(X^t), \lambda Y^t \notin O^t(X^t), \lambda \succ 1\} \qquad (7.6)$$

Shephard's (1970) output distance function provides another functional representation of production technology. The output distance function defined as a maximal proportional expansion of the output vector given an input vector at each period t is given by:

$$D_o^t(Y^t, X^t) = \min\{\delta : (Y^t / \delta) \in O^t(X^t)\} \qquad (7.7)$$

For $Y^t \in O^t(Y^t), D_O^t(Y^t,X^t) \leq 1$ and for $Y^t \in I^t(X^t), D_O^t(Y^t,X^t) = 1$. The distance functions, being radial distance measures, provide the tools with which we will recover the unobserved quantity vectors that we need for the profit decomposition.

7.2.2 Profit Decomposition Without Controlling for Quality

In this section we follow De Witte and Saal's (2009) approach and provide an input oriented profit decomposition between two time periods t and t+1 using Bennet indicators, average prices and quantities as weights to estimate the contributions of the quantity and price effect to profit change. Let a company's profit in period t, Π^t, be defined as a difference between its total revenues and total costs, $\Pi^t = P^t Y^t - W^t X^t$. Using Bennet indicators, $\overline{P} = 1/2(P^{t+1} + P^t)$,

$\overline{W} = 1/2\left(W^{t+1} + W^{t}\right)$, $\overline{X} = 1/2\left(X^{t+1} + X^{t}\right)$, $\overline{Y} = 1/2\left(Y^{t+1} + Y^{t}\right)$ profit change between period t and t+1, $\Pi^{t+1} - \Pi^{t}$, is decomposed as follows:

$$\begin{aligned} \Pi^{t+1} - \Pi^{t} &= \overline{P}\left(Y^{t+1} - Y^{t}\right) - \overline{W}\left(X^{t+1} - X^{t}\right) && \textit{quantity effect} \\ &+ \overline{Y}\left(P^{t+1} - P^{t}\right) - \overline{X}\left(W^{t+1} - W^{t}\right) && \textit{price effect} \end{aligned} \tag{7.8}$$

The quantity effect captures the contribution to profit changes from a change in output production and input usage, while the price effect shows the contribution to profit changes from a change in output and input prices. The quantity effect shows that profits may increase due to a rise in output production in excess of the corresponding input rise while the price effect shows that profits may also rise due to an increase in output prices in excess of the rise in input prices. The decomposition of profits into a quantity and price effect involves only observed quantity and price data.

In the second stage the quantity effect can be decomposed into a *productivity* and an *activity* effect as follows:

$$\begin{aligned} &\overline{P}\left(Y^{t+1} - Y^{t}\right) - \overline{W}\left(X^{t+1} - X^{t}\right) && \textit{quantity effect} \\ &= \left[\overline{W}\left(X^{t} - X^{B}\right) - \overline{W}\left(X^{t+1} - X^{C}\right)\right] && \textit{productivity effect} \\ &+ \left[\overline{P}\left(Y^{t+1} - Y^{t}\right) - \overline{W}\left(X^{C} - X^{B}\right)\right] && \textit{activity effect} \end{aligned} \tag{7.9}$$

This decomposition is depicted in Figure 7.1. $I^{t}\left(Y^{t}\right)$ represents the efficient input boundary, that is the locus of minimum input levels needed to produce a given level of output Y^{t} in period t. The quantity effect as decomposed in (7.9) makes use of the observed quantities X^{t} to X^{t+1} and of the unobserved quantities$\left(X^{A}, X^{B}, X^{C}\right)$. As can be seen in Figure 7.1, X^{A} and X^{B} denote the efficient input level that the unit could have used in period t and period t+1 respectively to secure out Y^{t} keeping to the input mix of X^{t}, while X^{C} represents the efficient input level that the unit could have used in period t+1 to secure out Y^{t+1} keeping to the input mix of X^{t+1}.

The productivity effect in (7.9) compares the distance from X^B to X^t in period t with the distance from X^C to X^{t+1} in period t+1. The difference in these two distances reflects productivity change of the unit as it captures how much closer or further from the 'fixed' efficient boundary of period t+1 the unit has moved over time. When we have $\left(X^t - X^B\right) > \left(X^{t+1} - X^C\right)$ we have a positive contribution to profit change, whereas when we have $\left(X^t - X^B\right) < \left(X^{t+1} - X^C\right)$ we have a negative contribution to profit change.

The activity effect in (7.9) measures the changes in the scale and scope of the activities of a company. When $\left(Y^{t+1} - Y^t\right)$ is positive it reflects a rise in output over time while $\left(X^C - X^B\right)$ when negative reflects a fall in the efficient level of input needed to secure the output. Thus both the output and the input differences in this case respectively lead to positive contributions to profit change between period t and t+1.

Finally in a third stage decomposition the productivity effect in (7.9) can be further decomposed into an *efficiency change* and *technical change* effect while the activity effect can be further decomposed into a *resource mix*, *output mix* and *scale* effect. Figures 7.1 and 7.2 depict the decomposition of the productivity and activity effect, which we now elaborate upon.

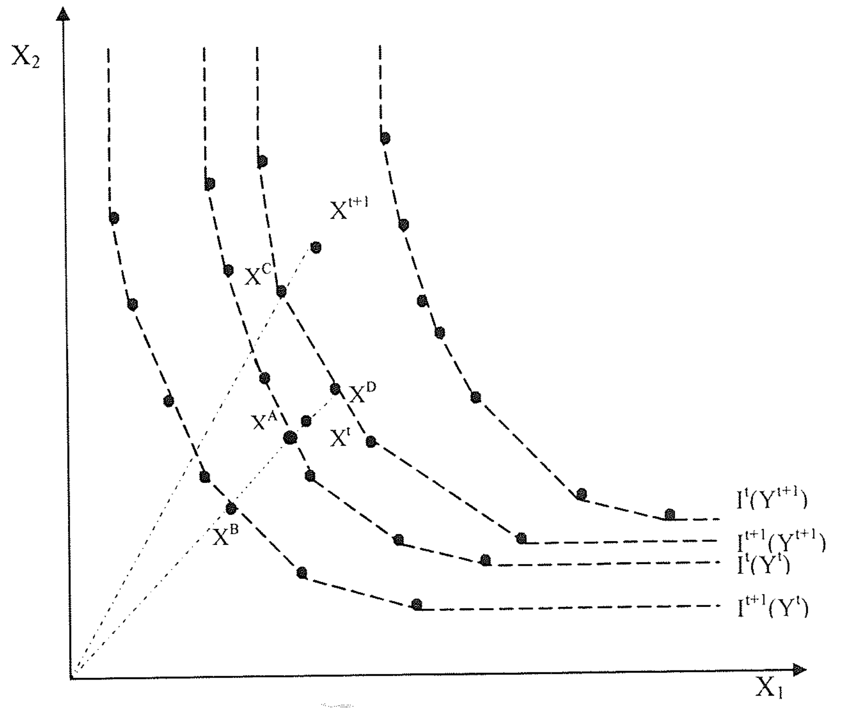


Figure 7-1 Productivity Effect

$$
\begin{aligned}
&\overline{W}\left[\left(X^{t} - X^{B}\right) - \left(X^{t+1} - X^{C}\right)\right] \quad \textit{productivity effect} \\
&= \overline{W}\left[\left(X^{A} - X^{B}\right)\right] \quad \textit{technical change} \\
&+ \overline{W}\left[\left(X^{t} - X^{A}\right) - \left(X^{t+1} - X^{C}\right)\right] \quad \textit{efficiency change}
\end{aligned}
\tag{7.10}
$$

Technical change is measured by the distance X^A to X^B. As can be seen in Figure 7.1 this difference reflects the distance between the efficient boundaries in periods t and t+1, controlling for output level. Technical improvement occurs when, $X^B < X^A$. Such an improvement in the efficient boundary from t to t+1 has a positive effect on profit change from t to t+1, whereas with technical regress, $X^B > X^A$, and there will be a negative impact on profit change.

Moving to the efficiency change term in (7.10) we note that the distance from X^A to X^t reflects the inefficiency of the firm in period t and similarly the distance from X^C to X^{t+1} reflects the inefficiency of the firm in period t+1. Thus, as illustrated in (7.10) a decline in the input price weighted cost of inefficiency in period t+1, relative to the equivalent cost in period t, has a positive impact on profit change. In contrast, a rise in the input price weighted cost of inefficiency in period t+1 relative to that in period t would have a negative impact on profit change.

The activity effect in (7.9) can be further decomposed as follows:

$$
\begin{aligned}
&\overline{P}\left(Y^{t+1} - Y^{t}\right) - \overline{W}\left(X^{C} - X^{B}\right) \quad \textit{activity effect} \\
&= \overline{W}\left(X^{D} - X^{C}\right) \quad \textit{resource mix effect} \\
&- \overline{P}\left(Y^{E} - Y^{t+1}\right) \quad \textit{product mix effect} \\
&+ \overline{W}\left(X^{B} - X^{D}\right) - \overline{P}\left(Y^{t} - Y^{E}\right) \quad \textit{scale effect}
\end{aligned}
\tag{7.11}
$$

The resource mix effect $X^D - X^C$ captures the impact on profits due to the change in the mix of inputs between period t and t+1 while keeping the output at the period t+1 level and also retaining efficiency in production (see Figure 7.1). When

$X^D - X^C$ is positive, the change in resource mix reflects a movement of input usage to one which reduces costs, thereby improving allocative efficiency. Similarly, we can infer from Figure 7.2 the product mix effect as the change in output mix from Y^E to Y^{t+1}. Note that Y^E reflects the output mix of period t but its level is that resulting from using the efficient input level X^D in period t+1 to secure the output mix of period t.

Finally the scale effect consists of two components, the input scale effect and the output scale effect, thereby capturing the impact of scale change on the firm's profitability. From Figure 7.1, we note that to produce efficiently the output of period t, Y^t using the best practice technology available in period t+1, the input level needed is X^B. In contrast when outputs change from Y^t to Y^{t+1}, while keeping the input mix and the technology constant to that of period t+1 the input required is X^D. The difference between X^B and X^D when positive means that efficient input level needed in constant technology has dropped as output changed from period t to t+1 and this has a positive impact on profit. As X^B and X^D have the same mix their difference simply reflect the difference in their scale size. In a similar manner, Y^t and Y^E have the same mix as can be seen in Figure 7.2 and their difference reflects the difference in their scale size. Y^t and Y^E are efficient output levels on t+1 technology using respectively input levels X^B and X^D already defined in Figure 7.1.

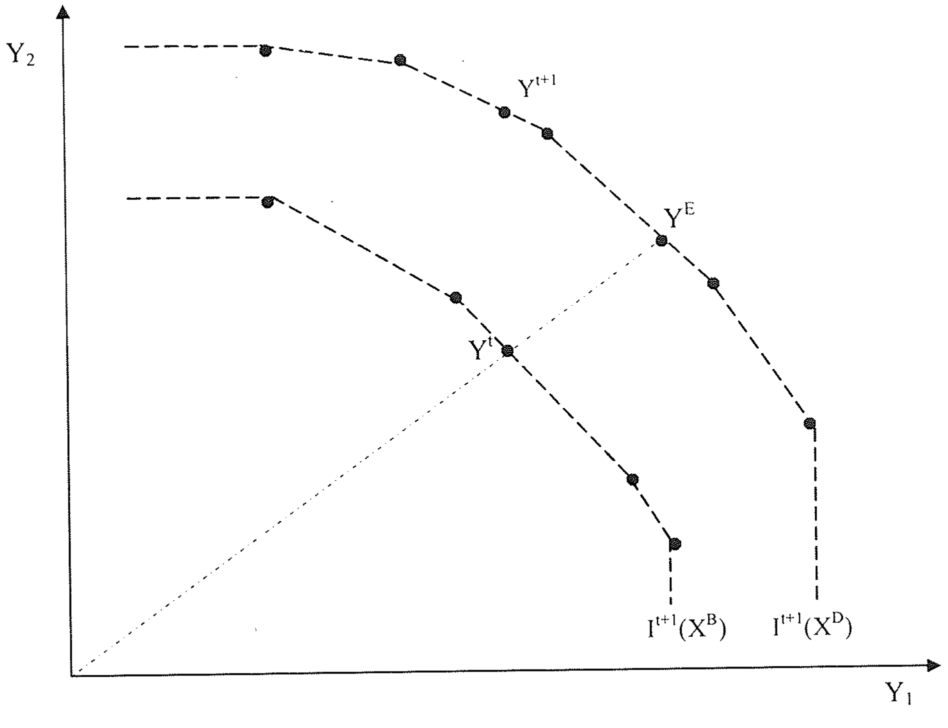


Figure 7-2 Activity Effect

7.2.3 Estimation With DEA

The second and third stage of the above profit decomposition requires the computation of the unobserved quantities $\left(X^A, X^B, X^C, X^D, Y^E\right)$. These unobserved input and output quantities can be estimated by means of the input and output distance functions as follows:

$$X^A = X^t * D_I^t\left(Y^t, X^t\right) \quad X^B = X^t * D_I^{t+1}\left(Y^t, X^t\right) \quad X^C = X^{t+1} * D_I^{t+1}\left(Y^{t+1}, X^{t+1}\right) \tag{7.12}$$
$$X^D = X^t * D_I^{t+1}\left(Y^{t+1}, X^t\right) \qquad Y^E = Y^t * D_O^{t+1}\left(\left(X^D, Y^t\right)\right)$$

The required distances and hence the quantities $\left(X^A, X^B, X^C, X^D, Y^E\right)$ as defined in (7.12) can be readily estimated using DEA. Let J, N, M and T denote, respectively, the total number of firms, inputs, outputs and time periods in the sample. Let ϕ denote a scalar, which represents the proportional contraction of the input vector, given the output vector and θ denote a scalar, which represents the proportional expansion of output vector, given the input vector. Let Y_j^t and X_j^t denote the $M \times 1$ output vector and the $N \times 1$ input vector respectively for the j-th firm in the t-th period $t = 1,2,...,T$. Let y^t and x^t denote respectively the $M \times J$ output matrix and the $N \times J$ input matrix in period t, containing the data for all the firms in the t-th period. The notation for period t+1 is defined similarly. We use the additional constraint $J1'\lambda = 1$ to allow for variable returns to scale technology. The reference technology for our DEA models is the sequential DEA technology which is defined in section 7.3. Sequential technology assumes that in any period t the technology of the previous periods remains feasible. By definition this technology does not allow for regress. Thus in period t the unobserved quantity X^A can be computed by the following linear programming problem:

$$\begin{aligned}
&\left[D_I^t\left(Y^t, X^t\right)\right]^{-1} = \phi^A = Min\,\phi \\
&\textit{subject to} \\
&Y_j^t \le \sum_{k=1}^{t} \sum_{j=1}^{J} y_j^t \lambda_j^t \\
&\phi\ X_j^t \ge \sum_{k=1}^{t} \sum_{j=1}^{J} x_j^t \lambda_j^t \\
&\lambda \ge 0 \\
&J1'\lambda = 1
\end{aligned} \qquad (7.13)$$

The variables $\lambda^k = \left(\lambda_1^k, \lambda_2^k, ..., \lambda_J^k\right)$ k=1…t whose optimal values are to be determined by the above model lead to the estimate the proportional reduction ϕ^A in X^t that would locate (X^t, Y^t) on the efficient frontier within the sequential, technology to period t. The unobserved quantity X^A for the firm having input output set (X^t, Y^t) is thus $X^A = \phi^A X^t$. The unobserved quantity X^A is computed as $X^A = \phi^A X^t$ for each firm in the sample in period t.

The unobserved quantity X^B can be computed by solving the following linear programming problem:

$$\left[D_I^{t+1}\left(Y^t, X^t\right)\right]^{-1} = \phi^B = Min\,\phi$$

subject to

$$\begin{gathered} Y_j^t \le \sum_{k=1}^{t+1}\sum_{j=1}^{J} y_j^{t+1}\lambda_j^{t+1} \\ \phi\; X_j^t \ge \sum_{k=1}^{t+1}\sum_{j=1}^{J} x_j^{t+1}\lambda_j^{t+1} \\ \lambda \ge 0 \\ J1'\lambda = 1 \end{gathered} \tag{7.14}$$

The unobserved quantity X^B is computed as $X^B = \phi^B X^t$ for each firm (X^t, Y^t) in the sample in period t.

The unobserved quantity X^C can be computed using the following linear programming problem:

$$\left[D_I^{t+1}\left(Y^{t+1}, X^{t+1}\right)\right]^{-1} = \phi^C = Min\,\phi$$

subject to

$$\begin{gathered} Y_j^{t+1} \le \sum_{k=1}^{t+1}\sum_{j=1}^{J} y_j^{t+1}\lambda_j^{t+1} \\ \phi X^{t+1} \ge \sum_{k=1}^{t+1}\sum_{j=1}^{J} x_j^{t+1}\lambda_j^{t+1} \\ \lambda \ge 0 \\ J1'\lambda = 1 \end{gathered} \tag{7.15}$$

The unobserved quantity X^C is computed as $X^C = \phi^C X^{t+1}$ for each firm (X^{t+1}, Y^{t+1}) in the sample in period t.

The unobserved quantity X^D can be computed by the following linear programming problem:

$$\left[D_I^{t+1}\left(Y^{t+1}, X^t\right)\right]^{-1} = \varphi^D = Min\,\varphi$$

subject to

$$\begin{gathered} Y_j^{t+1} \le \sum_{k=1}^{t+1}\sum_{j=1}^{J} y_j^{t+1}\lambda_j^{t+1} \\ \varphi X^t \ge \sum_{k=1}^{t+1}\sum_{j=1}^{J} x_j^{t+1}\lambda_j^{t+1} \\ \lambda \ge 0 \\ J1'\lambda = 1 \end{gathered} \tag{7.16}$$

The unobserved quantity X^D is computed as $X^D = \varphi^D X^t$ for each firm (X^t, Y^{t+1}) in the sample in period t.

The unobserved quantity Y^E can be computed using (7.16) and the following linear programming problem:

$$\begin{aligned}
&\left[D_o^{t+1}\left(Y^t, X^D\right)\right]^{-1} = \theta^E = Max\,\theta \\
&\textit{subject to} \\
&\theta\ Y_j^t \le \sum_{k=1}^{t+1}\sum_{j=1}^{J} y_j^{t+1}\lambda_j^{t+1} \\
&X_j^D \ge \sum_{k=1}^{t+1}\sum_{j=1}^{J} x_j^{t+1}\lambda_j^{t+1} \\
&\lambda \ge 0 \\
&J1'\lambda = 1
\end{aligned} \qquad (7.17)$$

The unobserved quantity Y^E is computed as $Y^E = \theta^E Y^t$ for each firm (X^D, Y^t) in the sample in period t.

7.2.4 Profit Decomposition After Controlling for Quality

Since the water and sewerage companies have carried out substantial capital investment projects to improve drinking water quality and environmental standards, it is important to include the impact of quality in a profit decomposition approach. As the substantial drinking water quality and sewerage treatment improvements over the 1991-2008 period (Maziotis, Saal and Thanassoulis, 2009) have been in response to increasingly stringent environmental regulation, including EU directives, it is reasonable to assume that quality improvements are exogenously determined (Saal and Parker, 2000). Therefore, quality is included in a profit decomposition approach as an exogenous factor and is intended to control for changes over the assesment period in water quality, environmental standards and characteristics that reflect differences between firms in terms of their operating environment (Stone & Webster Consultants, 2004). However, in more general contexts where regulation is not so tight it is possible for quality to be seen as a discretionary variable. Differences in output quality between firms may result in legitimate differences in required inputs to produce a given quantity of output. Moreover, variation in measured profitability, productivity and activity effects may result partially from these differences. This section therefore presents a profit decomposition approach which makes allowances for differences in output characteristics such as output quality between firms and across time, thereby extending Grifell-Tatje and Lovell's (1999) approach.

As earlier the inputs are represented by a positive input quantity vector $X = (X_1, X_2, ..., X_N)$ where N denotes the total number of resources and the positive vector of input prices can be defined as $W = (W_1, W_2, ..., W_N)$. However, the positive vector of output quantities $Y = (Y_1, Y_2, ..., Y_M)$ where M denotes the total number of outputs is now separated into a non-negative vector of output for high quality $Y_h = (Y_{1,h}, Y_{2,h}, ..., Y_{M,H})$ and a non-negative vector of output for low quality $Y_l = (Y_{1,l}, Y_{2,l}, ..., Y_{M,L})$ where H and L denotes the total number of outputs for high and low quality respectively and we assume that $Y = Y_h + Y_l$ and that more inputs are required to produce a given amount of high quality output than to produce the same amount of low quality output. The positive vector of output prices $P = (P_1, P_2, ..., P_M)$is similarly separated into a positive vector of output prices for high quality $P_h = (P_{1,h}, P_{2,h}, ..., P_{M,H})$ and a positive vector of output prices for low quality $P_l = (P_{1,l}, P_{2,l}, ..., P_{M,L})$ to reflect differences in output prices for quality between firms.

Therefore, given the assumptions that $Y = Y_h + Y_l$ and the output prices P_h and P_l the decomposition of profits into a quantity and price effect in equation (7.8) will become equation (7.8') as follows, using Bennet indicators, $\overline{P}_h = 1/2\left(P_h^{t+1} + P_h^t\right)$, $\overline{P}_l = 1/2\left(P_l^{t+1} + P_l^t\right)$, $\overline{W} = 1/2\left(W^{t+1} + W^t\right)$, $\overline{X} = 1/2\left(X^{t+1} + X^t\right)$, $\overline{Y}_h = 1/2\left(Y_h^{t+1} + Y_h^t\right)$, $\overline{Y}_l = 1/2\left(Y_l^{t+1} + Y_l^t\right)$:

$$\begin{aligned} \Pi^{t+1} - \Pi^t &= [\overline{P}_h\left(Y_h^{t+1} - Y_h^t\right) + \overline{P}_l\left(Y_l^{t+1} - Y_l^t\right)] - \overline{W}\left(X^{t+1} - X^t\right) \quad \textit{quantity effect} \\ &+ [\overline{Y}_h\left(P_h^{t+1} - P_h^t\right) + \overline{Y}_l\left(P_l^{t+1} - P_l^t\right)] - \overline{X}\left(W^{t+1} - W^t\right) \quad \textit{price effect} \end{aligned} \tag{7.8'}$$

The difference between equations (7.8) and (7.8') is in the output effect of the quantity effect and the output weights. Similarly in the price effect the output component and corresponding weights change. The input effect components remain the same between (7.8) and (7.8') as they are calculated using observed input quantities and input prices which have not changed. The quantity effect will now capture the contribution to profit changes from a change in output production of high and low quality and input usage, while the price effect will show the contribution to profit changes from a change in output prices for high and low quality and input prices.

Given that $Y = Y_h + Y_l$, and the output prices P_h, P_l the decomposition of the quantity effect into productivity and activity effect in equation (7.9) will now become equation (7.9').

$$[\overline{P}_h(Y_h^{t+1} - Y_h^t) + \overline{P}_l(Y_l^{t+1} - Y_l^t)] - \overline{W}(X^{t+1} - X^t) \textit{quantity effect}$$

$$= [\overline{W}(X^t - X^{B'}) - \overline{W}(X^{t+1} - X^{C'})] \quad \textit{productivity effect} \qquad (7.9')$$

$$+ [[\overline{P}_h(Y_h^{t+1} - Y_h^t) + \overline{P}_l(Y_l^{t+1} - Y_l^t)] - \overline{W}(X^{C'} - X^{B'})] \quad \textit{activity effect}$$

The productivity effect is now calculated using observed input quantities and input prices, $(X^t, X^{t+1}, W^t, W^{t+1})$ and unobserved quantities $(X^{B'}, X^{C'})$. The results for the productivity effect will now be different from those obtained in equation (7.9) because the recovery of the two unobserved input quantities $(X^{B'}, X^{C'})$ in equation (7.9') requires the inclusion of two output vectors, Y_h and Y_l instead of the aggregate one, Y, in the linear programming models (7.14) and (7.15). The quantities $X^{B'} = X^t * \phi^B$ and $X^{C'} = X^{t+1} * \phi^C$ where ϕ^B and ϕ^C are now optimal values as derived from models (7.14) and (7.15) respectively after substituting the two sets of output constraints (high and low quality) for the aggregate output set.

The output side of the activity effect in equation (7.9') now changes since it is calculated using the observed output quantities and prices, Y_h, Y_l, P_h, P_l. The activity effect now also reflects changes in output between high and low quality and the efficient level of input needed to secure the output changes. The results for the activity effect in equation (7.9') will differ from those calculated in equation (7.9) since it uses the unobserved input quantities $(X^{B'}, X^{C'})$ where estimates with DEA will need to include two output vectors, Y_h and Y_l instead of the aggregate vector Y, in the linear programming models (7.14) and (7.15).

The decomposition of the productivity effect into technical change and efficiency change in equation (7.10') is calculated using observed input quantities and input prices, $(X^t, X^{t+1}, W^t, W^{t+1})$ and unobserved quantities $(X^{A'}, X^{B'}, X^{C'})$. The results for technical change and efficiency change will therefore generally be different from those in equation (7.10) because the recovery of the three unobserved input quantities $\left(X^{A'}, X^{B'}, X^{C'}\right)$ in equation (7.10') requires the inclusion of two outputs, Y_h and Y_l instead of their sum, in the linear programming models (7.13) to (7.15), where $X^{A'} = X^t * \phi^A$ where ϕ^A is the optimal value of ϕ as derived from models (7.13) after substituting the two sets of output constraints (high and low quality) for the aggregate output set.

$$\begin{array}{ll} \overline{W}\left[\left(X^t - X^{B'}\right) - \left(X^{t+1} - X^{C'}\right)\right] & \textit{productivity effect} \\ = \overline{W}\left[\left(X^{A'} - X^{B'}\right)\right] & \textit{technical change} \\ + \overline{W}\left[\left(X^t - X^{A'}\right) - \left(X^{t+1} - X^{C'}\right)\right] & \textit{efficiency change} \end{array} \tag{7.10'}$$

Given that $Y = Y_h + Y_l$, and the prices P_h and P_l the decomposition of the activity effect into resource mix, product mix and scale effect in equation (7.11) will now become equation (7.11') as follows:

$$\begin{array}{ll} [\overline{P}_h\left(Y_h^{t+1} - Y_h^t\right) + \overline{P}_l\left(Y_l^{t+1} - Y_l^t\right)] - \overline{W}\left(X^{C'} - X^{B'}\right) & \textit{activity effect} \\ = \overline{W}\left(X^{D'} - X^{C'}\right) & \textit{resource mix effect} \\ -[\overline{P}_h\left(Y_h^{E'} - Y_h^{t+1}\right) + \overline{P}_l\left(Y_l^{E'} - Y_l^{t+1}\right)] & \textit{product mix effect} \\ + \overline{W}\left(X^{B'} - X^{D'}\right) - [\overline{P}_h\left(Y_h^t - Y_h^{E'}\right) + \overline{P}_l\left(Y_l^t - Y_l^{E'}\right)] & \textit{scale effect} \end{array} \tag{7.11'}$$

The resource mix effect in equation (7.11'), it is calculated using observed input prices and unobserved quantities $\left(X^{D'}, X^{C'}\right)$. The product mix effect in equation (7.11') is calculated using observed output prices and output quantities, P_h, P_l, Y_h, Y_l and the unobserved quantity $Y^{E'}$ and reflects the changes in the output mix for high and low quality. The scale effect in equation (7.11') is calculated using observed input prices and unobserved quantities $\left(X^{B'}, X^{D'}\right)$ and observed output prices and output quantities, P_h, P_l, Y_h, Y_l and the unobserved quantity $Y^{E'}$. This component reflects changes in the mix of output for high and low quality given efficient input usage. The results for the resource mix, product mix and scale effect will generally differ from those obtained through equation (7.11) since the recovery of the three unobserved input and output quantities $\left(X^{D'}, X^{C'}, Y^{E'}\right)$ needs to include two output vectors, Y_h and Y_l instead of the aggregate vector Y, in the linear programming models (7.15) to (7.17) where $X^{D'} = \varphi^D X^t$ where ϕ^D is the optimal value of ϕ as derived from models (7.16) after substituting the two sets of output constraints (high and low quality) for the aggregate output set. Similarly $Y^{E'} = \theta^E Y^t$ for each firm $(X^{D'}, Y^t)$ in the sample in period t.

The above modifications in the profit decomposition with adjustments for quality, equations (7.8')-(7.11') can be readily implemented if data for multiple output quality levels is available. However, in the UK water industry, all customers of a given water firm effectively pay the same price for water services regardless of output quality, as regulated water prices do not differentiate between quality of output. Moreover, given this regulatory practice, it is unsurprising that while total turnover data is available separately for water and sewerage services it is not disaggregated by quality of service. As a result, we do not in practice have different prices for high and low quality water and sewerage output types, even though we can observe quantity data reflecting differences in output quality. Hence, given that regulatory practice results in no quality related price differentials for a given company, we necessarily and appropriately proceed with the assumption that consumers pay the same price for high and low quality outputs. Thus, in our application we observe that $P = P_h = P_l$. It should be noted that in the general case the production of higher quality output may require more input of each type than the production of the same quantity of output of lower quality. Further, additional input types may be needed for producing higher quality output that are not necessary for producing output of lower quality. For example, different facilities and chemicals are needed at different stages of sewerage treatment, primary, secondary or tertiary. Prices for the different types of resources used for output of different quality may also differ. These factors should be taken into account in the assessments being undertaken. Our own model implicitly allows for different levels of output quality requiring different levels of input but only for inputs that are common to high and low quality output. This is true by virtue of the fact that the DEA model sets the mix of outputs of high and low quality against the input bundle being used by each comparative unit. However, for the purpose of this study we make the assumption that no additional input types are needed for producing higher quality output and that prices of inputs are independent of the mix of output quality. This is consistent with previous studies of the UK water and sewerage industry by Saal & Parker (2000, 2001 and 2006) and Saal et, al (2007). However, in our empirical application in the linear programming models, we imposed the weight restriction that the production of high quality output is at least as resource intensive as the same quantity of output of low quality. We therefore modify our earlier notation to reflect this empirical characteristic of the English and Welsh water industry.

As the technology set includes the set of all feasible inputs and outputs adjusted for high and low quality, the input and output set, input and output isoquant and input and output distance functions are by definition equivalent to those employed in (7.8') to (7.11'). However, given the single output price, $P = P_h = P_l$, profits decompose into a quantity and price effect as follows, using Bennet indicators, $\overline{P} = 1/2\left(P^{t+1} + P^t\right)$, $\overline{W} = 1/2\left(W^{t+1} + W^t\right)$, $\overline{X} = 1/2\left(X^{t+1} + X^t\right)$, $\overline{Y}_h = 1/2\left(Y_h^{t+1} + Y_h^t\right)$, $\overline{Y}_l = 1/2\left(Y_l^{t+1} + Y_l^t\right)$:

$$\begin{aligned} \Pi^{t+1} - \Pi^t = {} & \overline{P}[\left(Y_h^{t+1} - Y_h^t\right) + \left(Y_l^{t+1} - Y_l^t\right)] - \overline{W}\left(X^{t+1} - X^t\right) \quad \textit{quantity effect} \\ & + [\overline{Y}_h + \overline{Y}_l]\left(P^{t+1} - P^t\right) - \overline{X}\left(W^{t+1} - W^t\right) \quad \textit{price effect} \end{aligned} \qquad (7.8'')$$

However, we need to underline that since $Y = Y_h + Y_l$ the results from the first stage of the profit decomposition in equation (7.8'') will be exactly the same as in equation (7.8) since the decomposition of profits into a quantity and price effect is calculated using observed quantities and prices. Thus, equation (7.8'') reveals that in the absence of differentiated prices for different output qualities, it is not possible to gain further information with regard to the overall quantity effect, even if we differentiate between different output qualities.

Nevertheless, the differentiation of output quantities by quality does allow an alternative decomposition of the aggregate quantity effect, which is arguably superior because estimated technologies and distances will better reflect how quality influences input requirements, even if output price weights do not differ for low and high quality outputs. Thus, given that $Y = Y_h + Y_l$, $P = P_h = P_l$, the decomposition of the quantity effect into the productivity and activity effects in equation (7.9') becomes:

$$
\begin{aligned}
&\overline{P}[(Y_h^{t+1} - Y_h^t) + (Y_l^{t+1} - Y_l^t)] - \overline{W}(X^{t+1} - X^t) && \textit{quantity effect} \\
&= [\overline{W}(X^t - X^{B'}) - \overline{W}(X^{t+1} - X^{C'})] && \textit{productivity effect} \\
&+ [\overline{P}[(Y_h^{t+1} - Y_h^t) + (Y_l^{t+1} - Y_l^t)] - \overline{W}(X^{C'} - X^{B'})] && \textit{activity effect}
\end{aligned}
\tag{7.9''}
$$

The difference between equations (7.9'') and (7.9') is in the weights used to evaluate the changes in the output side of the activity effect since it is now calculated using the observed output prices, P, instead of P_h, P_l. Thus, we first emphasize, that the aggregate productivity effect obtained from a model differentiating output qualities is theoretically identical, regardless of whether we control for differences in output prices. In contrast, while the input side of the activity effect is theoretically identical to that obtained in equation (7.9'), the reliance on quality undifferentiated output prices implies an alternative empirically observable weighting of the output side of the activity effect. We nonetheless argue that given the empirical reality that consumers do not face quality differentiated output prices, (7.9'') clearly provides a superior decomposition of the quantity effect when compared to (7.9). This is because disaggregation of the output, Y, into the output for high and low quality, Y_h and Y_l allows unbiased estimation of the underlying production frontiers and, hence the quantities $(X^{B'}, X^{C'})$ and therefore eliminates the reliance on the potentially biased frontiers in equation (7.9) where (X^B, X^C) are estimated with the quality undifferentiated output vector Y.

Consideration of (7.10') reveals that the equivalence of the productivity effect, regardless of whether we assume quality differentiated or undifferentiated output prices, extends to its decomposition into technical change and efficiency change. Thus, even with the assumption of undifferentiated output prices we can be confident that a quality based decomposition of the output vector reveals superior unbiased estimates of technical change and efficiency change, especially when compared to the estimates obtained from (7.10) which relies on a quality undifferentiated output vector to yield estimates of the required unobserved input quantities $\left(X^{A}, X^{B}, X^{C}\right)$. In contrast, the unobserved input quantities $\left(X^{A'}, X^{B'}, X^{C'}\right)$ employed in (7.10'), rely on the differentiated output vectors Y_h and Y_l, and therefore allow a better specification of the underlying technology required to estimate technical change and efficiency change. We would therefore emphasize that differences in the results for technical change and efficiency change between equations (7.10') and (7.10) will suggest that there is a bias in frontier estimation in equation (7.10) where the estimated technology for input usage includes only one output, Y.

Given the assumption of quality differentiated output quantities but quality undifferentiated output prices, the decomposition of the activity effect into resource mix, product mix and scale effect in equation (7.11') will become:

$$\begin{aligned}
&\overline{P}[\left(Y_h^{t+1} - Y_h^{t}\right) + \left(Y_l^{t+1} - Y_l^{t}\right)] - \overline{W}\left(X^{C'} - X^{B'}\right) \quad \textit{activity} \quad \textit{effect} \\
&= \overline{W}\left(X^{D'} - X^{C'}\right) \quad \textit{resource mix effect} \qquad\qquad (7.11'') \\
&- \overline{P}[\left(Y_h^{E'} - Y_h^{t+1}\right) + \left(Y_l^{E'} - Y_l^{t+1}\right)] \quad \textit{product mix effect} \\
&+ \overline{W}\left(X^{B'} - X^{D'}\right) - \overline{P}[\left(Y_h^{t} - Y_h^{E'}\right) + \left(Y_l^{t} - Y_l^{E'}\right)] \quad \textit{scale effect}
\end{aligned}$$

The difference between equations (7.11'') and (7.11') is on the weights employed to evaluate changes in the product mix effect and the output scale effect since they are now calculated using the observed output prices, P, instead of P_h, P_l. The scale effect now captures the change in the efficient output levels for high and low output quality given efficient input usage. Also, given that $P = P_h = P_l$ the product mix effect will not reflect changes in the mix of output for high and low quality but only changes in the aggregate non quality differentiated mix of outputs. Nevertheless, we particularly emphasize that the resource mix effect and the input scale effect will be exactly the same as in equation (7.11') because they are calculated using observed input prices and unobserved input quantities $\left(X^{D'}, X^{C'}, X^{B'}\right)$. Thus, the resource mix effect in particular is invariant to the assumption of quality undifferentiated output prices, in a model that allows for quality differentiated output quantities. As before, the unobserved quantities $\left(X^{B'},X^{C'},X^{D'},Y^{E'}\right)$ in equation (7.11'') are recovered from the observed quantity vectors (X^{t},Y_h^{t},Y_l^{t}) and $(X^{t+1},Y_h^{t+1},Y_l^{t+1})$ by means of input and output distance functions and the linear programming models in (7.14)-(7.17) will still include two outputs, Y_h and Y_l. Therefore, the results for the resource mix, product mix and scale effect in equation (7.11'') will be different from those yielded by equation (7.11). These differences will underline that the disaggregation of the output, Y, into high and low quality, Y_h and Y_l leads to differences in the estimated frontiers and so there is a bias in the estimated frontiers in equation (7.11) where $\left(X^{B},X^{C},X^{D},Y^{E}\right)$ are estimated with only one output, Y. Furthermore, even if quality differentiated output prices are not available, the decomposition of the productivity effect and its components, technical change and efficiency change, the resource mix effect and the input price effect are invariant, in a model that allows for quality differentiated output quantities. Furthermore, the input side of the quantity effect, the input side of the activity effect and the input side of the scale effect are invariant to the assumption of quality undifferentiated output prices, in a model that allows for quality differentiated output quantities. In contrast, the output side of the quantity effect, the output side of the activity effect, the output side of the scale effect, the output price effect and the product mix effect will vary if quality differentiated output prices are available, in a model that allows for quality differentiated output quantities.

Finally, in the linear programming models (7.13)-(7.17) we further impose the weight restriction that the production of high quality output is at least as resource intensive as the same quantity of output of low quality.

7.3 Data and Empirical Implementation

Here we decompose the change in profits of English and Welsh water companies. Our model includes separate outputs for water and sewerage services, and the three inputs, capital, labour and other inputs. The data covers the period 1991-2008 for a balanced panel of 10 Water and Sewerage companies (WaSCs). Water connected properties and sewerage connected properties, Y_w and Y_s, are our outputs. They are drawn from the companies' regulatory returns to Ofwat. Water and sewage output prices were calculated as the ratio of the appropriate turnover in nominal terms, as available in Ofwat's regulatory returns, to measured output.

The first of three inputs, namely physical capital stock measure is based on the inflation adjusted Modern Equivalent Asset (MEA) estimates of the replacement cost of physical assets contained in the companies' regulatory accounts. However, as periodic revaluations of these replacement cost values could create arbitrary changes in our measure of physical capital, we cannot directly employ these accounting based measures. Rather, we use real net investment is therefore taken as the sum of disposals, additions, investments and depreciation, as deflated by the Construction Output Price Index (COPI). Following Saal & Parker's (2001) approach, we have averaged the resulting year ending and year beginning estimates to provide a more accurate estimate of the average physical capital stock available to the companies in a given year.

We subsequently employed a user-cost of capital approach, to calculate total capital costs as the sum of the opportunity cost of invested capital and capital depreciation relative to the MEA asset values,. We constructed the price of physical capital as the user cost of capital divided by the above MEA based measure of physical capital stocks. The opportunity cost of capital is defined as the product of the weighted average cost of capital (WACC) before tax and the companies' average Regulatory Capital Value (RCV). The RCV is the financial measure of capital stock accepted by Ofwat for regulatory purposes. The WACC calculation is broadly consistent with Ofwat's regulatory assumptions and is estimated with the risk free return assumed to be the average annual yields of medium-term UK inflation indexed gilts. The risk premium for company equity and corporate debt was assumed to be 2% following Ofwat's approach at past price reviews. We also allowed for differences in company gearing ratios and effective corporate tax rates, which were calculated as the sum of aggregate current and deferred tax divided by the aggregate current cost profit before taxation. Finally, following the approach in Ofwat's regulatory current cost accounts, capital depreciation was the sum of current cost depreciation and infrastructure renewals charge.

Moving to our second input, labour, the average number of full-time equivalent (FTE) employees is available from the companies' statutory accounts. Firm specific labour prices were calculated as the ratio of total labour costs to the average number of full-time equivalent employees. Finally our third input, namely

"Other costs" in nominal terms was defined as the difference between operating costs and total labour costs.[14] Given the absence of data allowing a more refined break down of other costs, we employ the UK price index for materials and fuel purchased in purification and distribution of water, as the price index for other costs, and simply deflate nominal other costs by this measure to obtain a proxy for real usage of other inputs. Finally, economic profits are calculated as the difference between turnover and calculated economic costs. Table 1 shows the aggregate statistics for our sample and all the data are expressed in real 2008 prices. To achieve this, we divided profits, turnovers, costs, output and input prices with the RPI index to express the changes in real terms setting the year 2008 as the base year.

As is well documented in past studies (see Saal & Parker 2000, 2001, Saal, Parker and Weyman-Jones, 2007, Maziotis, Saal and Thanassoulis 2009), the English and Welsh water and sewerage companies have been obliged to carry substantial capital investment projects in order to improve water and sewerage quality and environmental standards. Thus, we feel it is important to measure the impact of quality in our profitability, productivity and price performance measures. We therefore adjusted water and sewerage output for high and low water and sewerage quality respectively as follows.

Water quality is defined based on the data regarding drinking water quality and were drawn from the DWI's annual reports for the calendar years ending 1991-2007[15]. Following Saal and Parker (2001) water quality, Q_w, is defined as the average percentage of each WaSC's water supply zones that are compliant with key water quality parameters. Water supply zones are areas designated by the water companies by reference to a source of supply in which not more than 50,000 people reside. The drinking water quality can be defined either based on the sixteen water quality parameters or nine water quality parameters identified as being important for aesthetic, health reasons and cost reasons or based on the six water quality

---

[14] While it would be particularly desirable to disaggregate other input usage data further and in particular to allow for separate energy and chemical usage inputs, the data available at company level from Ofwat's regulatory return does not allow a further meaningful decomposition of other input usage.

[15] The DWI provides quality data based on calendar years, while all other information employed in this paper is based on fiscal years ending March 31st. We note this inconsistency in the data, but emphasize that the reported years overlap each other for 9 months. Thus, the year end to year end estimates of quality change obtained from the DWI data provide consistent estimates of quality change by the water companies, at a fixed point 9 months into each fiscal year.

parameters identified as being indicative of how well treatment works and distribution systems are operated and maintained. Due to changes in some of the drinking water quality standards and the new regulations, the DWI report for 2005 no longer included the two quality indices that compared companies' compliance for the sixteen or nine water quality parameters with the average for England and Wales. So we decided to base the drinking water quality on the six water quality parameters[16] that Ofwat also employs in its assessment. The parameters reflect how well treatment works and distribution systems are operated and maintained (Ofwat, 2006).

High drinking water quality, $Q_{w,h}$, is defined as the average percentage of each WaSC's water supply zones that are compliant with these six water quality parameters. Low drinking water quality $Q_{w,l}$ is defined as the average percentage of each WaSC's water supply zones that are not compliant with these six water quality parameters. The water output for high quality, $Y_{w,h}$, is calculated as the product of the water connected properties and high drinking water quality, $Y_{w,h} = Y_w Q_{w,h}$. The water output for low quality, $Y_{w,l}$ is defined as the product of the water connected properties and low drinking water quality, $Y_{w,l} = Y_w Q_{w,l} = Y_w(1 - Q_{w,h})$. Note that the sum of water output for high and low quality is equal to the water output, $Y_w = Y_{w,h} + Y_{w,l}$. The water output price is the same for high and low quality and it is defined as the ratio of water total turnover in nominal terms to the sum of water output for high and low quality.

Sewerage quality, Q_s, is defined based on the data regarding the percentage of connected population for which sewage receives various types of treatment, zero, primary, secondary or higher treatment. The sewage treatment data were taken from *Waterfacts* for the period 1990-91 to 1995-96 and the companies' regulatory returns for the fiscal years 1996-97 to 2007-08. We henceforward refer to data based on the ending year of the fiscal years. High sewerage treatment quality, $Q_{s,h}$, is defined as the percentage of connected population receiving at least secondary or higher sewerage treatment, while low sewerage treatment quality, $Q_{s,l}$, is defined as the percentage of connected population receiving zero or primary sewerage treatment.

---

[16] The six water quality parameters, which form the Operational Performance Index (OPI) are iron, manganese, aluminium, turbidity, faecal coliforms and trihalomethanes.

The sewerage output for high quality, $Y_{s,h}$, was calculated as the product of sewerage connected properties and the percentage of connected population receiving at least secondary or higher sewerage treatment, $Y_{s,h} = Y_s Q_{s,h}$. The sewerage output for low quality, $Y_{s,l}$ was calculated as the product of sewerage connected properties and the percentage of connected population receiving zero or primary sewerage treatment, $Y_{s,l} = Y_s Q_{s,l}$. Note that the sum of sewerage output for high and low quality is equal to the sewerage output, $Y_s = Y_{s,h} + Y_{s,l}$. The sewerage output price was the same for high and low quality and it was defined as the ratio of sewerage total turnover in nominal terms to the sum of sewerage output for high and low quality. Finally, Table 1 shows the aggregate statistics for our sample and all the data are expressed in real 2008 prices. To achieve this, we divided profits, turnovers, costs, output and input prices with the RPI index to express the changes in real terms setting the year 2008 as the base year.

Since our sample includes 10 WaSCs over an 18 year period, 1991-2008, we decided to modify the estimation with DEA as follows in order to deal with the small number of observations each year. Tulkens & Vanden Eeckaut (1995) proposed four different production sets using DEA in a panel data framework, the contemporaneous, sequential, intertemporal frontiers and window analysis. A contemporaneous production set assumes the construction of a reference production set at each point in time t, from the observations made at that time only. A sequential production set allows the current period technology set to be constructed from data of all the companies in all years prior to and including the current period. Thus, technologies in previous periods are "not forgotten" and remain available for adoption in the current period and therefore in equation 7.8, technical regress is not allowed, $X^A - X^B$ measures only technical progress (see Figure 7.3). An intertemporal production set assumes the construction of a single production set from the observations made throughout the whole observation period. Window analysis is a moving average pattern of analysis, in which each unit in each period is treated as if it is a different unit. The performance of a unit is compared with its performance in other periods, in addition to comparing it with the performance of other units in the same period.

Drawing on the foregoing and the sequential technology in particular, the reference technology for our DEA models is as follows. We have a balanced panel of ten observations (firms) for each year over 1991-2008. We decided to pool the data from 1991-1994 together in order to increase the number of observations from ten to forty. The first sub-panel includes periods $\{1991,1992,1993,1994\}$ and we use the observations from these years as a cross section to construct our reference technology and we refer to the corresponding frontier as our t = 1994 frontier. The second sub-panel contains periods $\{1991,1992,1993,1994,1995\}$ and we use the frontier constructed using the 1991-1995 data as our t+1 = 1995 frontier and so on until the last sub-panel which is actually the entire panel and includes periods $\{1991,1992,1993,1994,...,2008\}$. Thus in essence we use the sequential technology of Tulkens & Vanden Eeckaut (1995) except that our starting technology is the four-year period 1991-1994.

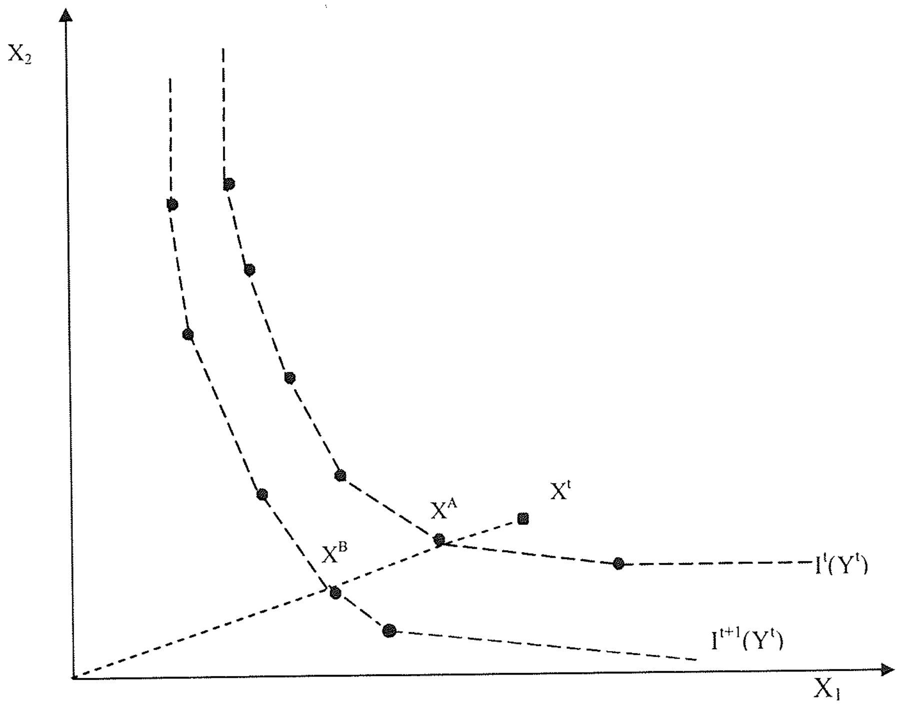


Figure 7-3 Technical Progress

7.4 Empirical Results

In this section we use the methodology outlined above to decompose the profit changes of the water and sewerage sector in England and Wales during the period

1994-2008. Before turning to our results, we first consider trends in aggregate WaSC turnover, costs and profits as reported at Table 7.1 where aggregate statistics for our sample are displayed. Focusing on economic profits, we notice that there was a substantial decrease in aggregate profits over the period 1994 to 1998, from 841.56 million pounds to 93.1 million pounds but a significant increase over the next two years until 2000, where aggregate profits reached their highest level during the entire period of study, 860.1 million pounds. In 2001, the first year of tightened price caps following the 1999 price review, the companies were obliged to reduce the prices charged to customers, and there was a substantial decline in aggregate profits and the industry made economic losses over the period 2001 to 2008, displaying the lowest level of economic loss in 2006, 45.5 million pounds. As far as aggregate turnover was concerned, it increased from 7,126 million pounds to 7,908 million pounds over the years 1994-2000 but it significantly decreased in 2001 at the level of 7,162 million pounds. Over the period 2001-2008, the aggregate turnover increased significantly from 7,162 to 8,494 million pounds. However, economic costs increased from 6,284 to 9,060.9 million pounds over the period 1994-2008 showing the highest level of increase in 2000/01. Thus, in aggregate, the increase in turnover after 2001 was outstripped by even greater increases in economic costs resulting in substantial economic loss.

	Units	1994	1995	1996	1997	1998	1999	2000	2001
Economic Profit	£000 000s (2008)	841.6	654.8	347.6	403.9	93.1	295.9	860.1	-613.8
Revenues	£000 000s (2008)	7,126.4	7,295.5	7,565.2	7,601.9	7,610.7	7,809.8	7,908.2	7,162.9
Total Economic Costs	£000 000s (2008)	6,284.8	6,640.6	7,217.6	7,198.0	7,517.6	7,513.9	7,048.1	7,776.8
Water Connected Properties	000s	16,665	16,897	17,785	17,966	18,102	18,191	18,304	19,302
High Quality Adjusted Water Connected Properties[1]	000s	15,101	15,604	16,305	16,550	16,745	16,866	17,237	18,412
Low Quality Adjusted Water Connected Properties[2]	000s	1,564	1,293	1,480	1,417	1,357	1,325	1,067	890
Sewerage Connected Properties	000s	21,298	21,479	21,710	21,816	21,886	21,908	22,123	22,274
High Quality Adjusted Sewerage Connected Properties[3]	000s	16,963	17,269	17,742	17,903	18,122	18,564	19,239	20,939
Low Quality Adjusted Sewerage Connected Properties[4]	000s	4,335	4,210	3,967	3,913	3,764	3,343	2,884	1,335
Capital	£000 000s (2008)	192,295	194,408	197,981	199,671	201,885	204,316	206,597	208,168
Number of employees	FTE	38,125	37,741	34,926	32,424	31,745	30,829	29,685	27,854
Other Inputs	£000 000s (2008)	999.5	997.7	1,020.3	987.9	918.7	897.3	970.1	958.6
Avg. Price for a Water Connected Property	£s (2008)	198.24	200	195.54	191.21	187.66	189.34	186.22	169.16
Avg. Price for a Sewerage Connected Property	£s (2008)	202.21	207.87	210.61	213.20	214.61	223.08	228.68	194.69
Avg. Price for a Quality Adjusted Water Connected Property	£s (2008)	219.01	216.27	214.49	208.96	204.17	205.30	198.75	176.90
Avg. Price for a Quality Adjusted Sewerage Connected Property	£s (2008)	229.87	236.89	236.23	236.59	232.95	234.3	229.5	175.07
Price for Capital	£s (2008)	0.017	0.019	0.022	0.022	0.024	0.024	0.021	0.025
Price for Labour	£000s (2008)	32.17	32.19	32.23	32.03	31.58	33.04	33.78	33.46
Price of Other Inputs[5]	(2008)	0.740	0.738	0.754	0.769	0.785	0.806	0.767	0.762

1. Calculated as the product of water connected properties and the average percentage of each WaSC's water supply zones fully compliant with key drinking water quality parameters
2. Calculated as the product of water connected properties and the average percentage of each WaSC's water supply zones not compliant with key drinking water quality parameters
3. Calculated as the product of sewerage connected properties and the percentage of population receiving at least secondary or higher sewerage treatment
4. Calculated as the product of sewerage connected properties and the percentage of population receiving zero or primary treatment
5. UK price index for materials and fuel purchased in purification and distribution of water

	Units	2002	2003	2004	2005	2006	2007	2008
Economic Profit	£000 000s (2008)	-116.5	-395.5	-632.7	-604.5	-45.5	-680.3	-566.3
Revenues	£000 000s (2008)	7,272.6	7,150.4	7,291.1	7,491.9	8,198.2	8,274.1	8,494.6
Total Economic Costs	£000 000s (2008)	7,389.1	7,545.9	7,923.9	8,096.3	8,243.6	8,954.5	9,060.9
Water Connected Properties	000s	19,444	19,573	19,702	19,821	19,972	20,087	20,061
High Quality Adjusted Water Connected Properties	000s	18,591	18,797	18,973	19,083	19,297	19,423	19,442
Low Quality Adjusted Water Connected Properties	000s	853	776	729	738	676	664	619
Sewerage Connected Properties	000s	22,498	22,648	22,809	23,017	23,456	23,602	23,795
High Quality Adjusted Sewerage Connected Properties	000s	21,842	22,141	22,330	22,647	23,186	22,935	23,072
Low Quality Adjusted Sewerage Connected Properties	000s	656	508	480	370	270	667	723
Capital	£000 000s (2008)	209,158	210,379	211,853	213,253	214,362	215,511	216,918
Number of employees	FTE	27,188	26,667	26,919	27,197	27,554	28,218	29,524
Other Inputs	£000 000s (2008)	973.7	969.9	939.6	824.6	819.3	841.6	781.4
Avg. Price for a Water Connected Property	£s (2008)	169.4	164.8	167.14	170.72	187.03	190.18	197.06
Avg. Price for a Sewerage Connected Property	£s (2008)	196.83	192.75	195.77	200	213.79	214.75	219.27
Avg. Price for a Quality Adjusted Water Connected Property	£s (2008)	177.93	172.01	173.68	178.06	193.87	197.24	204.35
Avg. Price for a Quality Adjusted Sewerage Connected Property	£s (2008)	168.39	162.85	165.37	168.01	178.43	181.21	185.7
Price for Capital	£s (2008)	0.023	0.023	0.025	0.026	0.026	0.028	0.029
Price for Labour	£ 000s (2008)	32.72	34.63	35.54	37.63	37.92	37.66	36.85
Price of Other Inputs	(2008)	0.765	0.768	0.797	0.889	0.957	0.971	1.000

Table 7-1 Aggregate Profits, Revenues, Costs, Outputs, and Inputs

7.4.1 Quality Unadjusted Results

Table 7.2 displays cumulative profit change and the drivers of profit change defined in equations (7.8) to (7.11) for the entire 1994-2008 period and the regulatory sub-periods 1994-2000, 2000-2005 and 2005-2008, but without making any allowances for any differences in the quality of outputs. Over the entire 1994-2008 period, the quantity effect, efficiency change, resource mix and scale effect contributed positively to profit changes, while the price effect, technical change and product mix effect contributed negatively to profit changes. Focusing on aggregate profit change, profits reduced by 1,407.9 million pounds over the period 1994-2008, which was the result of significant aggregate profit decrease during the years 2000-2005 and small aggregate profit increases during the years 1994-2000 and 2005-2008. In aggregate, profits increased by 18.5 and 31.8 million pounds respectively during the years 1994-2000 and 2005-2008 and reduced by 1,464.6 million pounds during the years 2000-2005.

	1994-2008	1994-2000	2000-2005	2005-2008
Profit change	-1,407.9	18.5	-1,464.6	38.1
Quantity effect	1,289.4	514.1	661.9	113.4
Output effect	1,080.4	482.6	413.7	184.1
Input effect	209.0	31.5	248.2	-70.7
Productivity	1,205.6	613.4	529.5	62.7
Technical Change	1,086.1	631.3	368.0	86.8
Efficiency Change	119.4	-17.9	161.5	-24.1
Activity effect	83.8	-99.3	132.4	50.8
Resource Mix	939.1	124.2	355.8	459.1
Product Mix	-2.1	47.1	-90.5	41.3
Scale Effect	-853.1	-270.6	-132.9	-449.6
Price Effect	-2,697.3	-495.5	-2,126.5	-75.3
Output Price Effect	287.7	299.2	-830.0	818.6
Input Price Effect	-2,985.02	-794.70	-1,296.44	-893.88

Table 7-2 Cumulative Profit Change and Its Decomposition (2008 pounds, millions)

Looking at the first stage of profit decomposition, where profit change was decomposed into a quantity and price effect (see equation 7.8), we conclude that over the entire period, the negative effect on cumulative profit change was attributed to a significant negative price effect which outstripped the positive quantity effect. The cumulative impact of the price effect led to a 2,697.3 million pounds reduction in profits offsetting the cumulative impact of the quantity effect which resulted in a 1,289.4 million pounds increase in profits.

At the first stage of profit decomposition, the price effect can be further decomposed into an output price and input price effect and the quantity effect into an output and input effect. During the years 1994-2008, output prices increased profits by 287.7 million pounds, however, greater increases in input prices contributed negatively to profit changes by 2,985.02 million pounds resulting in the overall negative entire price effect. Focusing on the sub-periods of our sample, we conclude that during the years 1994-2000, covering the end of the first price review after privatization and the entire 1995-2000 period covered by the 1994 price review, there was a small increase in output prices contributing positively to profit changes, 299.2 million pounds. However, substantial increases in input prices counteracted this as they reduced profits by 794.70 million pounds. Furthermore, the dramatically tightened 1999 price review obliged the companies to reduce their output prices and continuing increases in input prices resulted in a negative overall price effect which contributed negatively to profit changes, 2,126.5 million pounds between 2000 and 2005. During the years 2005-2008, output prices increased significantly, providing evidence that the 2004 price review was relatively loose and thereby contributing positively to profit changes, 818.6 million pounds, whereas increases in input prices moderated and reduced profits by 893.88 million pounds resulting in a small overall negative price effect.

In contrast to the high negative price effect, the overall positive quantity effect was attributed to a substantial increase in outputs contributing 1,080.4 million pounds to profit changes. Significant aggregate output increases occurred during the years 1994-2000 and 2000-2005, contributing positively to profit changes, 482.6 and 413.7 million pounds respectively, whereas small aggregate increase in outputs during the years 2005 -2008 increased profits by 184.1 million pounds. Focusing on aggregate input effect, the input effect increased profits by 209 million pounds over the period 1994-2008, which was the result of significant aggregate input usage reductions during the years 1994-2000 and 2000-2005 and small aggregate input usage increase during the years 2005-2008. In aggregate, input usage reductions increased profitability by 31.5 and 248.2 million pounds respectively during the years 1994-2000 and 2000-2005 and input usage increases reduced profitability by 70.7 million pounds during the years 2005-2008. It is worth mentioning that during the years 1994-

2000 and 2005-2008 small increases in aggregate profits were attributed to the substantial positive quantity effect which outstripped the negative price effect. However, the magnitude of the negative price effect, derived from both input and output price effects, during 2000-2005 resulted in a dramatic deterioration in economic profitability between 2000 and 2005, despite a substantial positive quantity effect amounting to 661.9 million pounds.

Looking at the second stage decomposition (see equation 7.9) we see in Table 7.2 that the positive quantity effect over the entire period was attributed to a significant positive productivity effect and a small but positive activity effect. During the years 1994-2008, the productivity effect increased profits by 1,205.6 million pounds, whereas the activity effect increased profits by 83.8 million pounds. Almost the entire productivity effect can be explained by technical change which contributed positively to profit change 1,086.1 million pounds, while the contribution of increased efficiency only amounted to 119.4 million pounds. In aggregate, the productivity effect significantly increased profits by 613.4 and 529.5 million pounds respectively during the years 1994-2000 and 2000-2005, while 62.7 million pounds contribution during the years 2005-2008 was much more modest.

At the third stage of profit decomposition, the productivity effect decomposes into the technical change and efficiency change effects (see equation (7.10)) and the activity effect decomposes into the resource mix, output mix and scale effects (see equation (7.11)). Focusing on the components of productivity effect in Table 7.2, technical change was positive during the years 1994-2008, increasing profits by 1,086.1 million pounds, showing the highest magnitude of increase during the years 1994-2000 and 2000-2005, 631.4 and 529.5 million pounds respectively. In contrast to the substantial positive technical change effect, efficiency change was small and negative during the years 1994-2000 and 2005-2008, but did substantially increase profits by 161.5 million pounds during the years 2000-2005.

Focusing on the decomposition of the activity effect in Table 7.2, it is concluded that in aggregate the positive activity effect was mainly explained by substantial positive resource mix which was unfortunately largely offset by the very substantial negative scale effect as well as the quite small negative product mix effect. Over the whole period, the resource mix effect contributed 939.1 million pounds to increased profits, whereas the scale effect and product mix effect reduced profits by 853.1 and 2.1 million pounds respectively. The resource mix effect increased significantly over the entire period and especially during the years 2000-2005 and 2005-2008 suggesting movement to a more cost efficient allocation of resources more in line with relative factor prices. Thus, over the whole period, capital input increased by 12.8%, whereas labour input decreased by 22.56% as can be seen in Table 7.1, indicating that the water industry became more capital-intensive and less labour-intensive. Moreover, the scale effect, resulting from respective increases in water and sewerage outputs of 20.37%, and 11.72%, did not lower costs and reduced profits significantly during the years 1994-2000 and 2005-2008 by 270.6 and 449.6 million pounds respectively. However, the negative impact of the scale effect declined during the years 2000-2005. Changes in the mix of outputs, the production of more output for water services than sewerage services increased profits by 47.1 and 41.3 million pounds respectively during the years 1994-2000 and 2005-2008 but decreased profits significantly by 90.5 million pounds during the years 2000-2005.

Overall, relating the results from the decomposition of profits into several factors in Table 7.2 with the regulatory cycle, we conclude that during the years 1994-2000 when price caps were tightened after the 1994 price review, profits increased. This increase in aggregate profitability was attributed to the positive cumulative quantity effect, and still increasing output prices which just offset substantial increases in input prices. There were also significant improvements in productivity mainly attributable to technical change, indicating that the most productive companies significantly improved their performance. Furthermore, there was evidence that changes in the mix of inputs and outputs had a positive impact on aggregate profitability until 2000. During the years 2000-2005 when profits substantially decreased, the cumulative impact of price effect as captured by a significant reduction in output prices due to the tightened 1999 price review and a high increase in input prices offset the positive quantity effect. However, there were still substantial productivity improvements attributed to both technical change and increased efficiency, indicating that both the most productive and the less productive firms had strong incentives to improve their productivity in order to regain economic profitability. Moreover, adjusting to a more cost efficient input mix also appeared to have lowered costs and increased profits. Finally, during the years 2005-2008, when profits increased very slightly, this was explained by a positive cumulative impact of the quantity effect, and substantial gains in output prices, which were nonetheless almost completely offset by large increases in input prices. Digging deeper into the quantity effect reveals that changes in the mix of input, outputs and technical change had a positive impact on aggregate profitability. However, no improvements in efficiency change and increases in the scale of operations (diseconomies of scale) significantly reduced aggregate profitability.

Over the whole period, the major positive determinants on the quantity effect and eventually on profit change as defined in equations (7.8) to (7.11) came from the technical change and the resource mix effect, whereas the efficiency change had a small positive impact on aggregate economic profitability. Technical change contributed positively to profit changes over the entire period, however, its magnitude reduced after 2000 and especially during the years 2005-2008 suggesting that significant productivity improvements by the frontier company were achieved before 2000. Also, the efficiency change effect became negative indicating that after 2005 the less productive companies did not improve their productivity towards the frontier company. Catch-up gains in productivity occurred during the years 2000-2005, when the profits substantial fell and the companies needed to improve their productivity in order to regain economic profitability. As far as the resource mix effect is concerned, it was large and positive over the whole period and its magnitude significantly increased after 2000, indicating that the water and sewerage industry moved to a more cost efficient allocation of resources more in line with relative factor prices.

Moreover, the major negative determinants on the quantity effect and eventually on profit changes came from the scale effect whose magnitude substantially increased during the years 2000-2008. This finding suggests that the mergers occurred in 2000/01 did not lower costs and as a result, the water and sewerage companies operated under diseconomies of scale contributed negatively to profit changes. This is mainly apparent during the years 2000-2005 when the profits substantially fell due to the negative impact from the mergers combined with the high increase in input prices. Also, during the years 2005-2008, the bigger negative scale effect offset the positive impact of technical change and resource mix effect resulting in a small increase in aggregate profitability. It is concluded that aggregate profitability should have fallen by less or increased by more if the bigger negative scale effect did not exist after 2000 despite the negative overall price effect.

7.4.2 Results After Controlling for High and Low Quality

We turn our discussion now to the results from cumulative profit change and its decomposition for the periods 1994-2000, 2000-2005 and 2005-2008 when we allow for differences in the quality of output. As explained in sections 7.2 and 7.3, the sum of water output of high and low quality was equal to the quality unadjusted water output and the sum of sewerage output of high and low quality was equal to the quality unadjusted sewerage output. The output price was the same regardless of the level of quality, high or low. Therefore, the results from the first stage of the profit decomposition in Table 7.3, the quantity and price effect will be exactly the same as those in Table 7.2, when quality is not included in our analysis. Differences between the quality unadjusted results and the results after controlling for high and low quality relate to the decomposition of the quantity effect into a productivity and activity effect, e.g. in the second and third stage of the profit decomposition.

Table 7.3 further depicts the results from the decomposition of the output effect into high quality and low quality output effect. The results indicate that over the whole period the water and sewerage companies moved to the production of more high quality of output than low quality of output contributing positively to the overall output effect and therefore to profit changes Over the whole period, high quality outputs increased profits by 2,067.1 million pounds. Significant aggregate high quality output increases occurred during the years 1994-2000 and 2000-2005, contributing positively to profit changes, 902.3 and 1,015.5 million pounds respectively, whereas small aggregate increases in high quality outputs during the years 2005-2008 increased profits by 149.3 million pounds. Focusing on the aggregate low quality output effect, it decreased profits by 986.7 million pounds over the period 1994-2008, which was the result of significant aggregate low quality output reductions during the years 1994-2000 and 2000-2005 and small aggregate low quality output increase during the years 2005-2008. In aggregate, low quality output reductions decreased profitability by 419.7 and 601.8 million pounds respectively during the years 1994-2000 and 2000-2005 and low quality output increases increased profitability by 34.8 million pounds during the years 2005-2008.

	1994-2008	1994-2000	2000-2005	2005-2008
Profit change	-1,407.9	18.5	-1,464.6	38.1
Quantity effect	1,289.4	514.1	661.9	113.4
Output effect	1,080.4	482.6	413.7	184.1
High Quality Output Effect	2,067.1	902.3	1,015.5	149.3
Low Quality Output Effect	-986.7	-419.7	-601.8	34.8
Input effect	209.0	31.5	248.2	-70.7
Productivity	1,137.4	586.4	480.1	70.8
Technical Change	1,031.3	575.8	340.0	115.5
Efficiency Change	106.1	10.6	140.1	-44.7
Activity effect	152.0	-72.4	181.8	42.6
Resource Mix	1,188.4	260.7	528.0	399.7
Product Mix	30.4	-60.9	81.8	9.5
Scale Effect	-1,066.8	-272.2	-428.0	-366.6
Price Effect	-2,697.3	-495.5	-2,126.5	-75.3
Output Price Effect	287.7	299.2	-830.0	818.6
Input Price Effect	-2,985.02	-794.70	-1,296.44	-893.88

Table 7-3 Cumulative High And Low Quality Adjusted Profit Change and Its Decomposition (2008 pounds, millions)

The positive quantity effect over the entire period can be entirely attributed to the significant positive productivity which offset the small but positive activity effect. During the years 1994-2008, the productivity effect substantially increased profits by 1,137.4 million pounds, whereas the activity effect increased profits only by 152 million pounds. The positive productivity effect can be entirely attributed to technical change which increased profits by 1,031.3 million pounds and offset the small but positive efficiency change which increased profits only by 106.1 million pounds. In aggregate, the productivity effect increased profits substantially by 586.4 and 480.1 million pounds respectively during the years 1994-2000 and 2000-2005, whereas its magnitude reduced during the years 2005-2008 since it slightly increased profits by 70.8 million pounds during the years 2005-2008. Focusing on the components of the productivity effect, technical change was large and positive during the years 1994-2000 and 2000-2005, increasing profits by 575.8 and 340 million pounds respectively, whereas it slightly increased profit changes for the years 2005-2008, 115.5 million pounds. In contrast to the substantial positive technical change, efficiency change was positive during the years 1994-2000 and 2000-2005, increasing profits by 10.6 and 140.1 million pounds respectively, while it remained negative during the years 2005-2008 reducing profits by 44.7 million pounds.

Focusing on the decomposition of the activity effect, it is concluded that in aggregate the positive activity effect was mainly explained by a high positive resource mix and small product mix effect, which outstripped a very large and substantial negative scale effect.. Over the whole period, the resource mix and product mix effect substantially contributed to increased profits by 1,188.4 and 30.4 million pounds respectively, whereas scale effect reduced profits by 1,066.8 million pounds. The resource mix effect contributed significantly to profit change over the entire period and especially after 2000 indicating that there was a steady shift to a more capital intensive resource allocation that was more cost effective given observed input prices. Thus, over the whole period, capital input increased by 12.8%, whereas labour input decreased by 22.56% as can be seen in Table 7.1, indicating that the water industry became more capital-intensive and less labour-intensive. Moreover, the scale effect did not lower costs and reduced profits significantly during the years 2000-2005 and 2008 by 428 and 366.6 million pounds. The substantial savings occurred by the resource mix effect were lost due to excessive mergers, the negative scale effect implying that during the entire period, the water and sewerage industry was operating under diseconomies of scale which affected negatively aggregate economic profitability. Changes in the mix of outputs, the production of more output for water services than sewerage services increased profits significantly by 81.8 and 9.5 million pounds respectively during the periods 2000-2005 and 2005-2008 but decreased profits by 60.9 million pounds during the period 2000-2005. As can be seen from Table 7.1, over the whole period there was an increase in output for water services of 20.37%, while the output for sewerage services increased by 11.72%.

The results after controlling for high and low quality indicate that during the years 1994-2000 when price caps were tightened after the 1994 price review, profits increased. This was attributed to the positive cumulative quantity effect, which offset the overall negative price effect. There were significant increases in the production of high quality output and reductions in the production of low quality output which outstripped the overall negative price effect, inputs prices increased greater than output prices. There were also significant improvements in productivity mainly attributable to technical change, indicating that the most productive companies significantly improved their performance, whereas gains in efficiency were positive but small. Furthermore, there was evidence that changes in the mix of inputs had a positive impact on aggregate profitability until 2000, whereas scale effect and output mix effect contributed negatively to profit changes. During the years 2000-2005 when profits substantially decreased, the cumulative impact of price effect as captured by a significant reduction in output prices due to the tightened 1999 price review and a high increase in input prices offset the positive quantity effect. However, there were still substantial productivity improvements attributed to both technical change and increased efficiency. Moreover, adjusting to a more cost efficient input mix and a shift to the production of more output for water services than sewerage services also appeared to have lowered costs and increased profits. However, the negative impact of scale effect on aggregate profitability became greater implying that mergers occurred in 2000/01 eventually reduced profits. Finally, during the years 2005-2008, when profits increased very slightly, this was explained by a positive cumulative impact of the quantity effect, and substantial gains in output prices, which were nonetheless almost completely outstripped by large increases in input prices. Changes in the mix of inputs, outputs and technical change had a positive impact on aggregate profitability. However, the negative efficiency change and scale effect, no gains in productivity by less productive firms and increases in the scale of operations (diseconomies of scale) significantly reduced aggregate profitability.

Over the whole period, the major positive determinants on the quantity effect and eventually on profit change as defined in equations (7.8") to (7.11") came from the technical change and the resource mix effect, whereas the impact of the efficiency change and product mix effect on profit changes was small. The results suggest that although technical change contributed positively on profit changes over the entire period, it started to fall after 2000 implying that the frontier companies achieved significant productivity improvements before 2000. The efficiency change effect became negative during the years 2005-2008 indicating that gains in productivity by less productive firms occurred during the years 1994-2005. As far as the resource mix effect is concerned, its magnitude significantly increased after 2000, indicating that the water and sewerage industry moved to a more cost efficient allocation of resources more by substituting labour with capital. Finally, the product mix effect was negative during the years 1994-2000 but slightly increased after 2000 indicating that the water and sewerage industry moved to the production of more output for water than sewerage services contributing positively to profit changes.

Moreover, the major negative determinants on the quantity effect and eventually on profit changes came from the scale effect whose magnitude substantially increased during the years 2000-2008. This finding suggests that the mergers occurred in 2000/01 had a negative impact on aggregate profitability. As a result, on average the water and sewerage companies operated under diseconomies of scale which contributed negatively to profits. This finding is apparent during the years 2000-2005 when the profits substantially fell due to the negative impact from the mergers combined with the high increase in input prices. Also, during the years 2005-2008, the bigger negative scale effect offset the positive impact of technical change and resource mix effect resulting in a small increase in aggregate profitability. Any substantial savings occurred by the resource mix effect were lost due to excessive mergers, indicating that over the whole period the water and sewerage companies were operating under diseconomies of scale which had a negative impact on aggregate profitability.

Looking at the two types of profit decomposition, it is concluded that without and after controlling for quality there were minor differences in the results. In both cases, the major determinant on the negative aggregate profitability is explained by the overall negative price effect which outstripped the overall positive quantity effect. The difference between the results from the two types of profit decomposition is on the magnitude of the productivity and activity effect. Without controlling for quality, the substantial positive productivity effect, which was explained by the large technical change outstripped the small positive efficiency change, offset the small positive activity effect, which was attributed to the significant positive resource mix effect offsetting the substantial negative scale effect and the small negative product mix effect. After controlling for quality, the impact of the productivity effect, both technical change and efficiency change, on the aggregate profitability reduced from 1,205.6 to 1,137 million pounds. Furthermore, the impact of the activity effect on profit change slightly increased, from 83.8 to 152 million pounds. Resource mix and product mix effect increased by 249.3 and 28.3 million pounds, whereas the magnitude of the scale effect increased even further, by 213.7 million pounds. Furthermore, after controlling for quality, we offered a further decomposition of the output effect into high quality and low quality output. The results showed that the water and sewerage industry moved to the production of more high quality of output than low quality of output contributing positively to the overall quantity effect and therefore to profit changes.

Moreover, without and after controlling for quality, the major determinants on the quantity effect and eventually, on profit change came from the technical change whose magnitude significantly reduced during the years 2005-2008, the resource mix effect, a shift to a more cost efficient allocation of resources by substituting labour with capital, and the negative scale effect whose magnitude substantially increased after 2000, implying that the mergers occurred in 2000/01 did not lower costs and therefore had a negative impact on aggregate profitability. Finally, in both types of the profit decomposition efficiency change was found to have a small positive impact on profit change, whereas after controlling for high and low quality, the product mix effect became positive but small, contributing positively to profit changes.

7.5 Conclusions

In this study, we firstly applied an input oriented profit decomposition approach following the approach of De Witte & Saal (2009). Then, by making allowances for differences in the quality of output we extended the profit decomposition approach of Grifell and Lovell (1999). We decomposed profit changes into various factors that were of significance for the regulator and the regulated firms such as price effect, technical change, efficiency change, resource mix, product mix and scale effect. We also adapted the sequential DEA approach of Tulkens and Vanden Eeckaut (1995) so that we could compute profit decomposition even when the number of observations is extremely limited. We applied our profit decomposition approaches to the water and sewerage Companies (WaSCs) in England and Wales over the period 1991-2008. The profit decomposition approaches, without and after controlling for quality demonstrated similar results and the differences between the results was on the magnitude of the productivity and activity effect. Furthermore, after controlling for quality, we offered a further decomposition of the output effect into high quality and low quality output.

The quality unadjusted results indicated that over the whole period the main source of negative profit change was driven by the substantial negative price effect which outstripped the overall positive quantity effect. In aggregate, significant increases in input prices contributed negatively to profit changes and small increases in output prices contributed positively to profit changes. Moreover, substantial improvements in technical change, resource mix and small gains in efficiency had a positive impact on aggregate profit change, whereas the significant negative scale effect and the small negative product mix reduced profits. During the first tight price cap regulation period, 1994-2000, there were small increases in profits mainly attributed to significant increases in technical change, the mix of efficient input usage and small increases in the mix of outputs produced. These factors which contributed positively to the quantity effect outstripped the increases in output and input prices. When prices were tightened during the years 2000-2005, profits substantially decreased. Any significant increases in technical change, efficiency and resource mix were outstripped by the higher increase in input prices together with the substantial negative scale effect and the significant reduction in output prices. Finally, when prices became more lax during the years 2005-2008, profits slightly increased. This was entirely attributed to the productivity and activity effect. Substantial improvements in the resource mix effect and small increases in technical change and the product mix outstripped the high increase in output and input prices, whereas the scale effect continued to be negative and its magnitude significantly increased after 2005 contributing negatively to profit change.

Furthermore, the results from the profit decomposition after controlling for high and low quality were similar to those obtained without controlling for quality. Over the whole period the main source of negative profit change was driven by the substantial negative price effect which outstripped the positive quantity effect. The overall positive quantity effect was attributed to substantial increases in outputs and a small but positive input effect. The positive output effect was attributed to a substantial increase in high quality outputs which outstripped the negative low quality output effect, which was the result of low quality output reductions during the years 1994-2005. Substantial improvements in technical change, resource mix and efficiency, and small increases in the output mix had a positive impact on aggregate profit change, whereas scale effect continued to influence negatively profit changes.

The difference between the two types of profit decomposition was on the magnitude on the productivity and activity effect and their components. During the years 1994-2000, profits slightly increased due to the significant improvements in productivity achieved by the most and less productive companies, the resource mix, which outstripped the increases in input and output prices together with the negative scale and product mix effect. After the tightened 1999/00 price review, during the years 2000-2005, profits substantially decreased. Any significant improvements in technical change, efficiency, the mix of efficient inputs used and the mix of outputs produced were outstripped by the high increases in input prices and the reduction in output prices, whereas the scale effect continued to have a substantial negative impact on aggregate profitability. During the years 2005-2008, where new looser price caps were introduced, there were small increases in profit changes. The main determinants for profit increases were the resource mix, technical change and product mix effect which outstripped the high increases in output and input prices and the substantial negative scale effect.

Therefore, it is concluded that without and after controlling for quality, the major determinants on the quantity effect and eventually, on profit change came from the technical change whose magnitude, however, substantially reduced during the years 2005-2008, the resource mix effect, a shift to a more cost efficient allocation of resources by substituting labour with capital, and the negative scale effect whose magnitude substantially increased after 2000, suggesting that the mergers occurred in 2000/01 did not eventually lower costs. Finally, in both types of the profit decomposition efficiency change was found to have a small positive impact on profit change, whereas after controlling for high and low quality, the product mix effect became positive but small contributing positively to profit changes. In both cases, any substantial savings occurred by the resource mix effect were lost due to excessive mergers, indicating that over the whole period the water and sewerage companies were operating under diseconomies of scale which had a negative impact on aggregate profitability.

Our methodology facilitated a backward-looking approach that allowed conclusions to be drawn with regard to the impact of price cap regulation on the financial performance of the regulated companies when the number of observations was extremely limited. This methodology enables regulators and regulated companies to identify the sources of profit variation such as price effects, productivity effects, changes in the mix of resources, outputs and the scale of operations and aid them to evaluate firstly the effectiveness of the price cap scheme and the performance of the regulated companies. Secondly and more significantly, profit decomposition enables the regulator to identify those sources of profits that can be passed along to the consumers e.g. any improvements in productivity that could pass to the consumers in terms of lower output prices. Moreover, our methodology can also be used by the regulated companies to identify the determinants of their profit changes and improve future performance, thereby leading to future profit gains.

Furthermore,, our methodology showed that even if quality differentiated output price is not available, the decomposition of the productivity effect and its components, technical change and efficiency change, the resource mix effect and the input price effect are invariant, in a model that allows for quality differentiated output quantities. Furthermore, the input side of the quantity effect, the input side of the activity effect and the input side of the scale effect are invariant to the assumption of the assumption of quality undifferentiated output prices, in a model that allows for quality differentiated output quantities. In contrast, the output side of the quantity effect, the output side of the activity effect, the output side of the scale effect, the output price effect and the product mix effect will vary if quality differentiated output price is available, in a model that allows for quality differentiated output quantities.

Finally, the results from the two types of profit decomposition have significant policy implications for the regulated water and sewerage industry in England and Wales and can be summarised as follows. Firstly, the substantial capital investment programs carried output by the water and sewerage companies since privatization leaded to the production of high quality output and the reduction of low quality output. Secondly, significant productivity improvements which contributed positively to profit changes were mainly attributed to technical change, whereas gains in efficiency were small. This finding is consistent with Cave's review (2009) findings which suggested that since privatization the main driver on productivity growth for the UK water and sewerage sector was attributed to technical change, however, our findings also suggest that technical was falling over time. Finally, the results from the profit decompositions showed that the resource mix effect was significantly large and positive over the whole period indicating that the water and sewerage industry moved to a cost efficient allocation of resources by substituting labour with capital and therefore contributing positively to profits. However, any substantial savings occurred by the resource mix effect were lost due to excessive mergers. The scale effect was negative over the whole period and substantially increased after 2000 indicating that the mergers occurred in 2000/01 had a negative impact on aggregate economic profitability. Therefore, this finding suggests that mergers were not profitable for WaSCs which is in contrast to Cave's review (2009) recommendations which suggested further mergers in the UK water and sewerage industry. We strongly believe that this finding is of great significance as it will allow further analysis on developing methodologies to explore the issue of economies of scale and scope and conclude about the most economically efficient structure and the existence of vertical integration economies in the UK water and sewerage industry (forward-looking).

CHAPTER 8 CONCLUSIONS, CONTRIBUTIONS AND IMPLICATIONS OF THE STUDY

8.1 *Introduction*

This chapter considers firstly the conclusions, contributions and implications of the thesis findings, as well as limitations and suggestions for future research. The conclusions are drawn after taking into account the results, the discussion of the findings and conclusions in previous chapters.

The aim of the thesis has been to evaluate the effectiveness of Ofwat's regulatory price cap scheme on the financial performance of the Water and Sewerage Companies (WaSCs) when the number of observations is small. This was achieved by analysing the relationship between profitability, productivity and price performance across WaSCs and over time with panel data and using index number techniques and Data Envelopment Analysis (DEA). Specifically, the first objective has been to measure changes in profitability, productivity and price performance across firms (relative comparative performance) and thereby identifying whether the price caps faced by the WaSCs were consistent with a regulatory system focusing on promoting both productive and allocative efficiency. The second objective has been to measure changes in profitability, productivity and price performance by less productive firms and benchmark firms over time. The third objective has been to measure changes in profitability across firms over time caused by other factors except for productivity and price changes such as resource mix, product mix and scale effect. Finally, the fourth objective has been to allow for exogenous factors such as quality in a profit decomposition analysis since the UK water and sewerage sector has carried out substantial capital investment programs to improve water and sewerage quality and environmental standards.

In achieving the first objective, across sectional (spatial) index number technique has been employed in Chapter 5 to measure differences in the level of productivity, price performance and profitability across WaSCs (relative comparative performance) over the period 1991-2008. A firm's actual economic profitability has been decomposed into two sources: a spatial multilateral Fisher productivity index (TFP) whose inverse has been interpreted as a regulatory excess costs index measuring the excess of a firm's actual costs relative to benchmark costs, and a newly developed regulatory total price performance (TPP) index, measuring the excess of regulated revenues relative to benchmark costs. The theoretical relationship between actual economic profitability, regulatory excess costs and regulatory price performance allowed a characterisation of the power of regulatory price caps. In achieving the fourth objective, water and sewerage quality was also included in our analysis and affected the regulatory excess costs and total price performance measures.

The second objective was accomplished by employing a panel index number technique in Chapter 6 that allowed us to measure differences in economic profitability, productivity and price performance across WaSCs over the period 1991-2008. The reconciliation of the unit-specific (across time for each firm) and spatial (across firms) profitability, productivity and price performance into relative profitability, productivity and price performance measures, allowed us to provide a single index that consistently measures performance change between both firms and over time. In achieving the fourth objective, quality was also included in our productivity and price performance measures, and therefore we were able to consistently decompose the unit-specific economic profitability change as a function of the quality unadjusted catch-up in productivity, the catch-up in quality regarding productivity, and the quality-unadjusted productivity and quality performance over time of the benchmark firm, and the quality unadjusted catch-up in price performance, the catch-up in quality regarding price performance, and the price performance and quality growth of the benchmark firm.

In achieving the third and fourth objective, profit changes has been decomposed into several factors that were of great significance for the regulator and the regulated companies such as the quantity and price effect, technical change and efficiency change, the resource mix, product mix and scale effect in Chapter 7 by employing both index number techniques and Data Envelopment Analysis (DEA). We included the impact of exogenous characteristics like quality in a profit decomposition analysis, captured as output for high and low quality and assuming that consumers pay the same price for high and low output quality. Finally, our sequential DEA technique (using data from all the previous periods) allowed us to estimate the productivity and the activity effect and their components when the number of observations was extremely limited.

8.2 *Conclusions, contributions, implications*

The key theoretical contribution in Chapter 5 was firstly to demonstrate across sectional (spatial) index number technique to measure differences in the level of productivity, price performance and profitability across firms (relative comparative performance) when the number of observations was extremely limited, making the approach directly applicable by regulators in setting price caps. Secondly, and more significantly, it also allowed the development of the theoretically consistent model of price cap regulation. It provided an analysis of whether price caps are consistent with the achievement of productive and allocative efficiency. We also showed that the relationship between actual economic profitability, regulatory excess costs and regulatory price performance indices can be used to categorize regulatory price caps as "weak", "powerful" or "catch-up promoting"

The results indicated that throughout the entire 1991-2008 period price caps were never "powerful", in the sense that they required less productive firms to immediately and fully catch-up to the most productive firm to regain economic profitability. While our results did suggest substantial regulatory tightening after 1994, we emphasized that the period 1991-2000 could be characterised as a period of "weak" regulation since allowed regulatory revenues almost always exceeded regulatory excess costs, thereby demonstrating that price caps during this period allowed firms to maintain economic profitability regardless of whether they made any progress in catching up to benchmark productivity levels. Since 2001 Ofwat had implemented "catch up promoting" price caps since average regulated revenues were always below average regulatory excess costs indicating that the firms were required to eliminate at least some excess costs in order to regain economic profitability. We also emphasized that as our results also clearly demonstrated a much closer alignment between allowed revenues and benchmark costs after 2001, Ofwat's approach during this period was not only appropriate, but should also be continued in the 2009 price review. Moreover our methodology also facilitated a backward-looking approach that allowed conclusions to be drawn with regard to the effectiveness of price cap regulation. More specifically, using our methodology, regulators and policy makers can determine if past regulatory decisions have not only promoted productive efficiency by providing appropriate efficiency incentives to firms, but also whether they have led to increased allocative efficiency by aligning consumer prices more closely with efficient costs.

The key theoretic contribution in Chapter 6 was firstly to employ a panel index number technique to allow for differences in economic profitability, productivity and price performance across firms (spatial) and time (unit-specific), where the number of observations was extremely limited. Secondly and more significantly, the reconciliation of the spatial, unit-specific and relative profitability, productivity and price performance measures allowed us to decompose the unit-specific index based number profitability growth as a function of the profitability, productivity, price performance growth achieved by benchmark firms, and the catch-up to the benchmark firm achieved by less productive firms. The inclusion of quality in our analysis allowed us to further decompose the unit-specific profitability growth as a function of the catch-up in quality regarding productivity and price performance achieved by less productive firms and the quality growth of the benchmark firm.

The results indicated that after 2001, the steady decline in average price performance, gains in TFP and relatively stable economic profitability suggested that Ofwat was more focused on passing productivity benefits to consumers, and maintaining stable profitability than in the earlier regulatory periods. Average economic profitability exceeded benchmark economic profitability during the years 1991-1994 and 1999-2008, showing high levels of catch-up relative to benchmark economic profitability after 2001 which was mainly explained by the relative decline in the economic profitability of the benchmark firm. The quality unadjusted productivity results indicated that significant productivity gains for the average firm relative to the benchmark occurred after 1995 which also continued after 2000. Our results suggested that all of this catch-up can be attributed to the post 1995 period, after Ofwat first tightened price caps, and most of it can be attributed to the post 2000 period, following the even more stringent 1999 price review. Moreover, looking at the average and benchmark quality unadjusted price performance we concluded that in the post 1999 price review period, the price performance of all firms was substantially lower than in the first 10 years after privatisation. By 2000, there was a convergence in average and benchmark TPP and during the years 2001-2004, there was little or no difference between average and benchmark TPP and during the years 2005-2008 average TPP exceeded benchmark TPP showing the highest levels of price performance catch-up in 2006 and in 2008.

The quality adjusted results for productivity and price performance changes suggested that while quality improvements have contributed to the productivity performance of the WaSCs, they have also contributed negatively to their price performance. The quality adjusted TFP results indicated that after 1997 and until 2002, average quality adjusted TFP increased more rapidly than benchmark quality adjusted TFP, therefore allowing average company to catch-up in quality adjusted productivity to the benchmark firm especially during the years 2000-2005. Even after 2002 the average company achieved still significant levels of catch-up in quality adjusted productivity until 2005, which must be attributed to input usage reductions. Furthermore, the considerable increase in average profitability relative to the benchmark firm must be attributed to this catch up effect.

Moreover, the quality adjusted TPP results suggested that after 1998, there was a steady erosion of average price performance relative to benchmark price performance suggesting that there was a considerable rebalancing of regulatory price decisions in favour of the benchmark firm, which was even more dramatically extended with the implementation of the 1999 price review in 2001. It is therefore appropriate to interpret these results as substantial positive evidence demonstrating that both the 1994 and 1999 price reviews resulted in considerable movement to a regulatory price cap system consistent with a yardstick regulation regime. We would moreover offer the suggestion that this better alignment of regulated prices with the principles of yardstick regulation is likely to have contributed significantly to both the catch-up in quality adjusted productivity and the catch up in economic profitability. Furthermore, after 2001 average quality adjusted TPP fell more than benchmark quality adjusted TPP suggesting that the broad convergence after 2000 between average and benchmark firm price performance which was observed in the quality unadjusted TPP results was no longer present.

Furthermore, our methodology facilitated a backward-looking approach that allowed conclusions to be drawn with regard to the impact of the price cap regulation on the productivity, price performance and profitability of the benchmark and less productive firms. However, we strongly believe that our methodology can be further used to aid regulators in setting X-factors under price cap regulation for regulated firms. Since X-factor requires the measurement of efficiency change (catch-up) and frontier shift (technical change), our approach provides evidence for catch-up (efficiency change) in productivity by less productive firms based on the consistent spatial productivity measures across companies at any given year and also provides evidence for the productivity growth of the benchmark firm (technical change).

The key theoretical contribution in Chapter 7 was firstly to decompose profit changes over time into various factors that are of great significance for the regulator and the regulated firms such as price effect, productivity effect and activity effect. Secondly and more significantly, we accounted for differences in output characteristics like water and sewerage quality in the profit decomposition analysis. Thirdly, as in previous studies, our sequential DEA technique allows measurement of the productivity and the activity effect and their components when the number of observations is extremely limited. The profit decomposition approaches, without and after controlling for quality demonstrated similar results and the difference between the results was on the magnitude of the productivity and activity effect.

It is concluded that without and after controlling for quality, the major determinants on the quantity effect and eventually, on profit change came from the technical change whose magnitude, however, reduced during the years 2005-2008, the resource mix effect, a shift to a more cost efficient allocation of resources by substituting labour with capital, and the negative scale effect whose magnitude substantially increased after 2000, suggesting that the mergers occurred in 2000/01 did not lower costs. Finally, in both types of the profit decomposition efficiency change was found to have a small positive impact on profit change, whereas after controlling for high and low quality, the product mix effect became positive but small contributing positively to profit changes. In both cases, any substantial savings occurred by the resource mix effect were lost due to excessive mergers, indicating that over the whole period the water and sewerage companies were operating under diseconomies of scale which had a negative impact on aggregate profitability.

Our methodology facilitated a backward-looking approach that allowed conclusions to be drawn with regard to the impact of price cap regulation on the financial performance of the regulated companies when the number of observations was extremely limited. This methodology enables regulators and regulated companies to identify the sources of profit variation such as price effects, productivity effects, changes in the mix of resources, outputs and the scale of operations and aid them to evaluate firstly the effectiveness of the price cap scheme and the performance of the regulated companies. Secondly and more significantly, profit decomposition enables the regulator to identify those sources of profits that can be passed along to the consumers e.g. any improvements in productivity that could pass to the consumers in terms of lower output prices. Moreover, our methodology can also be used by the regulated companies to identify the determinants of their profit changes and improve future performance, thereby leading to future profit gains.

Moreover, our methodology showed that even if quality differentiated output price is not available, the decomposition of the productivity effect and its components, technical change and efficiency change, the resource mix effect and the input price effect are invariant, in a model that allows for quality differentiated output quantities. Furthermore, the input side of the quantity effect, the input side of the activity effect and the input side of the scale effect are invariant to the assumption of the assumption of quality undifferentiated output prices, in a model that allows for quality differentiated output quantities. In contrast, the output side of the quantity effect, the output side of the activity effect, the output side of the scale effect, the output price effect and the product mix effect will vary if quality differentiated output price is available, in a model that allows for quality differentiated output quantities.

Finally, the results from the two types of profit decomposition have significant policy implications for the regulated water and sewerage industry in England and Wales and they can be summarised as follows. Firstly, the substantial capital investment programs carried output by the water and sewerage companies since privatization leaded to the production of high quality output and the reduction of low quality output. Secondly, significant productivity improvements which contributed positively to profit changes were mainly attributed to technical change, whereas gains in efficiency were small. This finding is consistent with Cave's review (2009) findings which suggested that since privatization the main driver on productivity growth for the UK water and sewerage sector was attributed to technical change, however, our findings also suggest that technical was falling over time. Finally, the results from the profit decompositions showed that the resource mix effect was significantly large and positive over the whole period indicating that the water and sewerage industry moved to a cost efficient allocation of resources by substituting labour with capital and therefore contributing positively to profits. However, any substantial savings occurred by the resource mix effect were lost due to excessive mergers. The scale effect was negative over the whole period and substantially increased after 2000 indicating that the mergers occurred in 2000/01 had a negative impact on aggregate economic profitability. Therefore, this finding suggests that mergers were not profitable for WaSCs which is in contrast to Cave's review (2009) recommendations which suggested further mergers in the UK water and sewerage industry. We strongly believe that this finding is of great significance as it will allow further analysis on developing methodologies to explore the issue of economies of scale and scope and conclude about the most economically efficient structure and the existence of vertical integration economies in the UK water and sewerage industry (forward-looking).

8.3 Limitations of the Present Study and Further Research

The findings of this research are with limitations, but such shortcomings can motivate potential further research. The chosen techniques employed in the thesis are appropriate and consistent with the scope, sample and data available for study. Chapters 5 and 6 employed cross sectional and panel index number techniques and Chapter 7 used both index number techniques and DEA methods.

However, different techniques such as Stochastic Frontier Analysis (SFA) could be used to estimate the productivity and the activity effect and their components in Chapter 7, thereby allowing both for a stochastic element in performance measurement and a comparison of results between index number, SFA and DEA based results. Furthermore, this thesis provided a backward-looking analysis on the impact of price cap regulation on the financial performance of Water and Sewerage Companies (WaSCs) without including the Water Only Companies (WoCs).

Based on our research findings and the limitations previously mentioned, the following subsequent directions for future research are suggested. We can provide a backward-looking analysis on the effectiveness of price cap regulation on the financial performance of Water Only Companies (WoCs) since access in such information is available. Moreover, in Chapter 7 we included the impact of quality in a profit decomposition analysis by making allowances for differences in the quality of output. Further research should be taken in order to estimate the exact impact of quality or other factors like quasi-fixed inputs in a profit decomposition framework based on the approach of Fare et al, (1995) and Ouellete & Vierstraete, (2004). Furthermore, in Chapters 5 and 6 we allowed the simultaneous measurement of firm specific productivity growth, as well as spatial relative productivity measures. This approach provided evidence not only with regard to the potential productivity catch up of laggard firms, but also the potential for further improvements in benchmark productivity levels. We therefore emphasize that such an approach would further aid regulators wishing to determine appropriate X-factors for regulated firms, as it would not only provide evidence for potential productivity catch-up, as in the current approaches, but would also provide evidence for further potential productivity improvements by benchmark firms (forward-looking). Furthermore, chapter 7 showed that the resource mix effect was significantly large and positive over the whole period indicating that the water and sewerage industry moved to a cost efficient allocation of resources by substituting labour with capital and therefore contributing positively to profits. However, any substantial savings occurred by the resource mix effect were lost due to excessive mergers. The scale effect was negative over the whole period and substantially increased after 2000 indicating that the mergers occurred in 2000/01 had a negative impact on aggregate economic profitability. Therefore, this finding suggests that mergers were not profitable for WaSCs which is in contrast to Cave's review (2009) recommendations which suggested further mergers in the UK water and sewerage industry. We strongly believe that this finding is of great significance as it will allow further analysis on developing methodologies to explore the issue of economies of scale and scope and conclude about the most economically efficient structure and the existence of vertical integration economies in the UK water and sewerage industry (forward-looking).

Finally, this thesis was concerned with decomposition of profits into several factors that are of great importance for the regulator and the regulated companies by employing index number techniques and DEA but further research should focus on cost decompositions, and thereby avoid the difficulties created by the non availability of quality differentiated output prices as well.

Nevertheless despite these potential limitations and future research, we conclude by arguing that this thesis has achieved its main goal. This is because the thesis has developed models that allow a robust and comprehensive performance analysis of the WaSCs despite their small numbers, thereby serving the needs of Ofwat in assessing the performance of water and sewerage companies. Moreover, the developed models may benefit other multi-function entities with limited data sets such as electricity and gas or health services. However, we would also further emphasize that our development of three alternative profit decomposition models, will further serve the regulated sector, which typically needs to assess not only the scope for efficiency savings of very large public organizations that have been privatized, but also the appropriateness of regulated prices and profits, and their impact on allocative performance in regulated industries.

References

Antonioli B., Filippini M., "The use of a variable cost function in the regulation of the Italian water industry", *Utilities Policy*, 10, 181-187 (2001)

Anwandter L. and Ozuna T. Jr., "Can public sector reforms improve the efficiency of water utilities", *Environment and Development Economics*, 7 (4), 687-700 (2002)

Armstrong M., Cowan S. and Vickers J., "*Regulatory reform, Economic analysis and British Experience*", The MIT Press (1999)

Ashton K. J., "Cost efficiency in the UK water and sewerage industry", *Applied Economic Letters*, 7, 455-458 (2000a)

Ashton K. J., "Total Factor Productivity growth and technical change in the water and sewerage industry", *The Service Industrial Journals*, 20 121-130 (2000b)

Aubert C. and Reynaud A., "The Impact of Regulation on Cost Efficiency: An Empirical Analysis of Wisconsin Water Utilities", *Journal of Productivity Analysis*, 23, 383-409 (2005)

Averch H., Johnson L.L., "Behavior of the firm under regulatory constraint", *The American Economic Review*, 52 (5), 1052-1069 (1962)

Ball V.E., Butault J-P and Nehring R., "U.S. agriculture, 1960-96, a multilateral comparison of total factor productivity", Electric Report from the Economic Research Service (2001)

Banker R.D., Charnes A. and Cooper W.W., "Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis", *Management Science*, 30, 1078-1092 (1984)

Battese G.E. and Coelli T., "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India", *Journal of Productivity Analysis*, 3, 153-169 (1992)

Berg and Sanford V. "Introduction to the fundamentals of Incentive Regulation", *In Australian Competition and Consumer Commission and Public Utility Research Center* (eds) (1997)

Bernstein I. J. and Sappington E.M. D., "How to determine the X in the RPI-X regulation: A user's guide", *Telecommunications Policy*, 24, 63-68 (2000)

Bernstein I. J. and Sappington E.M. D., "Setting the X factor in price cap regulation plans", *Journal of Regulatory Economics*, 16, 5-25 (1999)

Bishop M. and Kay J., "Does privatization work? Lessons from the UK", Centre for Business Strategy, London Business School (1988)

Bos D., "*Privatization, a Theoretical Treatment*", Clarendon Press Oxford (1991)

Byrnes J., Crase L., Dollery B. and Villano R., "The relative efficiency of urban water utilities in New South Wales and Victoria", *Resource and Energy Economics*, In Press (2009)

Cave M., "Independent review of competition and innovation in water markets: Final report" (2009)

Caves D., Christensen L.R. and Diewet E.W., "Multilateral comparisons of output, input and productivity using superlative index numbers", *The Economic Journal*, 92 (365), 73-86 (1982a)

Charnes A., Cooper W.W. and Rhodes E., "Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444 (1978)

Coelli T. and Walding S., "Performance Measurement in the Australian Water Supply Industry: A Preliminary Analysis", In *Performance measurement and regulation of network utilities* (ed. T. Ceolli and D. Lawrence), Edward Elgar (2006)

Coelli T.J., Pao P.D.S., O' Donnell C.J. and Battese G.E., "*An introduction to efficiency and productivity analysis*", Second edition, Spinger (2005)

Corton M.L. and Berg S.V., "Benchmarking Central American water utilities", *Utilities Policy*, 17, 267-275 (2009)

Cowan S., "Competition in the water industry", *Oxford Review of Economic Policy*, 13 (1) (1997)

CRI, "*Regulated Industries: The UK Framework*" (1996)

CRI, "*Regulation of the UK Water Industry*" (1998)

Cubbin J and Tzanidakis G., "Regression versus data envelopment analysis for efficiency measurement: An application to the England and Wales regulated water industry", *Utilities Policy*, 7, 75-85 (1998)

De Witte K. and Saal D.S., "Is a Little Sunshine All We Need? On the Impact of Sunshine Regulation on Profits, Productivity and Prices in the Dutch Drinking Water Sector", *Journal of Regulatory Economics* (2009)

DEFRA, "*Productivity of UK Agriculture: Causes and Constraints*", Final Report on Project No. ER0001/3 (2003)

Diewert E.W. and Lawrence D., "Regulating electricity networks: The ABC of setting X in New Zealand", In *Performance measurement and regulation of network utilities* (ed. T. Ceolli and D. Lawrence), Edward Elgar (2006)

Diewert E.W. and Nakamura A.O, *Essays in Index Number Theory, Volume I,* Contributions to Economic Analysis Series, No 217, North Holland, Amsterdam (1999)

Diewert E.W. and Nakamura A.O., "Index number concepts, measures and decompositions of productivity growth*", Journal of Productivity Analysis*, 19 (2), 127-159 (2003)

Diewert E.W., "Exact and superlative index numbers", *Journal of Econometrics*, 4 (2), 115-145 (1976)

Diewert E.W., "Fisher ideal output, input and productivity indices revised", *Journal of Productivity Analysis*, 3 (3), 211-248 (1992)

Elteto O. and Koves P., "On a problem of index number computation relating to international comparisons", *Statisztikai Szemle*, 42, 507-518 (1964)

Emery B. "Regulating the privatised Water Industry in the UK-Theory and Practice: Is the approach suitable for CEE countries?", European Water Markets Conference Bratislava - May 2003 (2003)

Erbetta F. and Cave M., "Regulation and efficiency incentives: Evidence from the England and Wales water and sewerage industry", HERMES Working paper 8 (2006)

Estache A. and Rossi M.A., "How Different Is the Efficiency of Public and Private Water Companies in Asia?", *The World Bank Economic Review*, 16 (1), 139-148 (2002)

Fabbri P. and Fraquelli G., "Costs and structure in the Italian water industry", *Empirica*, 27, 65-82 (2000)

Fare R. and Grosskopf S., "Understanding the Malmquist productivity index", Discussion Paper No 94-5, Southern Illinois University, Illinois (1994)

Fare R., Grosskopf S. and Roos P., "Productivity and quality change in Swedish pharmacies", *International Journal of Production Economics*, 39, 137-147 (1995)

Fare R., Grosskopf S., Lindgen B. and Roos P., "Productivity developments in Swedish Hospitals: A Malmquist Output Index Approach", Discussion Paper No 89-3, Southern Illinois University, Illinois (1989)

Farrell, M.J., “The Measurement of Productive Efficiency”, *Journal of the Royal Statistical Research Society, Series A*, CXX, Part 3, 253-290 (1957)

Filippini M., Hrovatin N. and Zoric J., “Cost efficiency of Slovenian water distribution utilities: an application of stochastic frontier models”, *Journal of Productivity Analysis*, 29, 169-182 (2008)

Fox K.J., “An Economic Justification for the EKS Multilateral Index”, School of Economics, University of South New Wales, Discussion Paper 2000/3 (2000)

Fox K.J., “An Economic Justification for the EKS Multilateral Index”, *Review of Income and Wealth*, 49 (3) (2003)

Fraquelli G and Moiso V., “Cost efficiency and economies of scale in the Italian water industry”, SIEP (2005)

Garcia S. and Thomas A., “The structure of municipal water supply costs: Application to a panel of French local communities”, *Journal of Productivity Analysis*, 16, 5-29 (2001)

Garcia-Sanchez I.M., “Efficiency measurement in Spanish local government: The case of municipal water services”, *Review of Policy Research*, 23 (2), 355-372 (2006)

Green R. and Pardina M.R., “Resetting Price Controls for Privatized Utilities, A Manual for Regulators”, EDI DEVELOPMENT STUDIES OF THE WORLD BANK (1999)

Grifell-Tatje E. and Lovell C.A.K., “Profits & Productivity”, *Management Science*, 45 (9), 1177-1193 (1999)

Grifell-Tatjé, E. and Lovell C.A.K., “Productivity at the post: its drivers and its distributions”, *Journal of Regulatory Economics*, 33, 133-158 (2008)

Gromley W.T., “Privatization Revisited”, *Review of Policy Research*, 13, 215-234 (1994)

Han, S-H and Hughes A.D., “Profit Composition Analysis: A technique for linking productivity measurement & financial performance”, New South Wales Treasury Working Paper, TRP 99-5 (1999)

Hill R.J., “Constructing price indexes across space and time: The case of the European Union”, *American Economic Review*, 94 (5), 1379-1410 (2004)

Hunt C. L. and Lynk L. E., “Privatization and efficiency in the UK water industry: An empirical analysis”, *Oxford Bulletin of Economics and Statistics*, 57 (3), 371-388 (1995)

Kirkpatrick C., Parker D. and Zhang Y., "State versus private sector provision of water services in Africa: An empirical analysis", Centre on Regulation and Competition Institute for Development Policy and Management, University of Manchester, UK (2004)

Kumbhakar S.C., "Production Frontiers, Panel Data and Time-Varying Technical Inefficiency", *Journal of Econometrics*, 46, 201-211 (1990)

Lim S.H. and Lovell C.A.K., "Profits and Productivity of U.S. Class I Railroads" (2006b)

Lin C., "Service quality and prospects for benchmarking: Evidence from the Peru water sector", *Utilities Policy*, 13, 230-239 (2005)

Littlechild, S.C., "*Regulation of British Telecommunications' Profitability*" London: HMSO (1983)

Lovell C.A.K., "The Decomposition of the Malmquist Productivity Indexes", *Journal of Productivity Analysis*, 20, 437-458 (2003)

Lynk E.L., "Privatization, joint production and the comparative efficiencies of private and public ownership: the UK water industry case", *Fiscal Studies*, 14 (2), 98-116 (1993)

Mawson P., Carlaw K.I. and McLellan N., "Productivity measurement: Alternative approaches and estimates", New Zealand Treasury Working Paper 03/12 (2003)

Maziotis A, Saal D.S. and Thanassoulis E., "Regulatory Price Performance, Excess Cost Indexes and Profitability: How Effective is Price Cap Regulation in the Water Industry?", *Aston Business School Working Papers*, RP 0920 (2009)

Mosheim R., "A Shadow Cost Function Model of the US Water Industry Incorporating Water Quality and Ownership Effects", In *Performance measurement and regulation of network utilities* (ed. T. Ceolli and D. Lawrence), Edward Elgar (2006)

National Audit Office, *"Pipes and Wires"* (2002)

Neary J.P., "Rationalising the Penn World Table: True Multilateral Indices for International Comparisons of Real Income", *American Economic Review*, 94 (5), 1411-1428 (2004)

Neary J.P., "True Multilateral Indexes for International Comparisons of Real Income: Theory and Empirics", London: Centre for Economic Policy Research, Discussion Paper No. 2590 (2000)

OFWAT, *"Future charges for water and sewerage services"*, Birmingham: Office of Water Services (1994)

OFWAT, "*Future water and sewerage charges 2000-2005; Final determinations*" (1999)

OFWAT, "*Future water and sewerage charges 2005-2010; Final determinations*" (2004)

OFWAT, *"Future water and sewerage charges 2010-2015; Final determinations"* Birmingham: Office of Water Services (2009)

OFWAT, "*Report on the levels of service for the water industry in England and Wales*", Birmingham: Office of Water Services (2006)

OFWAT, "*The development of the water industry in England and Wales*" (2005)

OFWAT, "*Water and sewerage service unit; costs and relative efficiency 2004-2005 report*" (2005)

OFWAT, *July Returns for the Water Industry in England and Wales*. CD-ROM. Birmingham: Office of Water Services (2006)

Ouellete P. and Vierstraete V., "Technological change and efficiency in the presence of quasi-fixed inputs: A DEA application to the hospital sector", *European Journal of Operational Research*, 154, 755-763 (2004)

Parker D., "Editorial: Lessons from Privatization", Economic Affairs (2004)

Parker D., "Price cap regulation, profitability and returns to investors in the UK regulated industries", *Utilities Policy*, l6 (4), 303-315 (1997)

Parker D., "Privatization and Rregulation of Ppublic Uutilities: Pproblems and Cchallenges for Ddeveloping Eeconomies", In *International Handbook on Privatization* (ed. D. Parker and D. Saal), Edward Elgar (2003)

Parker D., Dassi ler T. and Saal D., "Performance Benchmarking in Utility Regulation: Principles and the UK's Experience", In *International Handbook on Economic Regulation* (ed. M. A. Crew and D. Parker), Edward Elgar (20065)

Pastor, J.T. and Lovell C.A.K., "Circularity of the Malmquist Productivity Index", *Economic Theory*, 33, 591-599 (2005)

Picazo-Tadeo A., Saez-Fernandez F.J. and Gonzalez-Gomez F., "Does service quality matter in the performance of water utilities?", *Utilities Policy*, 16 (1), 30-38 (2008)

Portela M.C.A.S. and Thanassoulis E., "A circular Malmquist-type index for measuring productivity", Aston Business School Working Paper RP08-2, Aston University (2008)

Portela M.C.A.S., Thanassoulis E., Horncastle A. and Maugg T., "Productivity change in the Water Industry in England and Wales: Application of the Meta-Malmquist Index", Aston Business School Working Paper, Aston University (2009)

Ricketts M., "Further lessons from privatization", Institute of Economic Affairs (2004)

Saal D and Parker D., "The impact of privatization and regulation on the water and sewerage industry in England and Wales: a translog cost function model", *Managerial and Decision Economics*, 21 (6), 253 – 268 (2000)

Saal D. and Parker D., "Assessing the performance of water operations in the English and Welsh Water Industry: A lesson in the Implications of Inappropriately Assuming a Common Frontier", In *Performance measurement and regulation of network utilities* (ed. T. Ceolli and D. Lawrence), Edward Elgar (2006)

Saal D. and Parker D., "Productivity and price performance in the privatized water and sewerage companies of England and Wales", *Journal of Regulatory Economics*, 20, 61-90 (2001)

Saal D., Parker D. and Weyman-Jones Thomas, "Determining the contribution of technical efficiency, and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000", *Journal of Productivity Analysis*, 28 (1), 127-139 (2007)

Sahoo B. K. and Tone K., "Radial and non-radial decompositions of profit change: With an application to Indian banking", *European Journal of Operational Research,* 196, 1130-1146 (2009)

Salerian J., "Analysing the performance of firms using a decomposable ideal index number to link profit, prices and productivity", *The Australian Economic Review*, 36 (2), 143-55 (2003)

Sarafidis V., "An assessment of comparative efficiency measurement techniques", Europe Economics, Occasional Paper 2 (2002)

Shephard R.W., "*Theory of cost and production functions*", Princeton University Press, Princeton, NJ (1970)

Stone & Webster Consultants, "Investigation into evidence for economies of scale in the water and sewerage industry in England and Wales; Final Report" (2004)

Szulc (Schultz) B.J., "Indices for multiregional comparisons", *Przeglad Statystyczny (Statistical Review)*, 3, 239-254 (1964)

Thanassoulis E, Portela M.C.S and Despic O., "Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis", In *The Measurement of Productive Efficiency and Productivity Growth* (ed. H.O. Fried, K.C.A. Lovell and S.S. Schmidt) Oxford University Press (2008)

Thanassoulis E., "Comparative performance measurement in regulation: the case of English and Welsh sewerage services", *Journal of the Operational Research Society*, 53, 292-302 (2002)

Thanassoulis E., "DEA and its use in the regulation of water companies", *European Journal of Operational Research*, 127, 1-13 (2000a)

Thanassoulis E., "*Introduction to the theory and application of Data Envelopment Analysis; A foundation with integrated software*", Kluwer Academic Publishers (2001)

Thanassoulis E., "The use of data envelopment analysis in the regulation of UK water utilities: Water distribution", *European Journal of Operational Research*, 126, 436-453 (2000b)

Tulkens H. and Vanden Eeckaut P., "Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects", *European Journal of Operational Research*, 80, 474-499 (1995)

Tupper H.C. and Resende M., "Efficiency and regulatory issues in the Brazilian water and sewerage sector: an empirical study", *Utilities Policy*, 12, 29-40 (2004)

Vickers J. and Yarrow G., "Economic perspectives on privatization", *The Journal of Economic Perspectives*, 5 (2), 111-132 (1991)

Vickers J. and Yarrow G., "*Privatization, An Economic Analysis*", The MIT Press (1988)

Vitaliano F. Donald, "Estimation of the return on capital in municipal water systems", *National Tax Journal*, LVIII (4), 685-695 (2005)

Waters II, W.G. and Tretheway M.W., "Comparing Total Factor Productivity and Price Performance: Concepts and Application to the Canadian Railways", *Journal of Transport Economics and Policy*, 33 (2), 209-220 (1999)