Designing incentives in local public utilities, an international comparison of the drinking water sector

Kristof De Witte
Centre for Economic Studies
University of Leuven (KU Leuven)
Naamsestraat 69
3000 Leuven, Belgium
kristof.dewitte@econ.kuleuven.be

Rui C. Marques
Centre of Urban and Regional Systems
Technical University of Lisbon
Av. Rovisco Pais
1049-001 Lisbon, Portugal
rcmar@civil.ist.utl.pt

January 17, 2007
Preliminary draft

Abstract

Cross country comparisons avoid the unsteady equilibrium in which regulators have to balance between economies of scale and a sufficient number of remaining comparable utilities. By the use of Data Envelopment Analysis, we compare the efficiency of the drinking water sector in the Netherlands, England and Wales, Australia, Portugal and Belgium. After introducing a procedure to measure the homogeneity of an industry, order-m partial frontiers are used to detect outlying observations. By applying bootstrapping algorithms, bias-corrected first and second stage results are estimated. Our results suggest that the regulatory and benchmark incentive schemes have a significant positive effect to efficiency. By incorporating the environmental variables directly into the efficiency estimates, we first equalize the social, physical and institutional environment and secondly deduce the effect of incentive schemes on utilities as they would work under similar conditions. The analysis demonstrates that in absence of clear and structural incentives the average efficiency of the utilities falls in comparison with utilities which are encouraged by incentives.

JEL Classification: C14, L51, L95, C61

Keywords: Data Envelopment Analysis, Bootstrapping, Incentive Scheme, Water Industry, Efficiency Measurement

1 Introduction

The merits of competition are abundantly demonstrated in economic theory. However, a monopolistic configuration may still be desirable in certain activities. Particularly operations with large sunk costs or increasing returns to scale could lead to a natural monopoly. Irrespective of ownership, whether private or public owned utilities, every natural monopoly involves welfare cost to society by creating the *quiet life* of Hicks (1935), the *X-inefficiency* of Leibenstein (1966) or making excess profits. The problem is similar to a principal-agent problem under asymmetric information. The monopolistic utilities (the agents) have private information about their ability to transform inputs into outputs. As society (the principal) wants a guaranteed service at the lowest price possible, the utilities can extract information rents.

The objective of society is to minimize the extraction of information rents while assuring a satisfactory service. Policy makers can apply a range of incentive schemes in order to reach this goal. The different institutional frameworks (divestiture, concession, yardstick competition, etc.) reflect different regulatory and ideological views. Especially within the local public utilities, ideological views prevail, mainly if the water services are deemed services of general interest and not services of general economic interest and, therefore, should not be subject to the competition law.

In this article, we want to examine the role of incentive schemes in the drinking water sector. We investigate whether regulatory and benchmark incentive schemes ameliorate the efficiency of utilities which are encouraged by incentives. We try to make abstraction of ideological conflicts by only considering efficiency. Whatever the ideological background, no one can accept inefficiencies. Inefficiencies are merely resources which are left over on the table. We compare the incentive schemes of five different countries: benchmarking the drinking water sector as in the Netherlands, privatization as in England and Wales, a strong regulatory framework as in Australia, municipal provision with private sector participation as in Portugal or different levels of public management as in Belgium.

In methodological terms, this paper follows the literature on Data Envelopment Analysis (DEA). This non-parametric technique is particularly useful in the efficiency measurement of public utilities where knowledge of the production function is relatively scarce. However, the first DEA models, developed by Charnes et al. (1978) and Banker et al. (1984), did not allow for statistical inference. Only recently, by the work of Simar and Wilson (1998), was statistical inference introduced. We apply their methodology, which is based on bootstrapping, to determine the bias-corrected first and second stage results. These outcomes are compared to the ones arising from the more traditional Tobit regressions with censored and truncated samples. Order-m efficiencies are applied to detect the outlying observations in the sample. The bandwidth of the Kernel estimates is employed to stipulate the homogeneity of a country's drinking water sector. We conclude by incorporating environmental variables directly into the efficiency estimates. Besides applying a one-stage model, we therefore introduce a procedure which is based on the residuals of the Tobit regression.

The paper is organized as follows. In section 2 we describe the institutional frameworks in the Dutch, English and Welsh, Australian, Portuguese and Belgian drinking water sector. Section 3 briefly reviews the methodology and literature on the use of DEA in water services. In section 4 we specify the DEA model and determine the homogeneity in efficiency in the national drinking water sectors. Section 5 starts with an introduction on the bootstrap methodology as outlined by Simar and Wilson (2000) and continues with describing the first stage results. Section 6 determines by the use of censored and truncated Tobit regressions and by a bootstrapping algorithm the influential environmental variables. Section 7 provides the concluding remarks.

2 The institutional framework in the water sector

There exist several approaches to solve the principle-agent problem (see e.g. Laffont and Tirole, 1993). Every government wants a secure drinking water provision at a price as low as possible. However, countries have different ideological views on the extent of state intervention in the economy. This creates various incentive schemes. In this section, we compare the implemented incentive schemes in the Netherlands, England and Wales, Australia, Portugal and Belgium.
For the ease of understanding, we first define the concepts of benchmarking, yardstick competition and sunshine regulation. Benchmarking is the process of comparing the current performance of a utility with a reference performance. It is a tool to improve performance, but not a regulatory method per se. The regulatory methods include the consequences and the effects of the use of benchmarking. Yardstick competition or competition by comparison is a regulatory method. The two existing types of yardstick competition in the water sector are the 'price yardstick competition' and 'sunshine regulation' (Marques, 2006). *Price yardstick competition* intends to define the tariffs and mainly consists of price cap or revenue cap regulation where the factor X in their formulas are determined by benchmarking techniques. *Sunshine regulation* intends to 'embarrass' the utilities that reveal a worse performance by the public discussion of their scores. Even when sunshine regulation is not triggered in a compulsory way (by a sector-specific regulator) the public display of the efficiency levels by the government or by the self-regulators generates a competitive pressure. Besides providing transparency, sunshine regulation leads to a global efficiency growth in the water sector, preventing the quiet life and the X-inefficiency. In the remaining of this paper, we identify sunshine regulation with benchmarking and a regulatory process with yardstick competition.

2.1 The Netherlands

In the late 1990s, the Netherlands were engaged in a debate about the privatization of water services. The issue was driven by the Ministry of Economic Affairs, which published in 1997 a study on prospects for utilizing market forces in the drinking water sector (Dijkgraaf et al., 1997). It concluded that privatization might reduce the price of water services by, at least, 10 percent. The water sector (i.e. the drinking water companies and the waterboards which are responsible for wastewater treatment) was strongly opposed to the privatization idea. Nevertheless, as the government was looking for more cost transparency of the operational activities in the sector, already in 1997 the drinking water companies started with a voluntary benchmark, organized by the the Dutch umbrella organization for drinking water companies, VEWIN (Vereniging van Waterbedrijven in Nederland). The Dutch water companies tried to escape government regulation by using self-regulation and, in particular, benchmark studies.
There are several approaches to compare the performance of a utility with a reference performance. One could, for example, establish a frontier production function for a utility and then calculate efficiency scores relatively to the frontier (see infra). Dijkgraaf et al. (2005) used this approach for the Netherlands. The VEWIN study detects the best-practices by a Balanced Scorecard, in which performance indicators are defined for the operating, financial, environmental and innovative perspective. VEWIN reorganized the benchmark in 2000 and 2003. The results were remarkable. During the 1997-2002 period, the efficiency increased by 9%. The re-measurement, three years later, pointed out that the efficiency gains over the 1997-2005 period increased to 21% (Waterspiegel, 2006).
Thanks to the increased transparency and efficiency by the voluntary benchmark, the Dutch government decided, in 2003, to protect the drinking water sector as a public domain. Water services are provided by government owned public limited companies (PLCs). However, through a series of mergers, stimulated by the provincial governments, many PLCs have grown to a size where they supply a substantial part of a province or more. The scale increase was initially instigated and enforced by the provinces, as they consider 100.000 connections as the minimal

size for the companies to guarantee the best services and quality at the lowest price. In the 1960s, the Netherlands counted about 200 water supply companies while in 1980 the number was reduced to about 100. There was a further reduction to 60 in 1990. In 2000 there were only 20 PLCs left for about 16 million inhabitants (Kuks, 2001). The number further declined to 13 drinking water companies at the end of 2006. Indeed, the increase in efficiency in the Netherlands, as mentioned above, can be related to the scale efficiency earnings.

The effectiveness of self-regulation by the use of benchmark studies depends on the quality of the benchmark. As the number of participating companies decreases, it becomes easier for the remaining utilities to invoke exogenous influences when they perform less efficiently according to the benchmark study. Moreover, after several years of benchmarking the same utilities, the novelty is gone. Therefore, in 2005, VEWIN started an international benchmark co-operation together with some Scandinavian companies. The objectives were twofold. Firstly, an international benchmark study could increase the learning effects, as new companies with different cultural, juridical and even technical backgrounds were included in the comparison. Secondly, the increased number of companies in the sample could compensate for the decreased number of national companies (Waterspiegel, 2005).

2.2 England and Wales

As early as in 1984, the Thatcher Government advanced plans to privatize the drinking water sector in England and Wales. After a public outcry, the plans were suspended until the reelection in 1987. By the Water Act of 1989, the ten regional water authorities which were responsible for water quality, supply and sanitation, since the nationalization of the water industry in 1974, were privatized and floated on the London stock exchange. The Water Act gave the newly established PLCs a 25 year concession for sanitation and water supply. The existing 29 private water companies were also licensed and continued to operate in their respective area (Lobina and Hall, 2001).

Privatization entailed a change in ownership, financing and regulatory structure of the industry. Three regulatory agencies were created: an environmental regulator (Environment Agency), a drinking water quality regulator (Drinking Water Inspectorate) and an economic regulator (the Office of Water Services, OFWAT). For our purpose only OFWAT is relevant. OFWAT uses a price-cap regulation which limits the annual growth rate of the water price for every water company by a factor K. The variable K is calculated as the growth rate of the Retail Prices Index (RPI) minus a productivity factor (X). The factor X is determined by comparing the performances of the water utilities. This yardstick regulatory method is done by employing econometric models and detailed assessment of individual company performance. The price cap regulation creates an incentive to increase efficiency and innovation as this will reduce expenditures in addition to the revenue allowed by the price-cap.

Originally, the price-cap was to be set every ten years. The lengthiness of the period would reduce the regulatory interference to a minimum. But as accurate forecasting of input costs turned out to be very difficult, OFWAT carried out Periodic Reviews at five-year intervals. Moreover, as OFWAT had to estimate the productivity gains and monitor the capital investment programs and the levels of service depicted, it needed to overcome the asymmetric information problem. This created the rationale for growing information requirements. Water companies in England and Wales are now more tightly regulated than any other in the privatized industries (Bakker, 2003).

An indirect effect of the improved quality and quantity of comparative information is the easy identification of potential take-over targets. The 1989 Water Act did not forbid take-overs of a water company. However, to provide the industry with a period of stability, special or 'golden' shares were issued for each of the ten privatized companies. Since there were no 'golden' shares in the already existing 29 private companies, a first wave of take-over and merger activity was

undertaken by French multinationals, already in the late 1980s. By the expiry of the government's 'golden' shares in December 1994, a second wave of merger activity arose. Only mergers which could prevent the ability of OFWAT to make comparisons were suppressed. After these two waves, the number of independent water utilities operating under independent licences decreased from 39 in 1989 to 27 a decade later (Sawkins, 2001). A tighter 1999 price review and a slowdown of the world economy prevented more mergers.
The effects of privatization are not univocal. There seem to be 'believers' and 'non-believers'. The believers argue that investments have increased significantly, compliance with environmental laws has improved and services are ameliorated. The non-believers point to non-efficient investments, the high water losses due to a still bad condition of infrastructure, the social implications of more individual meters, sharp price increases (in nominal as well as in real terms) and excessive profits (e.g. Dijkgraaf et al., 1997; Lobina and Hall, 2001).

2.3 Australia

The regulatory framework of the Australian water sector has several appealing characteristics. The Australian governments, both at state and federal levels, were able to take advantage of the strengths and weaknesses of the UK and US older regulatory models (Williamson and Mumssen, 2000). This is mainly thanks to the regulatory procedures adopted which are close to the American ideas of transparency, enactment and accountability and to the typical UK performance incentives through benchmarking and yardstick competition. Note that Australia has been the pioneer of benchmarking in the water industry. Even prior to the first American and English studies, there were already research documents and workshops about this theme in Australia. We mention the Steering Committee meetings of the Urban Water Research Association of Australia (UWRAA) in Perth in 1992, the IIR workshop in Sidney in 1993 or the UWRAA publications (e.g. Manning and Molyneux, 1993; Eggleton, 1994). The Water Services Association of Australia, replacing the UWRAA in 1995, has been carrying out several benchmarking studies. It comprises the largest Australian water services and performs a relevant role in the spread of the sector best practices.
In Australia, any industry or sector, public or private, irrespective of monopoly power, must be regulated by an independent regulatory authority. Moreover, since 1994, the Australian Government Council, in the scope of the National Competition Policy, has decided to reform the water industry and defined a clear policy and strategy for these sectors to fulfil in 10 years (until 2005). Among other measures, the reforms in these sectors aimed at its corporatization and sustainability, defining, for example, the legality of the user / payer principle and the total costs recovery.
There are some slight differences among the regulatory frameworks in the different states. We will discuss the institutions in some of the Australian states. In New South Wales, the Independent Pricing and Regulatory Tribunal (IPART), created in 1992 as an independent multisectorial regulatory authority, regulates the metropolitan water and wastewater services. The remaining water utilities (non-metropolitan) are directly regulated by the municipalities. IPART defines the price caps these entities can implement for a period of two years (IPART, 2000). The water industry of the State of Victoria is regulated by the Essential Services Commission (ESC), which took the place of the Office of the Regulator General in 2002. Until 2004, unlike other infrastructure services, ESC was not responsible for setting the water and wastewater tariffs. The regulation of the sector was only performed with regard to the quality of service by means of benchmarking (i.e. sunshine regulation). ESC sets the tariffs for a three-year period from 2005 on. By the mid-1990s there was a strong discussion about the possible privatization of the Melbourne water utility, although the option was only for a restructuring process where the original company was separated into four, one for 'bulk' water and three for water distribution to allow for yardstick competition between them. The Essential Services Commission of South

Australia (ESCOSA) replaced, in 2001, the South Australian Independent Industry Regulator. This regulatory authority, multisectorial and independent as well, does not have the responsibility for tariff setting, which is a direct governmental function. However, ESCOSA analyzes and provides advice on them (SA government, 2004; ESCOSA, 2004). Remarkably, the South Australia State, alongside the Australian Capital Territory (ACT), have the water and wastewater services with more concentration (a unique company) and where the presence of the private sector is the most significant (partnership with a private company for the system's operation). In ACT, the independent and multisectorial economic regulation is assigned to the Independent Competition and Regulatory Commission which, since 1998, defines the tariff system of the unique water and wastewater service, the ACTEW, by means of revenue cap regulation for a regulatory period of four years.

2.4 Portugal

In Portugal, except for Lisbon, the water service responsibility belonged until the 1990s exclusively to the municipalities. Only since 1993 has the private sector participation been allowed. The reform performed in this year created the 'multimunicipal' systems. These arrangements provide 'bulk' water to at least two municipalities and require a predominant investment by the State for reasons of national interest. All the remaining structures are called 'municipal systems', even though they could be managed by an association of municipalities. This regulatory reform, which is a milestone for the Portuguese water sector, includes the possibility of direct operation and management of the multimunicipal systems by the State, the municipalities or their associations. It allows for concessions of the municipal systems management and operation to companies, irrespective of capital shareholder, or to users associations. In 1998, the establishment of municipal companies was regulated according to three frameworks, corresponding to only one municipality, more than one municipality (intermunicipal company) and to one or more municipalities with a private partner with minor shareholding (mixed company). The latter is subject to a public tender. A state public company, EPAL, is responsible for the water service of Lisbon but it embodies an atypical situation in Portugal.

The strategy designed by the different governments from the 1990s onwards led the Portuguese water sector to be considered similar to the French model, despite having some particular features. Firstly, there was unbundling both in water and wastewater. About 70% of the municipalities import (or will import) water from other companies and approximately 50% of them export wastewater. Secondly, the State as entrepreneur emerged as the main player not only in the 'bulk' water supply but also in direct water distribution. The State competes with private companies in public tenders, particularly with some multinational companies. Finally, an industry-specific regulator for the water sector was established (The Institute for the Regulation of Water and Waste, IRAR).

In the last decade a growing trend towards corporatization characterized the Portuguese water sector. Private sector participation has been increased, with private water supply to about 20% of the population in 2006. In addition, municipal companies have been spreading out. In the beginning of 2006 there were around 20 municipal companies covering roughly 20% of the population. At the end of 2006, there were 299 water services in Portugal, from which 50 are companies of private type still belonging to the municipalities.

Notwithstanding the significant investment carried out over the last years, a lot of work has yet to be done. Currently, even though some occasional problems of services coverage still occur (in particular within the sewage treatment), the main challenges faced by the water sector concerns the management and operation efficiency and the systems' effectiveness associated with a possible sector restructuring. High water losses, excessive staff in the urban systems and lack of staff in the rural ones, inadequate tariff systems, inefficient assets management and unsatisfactory customer service represent some of the problems which require solutions in the short-term.

Table 1: Staff per customer in selected water services(2001)

Water service	Staff per customer (no. 10^{-3} cons.)	Ranking	Staff per volume (no. 10^{-6} m^3)	Ranking	Staff per main length (no. 10^{-1} km)	Ranking
Leiria	3.60	2	30.48	4	1.84	2
Setúbal	4.19	4	29.36	1	5.98	4
Viana	3.43	1	30.22	3	2.30	3
C. Branco	3.69	3	29.91	2	1.12	1

Source: Marques (2005)

Moreover, even if there are municipal systems in the coastal area and in the largest cities with an adequate size, in the countryside there are others which require amalgamation in order to be sustainable.

The role performed by IRAR has been widely restricted by its institutional design, particularly because it is not an independent regulatory authority. The ambiguous bounds between the State as producer and as regulator have prevented an effective regulation policy. By means of a sunshine regulation, IRAR is trying to increase the quality of service regulation. IRAR developed a set of performance indicators (PIs) to be published annually such that the operators who have a less good performance are expected to be 'embarrassed' and, consequently, to correct the deviations. As this regulatory model was only implemented in 2005, it is still impossible to make a rigorous assessment of its application.

The option for sunshine regulation as the backbone of the Portuguese regulatory model is justified by reasons such as the existing high inefficiency levels, the market structure, the politicians' interference in the regulatory process and the lack of transparency (Marques, 2006). The adoption of this approach is understood if one considers the regulator's responsibilities (only for the concessionary services either with 'bulk' water supply or direct water distribution), the existence of a contractual regulation (franchising) for these operators and the sector's fragmentation. Nevertheless, sunshine regulation will not lead to extraordinary results, especially regarding the economic efficiency if, on the one hand, it is only based on PIs and if, on the other, it just comprises the concessionary companies. Although the PIs are easily computed and have a transparent meaning, they can be misleading when taken by themselves. These indicators only assess one aspect of productivity as they rely on a single input and on a single output. Hence, under a global viewpoint, when there is a complex combination of inputs and outputs able to substitute one another, its usefulness can be slight or even harmful. Table 1 provides evidence of that situation.

2.5 Belgium

Although Belgium is a federal country and the drinking water supply has been a regional policy since 1980 (i.e. for the Flemish and the Walloon government), price regulation remains a federal issue. Within the drinking water sector, the decisions by the pricing commission are considered as rather *ad hoc* and only based on the current costs. The commission would disregard long-term visions and harmonization of the fragmented drinking water structure (Vlaams Parlement, 2001).

By law, drinking water supply is the responsibility of the municipalities. As a first organizational structure, in the Flemish as well as in the Walloon part, municipalities have organized themselves in a first organizational structure, the so-called 'intercommunales'. *Intercommunales* are a typical Belgian structure which gives the organized municipalities corporate personality. There are, respectively, 7 and 18 intercommunales in the Flemish and the Walloon part. If the municipalities, whether or not united in intercommunales, refrained from supplying drinking water to their inhabitants, the regional drinking water company (former national) provides water to this area. This regional company is called the 'Vlaamse Maatschappij voor Watervoorziening' (VMW) in the Flemish region and the 'Société Wallonne De Eaux' (SWDE) in the Walloon region. A third organizational structure corresponds to municipal drinking water suppliers. These companies have technical and financial autonomy, but their corporate personality is part of the legal personality of the municipality. They contrast with a fourth structure, municipal services, which are part of the municipal payroll and do not have financial autonomy. These services are still active in 57 Walloon municipalities.

SVW, the Flemish umbrella organization of drinking water companies, implemented a benchmark study in 2000. However, the study was never disclosed in an external report. This prevented the important 'naming and shaming' effect of public benchmarking. The question remains to which extent the drinking water companies are performing efficiently and whether they will learn from the 'best-in-class' if the results of the benchmark are not disclosed. Except for this occasional study, no other efficiency incentives are provided to the water sector (Keirsebilck and Gellynck, 2006).

The Walloon government tries to reduce the number of municipal drinking water suppliers and services by encouraging them to merge with the regional company. Yet, there remain many small local drinking water suppliers. The inducement to increase the operational scale is the only incentive to increase the efficiency of the Walloon drinking water companies.

Nevertheless, despite the encouragements of the Walloon government, the number of drinking water utilities remain very high in the Walloon as well as in the Flemish region. The drinking water companies are reluctant to increase the scale of operations which should be optimal on the principle of subsidiarity. They prefer to exchange ideas in numerous sector organizations (De Witte, 2006).

3 International benchmarking by DEA

In this study, we will 'benchmark' Dutch, English and Welsh, Australian, Portuguese and Belgian drinking water utilities against each other. One way to obtain a comparison of current performance against a reference performance (and hence to benchmark) is to assume a common frontier technology, allowing utilities from different countries to support the envelope. An alternative (as employed in the Netherlands, England and Wales and recently in Portugal) is to establish a national frontier production function, i.e. only a country's own firms may be best practices (Caves et al., 1982). We use Data Envelopment Analysis (DEA) to estimate the production frontier. We first pass some advantages of cross-country comparisons.

3.1 Cross-country comparisons

Regulators balance between economies of scale (i.e. mergers in the drinking water sector) and a sufficient number of remaining comparable companies. In this respect, cross-country comparisons offer some advantages. Firstly, studies which compare the efficiency of drinking water companies in different countries offer the possibility to escape the unsteady equilibrium between economies of scale and the number of comparators. Secondly, one can use a larger database to benchmark the national best practices. The possibility that a national best practice remains the reference in an enlarged data set decreases, which provides additional incentives to the best performing firms of a country. A third advantage arises from the potentially closer approximation to the world best-practice frontier (Estache et al., 2004). We develop a fourth advantage of cross-country comparisons. We would like to examine the effectiveness of incentive schemes objectively. Therefore, in an international dataset, we measure the efficiency of the water utilities by the use of DEA. After correcting bias in the efficiency estimates and after taking into account environmental factors, which are out of control of the management of the water utility, we calculate the average efficiency of the country. The incentive scheme of the country which

has the highest average efficiency will be considered to be superior.
However, international benchmarking raises some particular difficulties. The most intricate issue is the lack of comparability of the data as national regulators define concepts slightly different. Even in national benchmark studies, interpretation of definitions and measurement of variables could differ. Moreover, exchange rate fluctuations are important when comparing monetary units. A second concern is the unequal extent of outsourcing in the countries as this influences the number of employees (and the staff cost) in a company. A third issue is country's specific differences beyond the control of the firms. Dissimilarities such as wage rates, taxes or rates of return on capital could induce different policy options (Jamasb and Pollitt, 2001). During our cross-country comparison, we take into account all these concerns.

3.2 Determining efficiency

Efficiency measurement has an indisputable importance in any sector (Färe et al., 1985). It acquires a special significance in the water sector due to its particular characteristics (e.g. monopoly and asymmetric information). The efficiency computation aims at different goals according to the actors that perform it and to the different contexts of each country's water sector.
First of all, efficiency measurement provides relevant information to the water services management. Therefore, it could be used as a strategic tool to identify best practices and success cases and to monitor performance. The water services have all the interest in promoting studies of this kind, even if they have to associate and cooperate with each other by sharing information. This cooperation causes, most of the times, some constraints and prevents the development of efficiency measurement. From this perspective, the number of known studies is still limited. Moreover, on the one hand, there are few countries or regions with a significant number of water services with homogeneous information or conditions that enable them to perform an efficiency measurement. On the other hand, efficiency studies are rarely made public for commercial or image reasons.
Secondly, especially in recent years, efficiency research has been associated with the water services economic regulation. One of the main objectives of this kind of regulation is to improve the efficiency and productivity of the regulated companies. We described some regulatory models in section (2).
Thirdly, the efficiency measurement allows for the study of the water structure and organization. The horizontality and verticality can be assessed towards their efficiency potential. Questions such as the municipalization, regionalization or nationalization, which provide different scale earnings, should be discussed bearing in mind the efficiency measurement. The same should occur when deciding about the merging of water services with sewerage services.
The last and perhaps most ancient objective of efficiency studies regards the examination of the water services ownership (and management). These studies investigate the influence in efficiency of public and private ownership. Table 2 presents the main studies published on water services DEA efficiency.

From 1985 until the beginning of 2006, around 40 DEA applications to the water services were carried out around the world. The case-studies which were made public amount to 30. The most frequently cited studies are referred in table 2 and will be briefly described next. The objectives of these studies are diverse, although most of them focus on the water services (WS) performance measurement with regulatory aims. The protagonists are generally academic or regulatory authorities. The models entail 13 countries, namely the USA, Australia, UK, Denmark, Norway, Japan, Italy, Mexico, Portugal, Spain, Belgium, The Netherlands and Brazil. From the case-studies, 12 comprise the water supply, the sewerage and the sewage treatment altogether; 12 only the water supply; 3 the sewerage and 3 the sewage treatment separately. The

Table 2: Main DEA studies of water services

Study	Object	Focus	Results
ACT (1995)	Aust.; E&W	ACTEW performance	Significant inefficiency level
Aida et al. (1998)	108 WS from Japan	Market structure	Smaller size, more efficient
Ancarani et al. (2000)	37 WS from Italy	Italian WS performance	Inefficiency balanced by effectiveness and high quality
Ancarani (2000)	154 WS from Sicily (It.)	Market struc.-ownersh	Presence of scale and scope economies
Anwandter and Ozuna (2002)	110 WS from Mexico	Market structure	Municipalization and regulation without positive results
Bosworth et al. (1996)	10 WS from E&W	WS Regulation	Significant inefficiency level
Brynes (1986)	143 WS from the USA	Ownership	Results depend on the model
Brynes et al. (1986)	127 WS from the USA	Ownership	Indifference between public and private
Cubbin and Tzanidakis (1998)	29 WS from E&W	WS Regulation	Differences according to the computation method
Dijkgraaf et al. (1997)	WS from Netherlands	Dutch WS perfor.	WS inefficiency of 15%
KS (2003)	96 WS from Denmark	Danish WS perfor.	Sign. potential of technical and efficiency earnings
Lambert et al. (1993)	271 WS from the USA	Ownership	Public more efficient
Liang (2003)	11 WS from Australia	WS performance	Significant average inefficiency
London Economics (1995)	30 WS E&W; 6 Aust	WS performance	Aust. with high efficiency benefit from their consumption
Marques and Monteiro (2003)	45 Portuguese WS	WS performance	Private more productive
Marques and Monteiro (2004)	56 Portuguese WS	WS performance	High efficiency earnings potential
Marques and Monteiro (2005)	70 Portuguese WS	WS performance	High efficiency earnings potential
Norman and Stoker (1991)	25 WS from E&W	Market structure	Efficiency as the most important aspect
Thanassoulis (1997 and 2000a, b)	32 WS from E&W	WS Regulation	Sign. cost savings and DEA advantages in regulation
Tupper and Resende (2004)	20 Brazilian state WS	WS Regulation	Significant cost savings and YC potential
Wood et al. (1997)	WS from E&W	WS performance	Significant inefficiency level
Woodbury and Dollery (2003)	73 WS from NSW (A)	WS performance	Average inefficiency of 26.5%

30 studies mentioned correspond to 38 distinct models. These are mostly input-oriented. Only two studies concern non-oriented models. Without including the units, the studies comprise 23 inputs, 22 outputs and 20 different explanatory factors. The most frequently adopted inputs are the staff, the OPEX, the energy and the mains length. The leading outputs are the distributed (revenue) water volume, the customers number and the mains length, while the chief explanatory factors are the water source (or the associated treatment), the water volume distributed by type of customer and the density of inhabitants (or customers). Table 3 systematizes the inputs, outputs and explanatory factors which are used more than three times by at least more than one author.

The various benchmarking techniques can be classified in two major subgroups. Fully parametric methods, such as Stochastic Frontier Analysis (SFA) and (Corrected) Ordinary Least Squares (COLS), have the advantage of allowing for statistical noise, but have the disadvantage of requiring strong assumptions as to the form of the production set Ψ (which is the set of physically attainable points (x, y)) and the distribution of inputs and output vectors (x, y) over Ψ. These assumptions are called the probability model (Simar and Wilson, 2006). In semi-parametric approaches, some properties of the probability model are unspecified. Fully non-parametric methods such as DEA and Total Factor Productivity (TFP) do not require a specific analytical form which describes the frontier. This creates flexible estimators which are easily computed by the use of linear programming. Today, the statistical properties of non-parametric estimators are well established. However, as we show below, the non-parametric methods usually deem all random 'noise' to represent inefficiency (Coelli et al., 2001).

In this article we will follow the DEA methodology. In the next subsections, we explain the DEA model and indicate some advantages and disadvantages of the non-parametric technique.

3.3 Data Envelopment Analysis model

The DEA approach constructs the above mentioned non-parametric frontier as the piecewise linear combination of all efficient Decision Making Units (DMUs) in a sample. The larger the sample size, the closer the frontier will be located to the 'true' frontier.

The generic DEA model was proposed by Charnes, Cooper and Rhodes (CCR) (1978). As their model assumed constant returns to scale (CRS), Banker, Charnes and Cooper (1984) extended this to variable returns to scale (VRS). The extension involves the introduction of a convexity constraint ensuring that DMUs are only compared with 'similar' DMUs (e.g. similar size). The essential characteristic of the CCR-model is the reduction of a multiple-output / multiple-input situation for each DMU, to that of a virtual-output / virtual-input. The technical efficiency measure is calculated as this ratio of weighted outputs to weighted inputs.

Table 3: Inputs, outputs and explanatory factors adopted in the bibliographic references

Variable	Input	Output	Explanatory factor
OPEX	18		1
CAPEX	10		1
Total cost	7		
Customers number	1	17	
Mains length	15	10	
Water source/treatment	1	2	14
Staff	16	2	
Energy	11	1	
Distributed water volume		28	
Volume by customer class		3	12
Reagents costs	4		
Miscellaneous costs	5		
Other OPEX (without staff)	6		
Customers / Population density			8
Revenues		4	
Peak factor		2	6
Water losses	4	2	5

Assume there are n DMUs to be evaluated. Each of the n DMUs consumes varying amounts of m different inputs, to produce s different outputs. In particular, DMU_j consumes an amount x_{ji} of input i and produces y_{jr} of output r. We label inputs and outputs which are evaluated with an $'o'$ subscript. We will apply the input-oriented model which searches for the minimum proportion of input usage that could feasibly produce the same amount of outputs. The CRS-DEA model with input-orientation, can be expressed as:

$$\theta = \max_{\mu,\nu} \frac{\sum_{r=1}^{s} \mu_r y_{ro}}{\sum_{j=1}^{m} \nu_j x_{jo}} \tag{1}$$

subject to

$$\frac{\sum_{r=1}^{s} \mu_r y_{ri}}{\sum_{j=1}^{m} \nu_j x_{ji}} \leq 1, \qquad i = 1, 2, \ldots, n$$

$$\mu_r > 0, \nu_j > 0, \qquad \text{for all r, j}$$

In order to obtain the technical efficiency score θ for each of the n DMUs, the optimization problem needs to be repeated n times (Seiford and Thrall, 1990). The set of normalizing constraints (one for each DMU) reflects the condition that the virtual output to the virtual input ratio of every DMU is less than or equal to unity. DMU_o is efficient if and only if its efficiency score $\theta_o = 1$. An inefficient DMU is denoted by $\theta < 1$. The input efficiency measure is the reciprocal of Shephard (1970) input distance function.

Remark that if the number of DMUs in the sample increases from n to $n+p$, the only change in the model is the addition of p normalization constraints. Due to the implying reduction of the feasible solution set, the new optimal solution for any existing DMU must be less or equal to the previous optimal solution. Therefore, if we join separate datasets, the efficiency scores of the individual DMUs cannot increase in comparison with the separate analysis (Zhang and Bartels, 1998). This is an important aspect in international benchmark studies as the combination of national databases increases the number of observations.

3.4 Strengths and weaknesses of DEA

It is important to consider the strengths and the weaknesses of the DEA-methodology since other estimation techniques could yield different results. First of all, the non-parametric background of DEA does not impose an underlying functional form to the best practice frontier.

Secondly, by construction DEA can easily handle multiple inputs and outputs (even with different measurement units). Thirdly, the weights of the input and output variables are endogenously determined. Hence, we do not have to assume *a priori* a weighting scheme. A fourth advantage is the identification of 'benchmarks' as comparison elements for inefficient DMUs. Fifthly, it is possible to decompose the efficiency score in scale, congestion and pure technical efficiencies. A final advantage is the fact that the DEA technique is 'conservative'. It can be shown that the DEA estimates are upward biased and allow for specialization in an input or output variable (see below). This is a remarkable feature for regulatory use, since the regulated companies cannot complain so much.
However, the strengths of DEA lay the foundation of its weaknesses as well. As DEA is an empirically based estimation technique, it is sensible to outliers, error measurements and random influences in the data. DEA deems any deviation from the efficiency frontier to be the result of inefficiency. From the endogenous weighting system follows a second shortcoming. Unless the number of inputs and outputs is small relative to the number of DMUs, a typical large number of observations will be rated as efficient. Specialized units could be considered as efficient due to a single input or output, even though that input or output may be seen as relatively unimportant (Andersen and Petersen, 1993). Although some recent progress, a final disadvantage is the weak statistical accountability (Kittelsen, 1993; Simar and Wilson, 2000). We will tackle each of these drawbacks in the next sections.

4 Data and indicators

Choosing the input and output variables is the most important stage in any DEA assessment. The results are highly influenced by this choice. Kittelsen (1993) proposed a statistical procedure to analyze the selection of the variables. We apply his proposal to the choice of the orientation (input or output) and the option of returns to scale. As mentioned before, the main problem in international comparisons is the comparability of the data. We partly tackle this problem by considering only non-monetary variables which are less affected by purchasing power and exchange rates. However, as definitions could still slightly differ between countries, we should examine our results very carefully. In international comparisons, it is appealing to estimate the technical efficiency of companies. The goal of maximizing the technical efficiency will not be in conflict with any other goals. Indeed, inefficiencies are as resources which remain on the table. No one benefits from inefficiencies (Mobley and Magnussen, 1998).

4.1 The data

The data are obtained from various sector's organizations. One has to be very careful by the slight differences in definitions. As we are not competent to make these specifications uniform, we just copy the data from the national databases. The Dutch data are deduced from the 'Benchmark' studies and the annual 'Water Supply Statistics' organized by VEWIN. The latest year available is 2005.
The English and Welsh data are obtained from the 'June Return' by OFWAT. The 'June Return' collects information from each of the water companies. Most data tables contain information from the 1997-2005 period. As volumetric data are expressed in Mega liters per day (Ml/d), the conversion to m^3 per year could create possible rounding off errors.
The Australian data were obtained by means of WSAA facts. These documents are published annually (since 1996) by the Water Services Australian Association that compiles and audits the data. Some remaining elements needed to the research were picked from the companies websites, namely by means of their account and performance reports.
The data of the Portuguese water services were collected directly by the annual account and activity reports produced by the utilities. As some technical data were sometimes missing in

Table 4: Summary statistics, average country values 2005

	number of DMUs	average number of employees	average length of mains (km)	average volume of water (m^3)	average number of connections	connections per employee
the Netherlands	13	379	8,867	87,538,462	565,462	1490
England - Wales	23	1,306	14,540	242,703,893	913,975	699
Australia	17	464	5,450	118,735,000	340,330	862
Portugal - public	29	193	778	9,719,033	70,551	366
Portugal - private	15	91	590	4,948,958	35,793	391
Belgium*	25	226	3,550	23,924,449	136,592	604

* 2004 data

Table 5: Kittelsen test - orientation

Variable	Mean	Median	Std. Dev. Mean
Input-oriented - E^0	0.6324	0.5786	0.2389
output-oriented - E^1	0.6162	0.5733	0.2551

Method	df	Value	Probability
t-test	208	0.4764	0.6343
Anova F-statistic	(1, 208)	0.2270	0.6343
Wilcoxon signed-rank		0.6280	0.5300

the reports, the companies were contacted in order to provide them. Finally the quality of the data was checked with the information of the Portuguese Association of Water and Sewerage Services and the information on the utilities' websites.
Data on the Belgian water industry are compiled by Belgaqua, the Belgian umbrella organization, since 1993. In contrast to the other countries, the most recent year available is 2004.
Summary statistics for the various countries are presented in table 4.1. The difference in utility size is large, as revealed by the averages in the different columns. An average English and Welsh water company counts 14 times more employees than a private Portuguese firm. Also the productivity, measured by the number of connections per employee, differs significantly. A Dutch employee handles 4 times more connections than his Portuguese colleague.

4.2 Model specification

DEA models should, as much as possible, reflect the consumed resources and the produced outputs. The inputs of our DEA model consist of labor and capital. We proxy labor input by the number of employees (in full time equivalents). Measuring labor in a single aggregate variable implicitly assumes a uniform skill distribution across firms. Ideally, we should make a distinction between three categories: unskilled labor, skilled labor and management (Estache et al., 2004). However, this disaggregation seems not to be available. By including per capita Gross Regional Product in the second stage (see below), we try to control the differences in skill distributions. The length of mains (in kilometers) is used as a proxy for capital inputs. We prefer the length of mains to the 'capital expenditures' as it is easier to measure and less prone to inaccuracies

Table 6: Kittelsen test - returns to scale

Variable	Mean	Median	Std. Dev. Mean
CRS - E^0	0.5464	0.4989	0.2351
VRS - E^1	0.6324	0.5786	0.2389

Method	df	Value	Probability
t-test	208	2.6293	0.0092
Anova F-statistic	(1, 208)	6.9134	0.0092
Wilcoxon signed rank		2.6789	0.0074

Table 7: Homogeneity in efficiency

	Bandwidth	Average efficiency
the Netherlands	0.1813	0.8330
England and Wales	0.1455	0.7973
Australia	0.2244	0.7854
Portugal	0.2550	0.7467
Belgium	0.1267	0.8411
Portugal - public	0.2993	0.7232
Portugal - private	0.0040	0.9864

from variations in estimating current construction and exchange rates.
The outputs in the model reflect the main activities from the drinking water companies. The companies have to deliver water to their customers. We use this volume of delivered water as a first output indicator (in m^3). The second output variable is the number of connections.
The relative nature of DEA makes it, as in every empirically oriented methodology, vulnerable to problems with the degrees of freedom. The number of degrees of freedom will increase with the number of DMUs in the dataset, and decrease with the number of input and output variables. Banker et al. (1989) suggest a rough rule of thumb. Let m be the number of inputs and s be the number of outputs used in the analysis, then the sample size n should satisfy $n \geq \max\{m \times s, 3(m+s)\}$. This rule of thumb is satisfied in our analysis.

We use the procedure as described by Kittelsen (1993) to decide on the orientation of the DEA model. Since the sample size is wide enough, the results are not biased. Kittelsen tests whether a change in model specification significantly changes the results. If we denote the efficiency of company i measured by an input and output-oriented DEA model by, respectively, E_i^0 and E_i^1, the hypotheses can be formulated as:

$$H_0 : E_i^0 = E_i^1 \qquad H_1 : E_i^0 < E_i^1 \tag{2}$$

Several statistics are proposed to test these hypotheses. We compare the mean efficiencies by Fisher's F-distributed statistic and the ordinary t-test. The median efficiencies are compared by the Wilcoxon signed-rank test. The efficiency scores are computed by the use the statistical program R and its package 'FEAR' developed by Paul Wilson (2005). The results are presented in table 5. From the statistics, we conclude that the input-oriented model does not significantly differs from the output-oriented model. As the water utilities have to provide all the customers and they cannot encourage the consumption (demand side management policy), the input-oriented approach is preferred. In the remaining of this paper, we will only compute the input-oriented DEA-scores.
A second application of the Kittelsen procedure is to determine the returns to scale. Let E_i^0 and E_i^1 denote, respectively, the efficiency of company i in an input-oriented DEA-model with constant and variable returns to scale. We conclude from the test results in table 6, that the CRS-model significantly differs from the VRS-model. We have to make a choice with respect to the returns to scale. We prefer to apply VRS as this assumption is less stringent and ensures that DMUs are only compared with 'similar' DMUs. Besides, the water utilities cannot change their size in short-term.

4.3 Homogeneity in efficiency

By restricting the dataset to companies of the same country, we obtain a 'national efficiency comparison'. In this case, every DMU is compared with companies of its own nationality. Hence, as in De Witte (2006), we can interpret the average 'national' efficiency as a measure for the homogeneity in efficiency of a country's drinking water sector. Indeed, by construction DEA detects the relatively most efficient firms which determine the efficiency of the relatively less efficient companies. If all companies in the dataset are rather similar (homogeneous), the individual DEA efficiency scores will be higher. This results in a higher average efficiency of the country.

Zhang and Bartels (1998) point out that one cannot simply compare the average efficiencies. The technical efficiency score of a DMU will on average decrease as the sampling size increases. In order to equalize the size of the datasets, we will resample the efficiency scores. For every country, we first compute the input-oriented DEA-VRS efficiency scores as described in section (4.2). Secondly, by the use of the statistical package R and its code 'FEAR' (Wilson, 2005), we determine the bandwidth suitable for use in Kernel estimates of densities of efficiency estimates that are bounded above at one. One approach to compute the bandwidth is the unbiased cross-validation which minimizes the estimate of the mean-integrated square error (see Simar and Wilson, 2006). The bandwidth of every country is presented in table 7. In a last step, we resample the original DEA-VRS efficiency scores. We obtain n values drawn from a Kernel estimate of the bounded density of the efficiency estimates. We set n equal to 44, the size of the largest dataset. The average resampled efficiency is presented in the third column of table 7. The sample size bias could have been avoided by the use of order-m efficiency scores as well (see below). We apply the order-m efficiencies to determine the outliers below.

It turns out that the efficiency in the Belgian drinking water sector is the most homogeneous one. Belgium is closely followed by the Netherlands. In those two countries, it should be relatively easy for policy makers to adopt new laws which are generally approved by all water companies. Portugal ends as the most heterogeneous country in efficiency. Nevertheless, the high heterogeneity can especially be attributed to the public sector. The efficiency of private Portuguese drinking water companies seems to be very similar to each other.

In determining the efficiency of an industry as a whole, the average efficiency of all DMUs can have a reduced meaning. Farrell (1957) points out that the industry average should be computed as a weighted average based on the outputs (or on the inputs). However, when several outputs (inputs) exist Farrell does not refer how they are weighted and if we should use the observed outputs (inputs) or the target outputs (inputs). Another measure to estimate the average efficiency of an industry (called structural efficiency) is to consider the average DMU as suggested by Førsund and Hjalmarsson (1979) in the set of observations. As we created pseudo-samples in this homogeneity exercise, we are not able to weight the efficiency scores. We introduce a weighting system if we discuss the first stage results.

4.4 Outlier detection

Theoretical background

A major drawback of DEA can be found in its deterministic nature in which the frontier model assumes that

$$Prob((x, y) \ \epsilon \ \Psi) = 1 \tag{3}$$

where Ψ denotes the attainable set ($\Psi = \{(x, y) \ \epsilon \ \mathbb{R}_+ | x \text{ can produce } y\}$). DMUs located in the interior of Ψ operate technically inefficient, while firms on the boundary of Ψ are technically efficient. Equation (3) states that deterministic models do not allow for outliers. Sexton et al. (1986) conclude that "DEA results are likely to be unstable because its evaluations are based on outlier observations". This could be the case if the outlying points are determined as efficient observations and hence make part of the frontier. The bigger the number of efficient DMUs, the weaker the relationship between the number of DMUs and the dimensionality of the model (inputs + outputs). Likewise, with a tighter conditioning of the technology that characterizes the model (VRS instead of CRS, or weak disposability instead of strong disposability), the number of technically efficient DMUs will be larger. However, it is easily observed that not all the efficient observations have the same 'importance' in the sample.

Indeed, some efficient DMUs are outliers. Outlying observations could be attributed to measurement errors, noise and influential observations (i.e. atypical data). As our data are obtained from national regulators and sector organizations, we consider the data as more or less free from measurement errors. In order to eliminate noise from the dataset and draw statistical inference, we apply bootstrapping procedures in the next section. Influential data affect the efficiency results of a significant number of other DMUs (Charnes et al., 1985). In other words, part of the efficient DMUs are the peers of other DMUs, while the remaining efficient DMUs are just peers of themselves or of a reduced number of DMUs. Actually, the identification of influential efficient DMUs becomes fundamental in DEA analysis, specially if they can be considered outliers, and for that reason can be taken out of the sample, or if they are regarded as 'true' benchmarks, and are therefore essential to the benchmarking analysis. The opposite case of outliers presence, but with inefficient DMUs, has little effect in the analysis, except with regard to that DMU itself. We will neglect this case here. As the best way to identify outliers (to be sure of them) is to consider several techniques, we explain and apply five outlier detection procedures.

One of the easiest formulas of determining outliers or the influential DMUs is the computation of the 'peer count index'. As suggested by Charnes et al. (1985) the computation of the number of times a given efficient DMU is peer of an inefficient DMU can be taken as an indicator that proves that a DMU is atypical either as outlier or as best practice.

Andersen and Petersen (1993) developed a second methodology to sort the efficient DMUs, by means of which the super-efficiencies of the efficient DMUs are computed. Super-efficiency calculates to what extent the efficient DMUs can increase their inputs by keeping themselves technically efficient (input-oriented), or vice-versa, reduce their outputs and at the same time continue to be efficient (output-oriented). In numerical terms, the procedure consists in taking out the efficient DMUs themselves at the moment of their evaluation. With a reference set which does not include DMU_i, its efficiency can be greater than 1. Observations with high values are suspected to be outliers.

Yet, in spite of sorting the efficient DMUs with regard to the efficiency surpluses, super-efficiency does not say anything about their sorting according to the importance of the efficient DMUs as reference or benchmarking element for the inefficient DMUs of the sample. A hypothesis of measuring the suitability of the efficient DMU to be best practice consists of computing the indicator ρ, called *the peer index*, of the efficient DMU_j for the input k, represented by the following expression (Torgersen et al., 1996):

$$\rho_j^k = \frac{\sum_{i \varepsilon N} \lambda_{ij}(x_{ki}^P - x_{ki})}{x_k^P - x_k} \tag{4}$$

where λ_{ij} regards the weight of the efficient DMU_j for the inefficient DMU_i, x_{ki} the input k of DMU_i and x_{ki}^P represents the target (score at the frontier) for the input k of DMU_i. The measure ρ_k^k expresses the percentage of the potential reduction of an input k that is represented by the inefficient DMUs which depend on the efficient DMU_j. The higher the ρ_k^k, the larger the possibilities of employing that DMU for benchmarking or in other perspective the larger the possibility that it is an outlier.

To find influential and deviant observations, Wilson (1993) uses in his descriptive model the relative change due to the deletion of i observations from the sample. As a multi-output extension of the geometric influence function $R_L^{(i)}(XY)$ of Andrews and Pregibon (1978), the graphical analysis of log ratios $(\log(R_L^{(i)}(XY)/R_{min}^{(i)}))$ examines the separation between the smallest ratios. This ratio is computed for each of the possible subsets L of size i. The choice of i, the stopping point of the analysis, is arbitrary but involves a dramatically increasing computational burden (as there are $\binom{n}{i}$ combinations). Nevertheless, to avoid a 'masking effect' by which one outlier could be hidden behind another with similar values, i should be large enough.

A fifth and more recent methodology for detecting outlying observations is based on the 'expected frontier of order-m'. The concept is developed by Cazals et al. (2002) and extended to outlier

Table 8: Outlier detection: traditional methods and Wilson (1993)

Peer count index		Super-efficiency		Peer index	Employees	Mains	Wilson
Brisbane	1	Brisbane	1,033	Brisbane	0,011	0,009	Anglian
Calamine	46	Calamine	1,025	Calamine	0,090	0,070	Dwr Cymru
Chimay	9	Chimay	1,210	Chimay	0,020	0,014	Severn Trent
City West	44	City West	1,243	City West	0,068	0,070	South West
Dwr Cymry	24	Dwr Cymry	8,296	Dwr Cymry	0,040	0,075	SWDE
DZH	20	DZH	1,114	DZH	0,103	0,132	Thames
Evides	16	Evides	1,057	Evides	0,143	0,184	V.M.W.
Hulpe	10	Hulpe	1,200	Hulpe	0,005	0,009	Vitens
IWVA	29	IWVA	1,749	IWVA	0,077	0,026	
Lisboa	35	Lisboa	3,720	Lisboa	0,074	0,042	
Thames	14	Thames	infeasible	Thames	0,128	0,161	
TWM	52	TWM	1,094	TWM	0,063	0,080	
Waimes	4	Waimes	1,250	Waimes	0,000	0,001	
WLB	53	WLB	1,203	WLB	0,114	0,071	
Yarra	12	Yarra	1,052	Yarra	0,065	0,056	

Table 9: Outlier detection: Simar (2003)

	m=10	St.E.		m=40	St.E.		m=60	St.E.		m=80	St.E.
Dwr Cymry	19.57	12.27	Dwr Cymry	7.722	7.63	Dwr Cymry	4.660	5.90	Dwr Cymry	3.681	5.24
Lisboa	10.06	6.35	Lisboa	4.560	3.12	Lisboa	3.415	2.92	Lisboa	2.748	2.63
Thames	6.505	6.06	Thames	2.767	1.64	City West	2.270	1.70	City West	1.979	1.59
City West	5.710	3.45	City West	2.623	1.85	WLB	1.984	1.07	Thames	1.693	0.78
Coliban	5.461	3.33	WLB	2.256	1.18	Thames	1.875	1.04	A.W.W.	1.686	0.84
South East	4.455	3.71	A.W.W.	2.123	1.00	A.W.W.	1.812	0.90	WLB	1.648	0.95
Yarra	4.391	3.54	Brisbane	2.079	0.91	Gold Coast	1.725	0.87	Gold Coast	1.566	0.81
Brisbane	4.332	3.06	Gippsland	2.066	1.34	Brisbane	1.715	0.75	Brisbane	1.528	0.69
Gippsland	4.175	2.52	Gold Coast	1.986	1.00	Gippsland	1.697	1.05	Sidney	1.509	0.59
Sidney	4.102	4.23	Portsmouth	1.973	1.05	Portsmouth	1.666	0.87	Three Val.	1.432	0.44
WLB	4.097	1.88	Coliban	1.947	2.03	Sidney	1.588	0.65	Portsmouth	1.409	0.72
Yorkshire	4.045	5.12	South East	1.875	0.98	South East	1.573	0.68	South East	1.408	0.58
United Util.	3.993	4.04	Porto	1.742	1.09	Three Val.	1.498	0.49	United Util.	1.368	0.48
Portsmouth	3.886	1.71	United Util.	1.737	0.76	United Util.	1.483	0.64	Yorkshire	1.342	0.44
Severn Trent	3.730	3.89	Sidney	1.719	0.68	Oeiras	1.461	0.68	South Staffs	1.327	0.39

detection by Simar (2003). Instead of using all the observations to determine the efficient frontier (i.e. a full frontier), the order-m partial frontier uses a sample of size m which is drawn from the total sample with size n. Whereas a full frontier indicates for all firms which produce at least level y of outputs the minimum achievable lower boundary of inputs, the expected frontier function of order-m is the expected minimal input achieved by any m firms drawn from the population of firms which produce at least y outputs (Simar, 2003). In this partial frontier, the value of m can be considered as a trimming parameter, since the estimator does not envelop all data points. This makes the order-m frontier more robust to outlying observations. With an order-m input oriented frontier, an observation which lies far above the frontier (i.e. a value considerably larger than 1) will be determined as an outlier. Observations near, on or below the frontier are considered as efficient. The relative (in)efficiency of a DMU is influenced by the value of m (see below). The order-m method is little sensitive to the 'masking effect'. This could be attributed to its construction: the efficiency is computed as a conditional mean of a minimum among a random sample of m DMUs with the same characteristics (variables) as the DMU analyzed (Beguin and Simar, 2004):

$$\phi_m(y) = E[\min(X^1, \ldots, X^m)|Y \geq y] \tag{5}$$

Detecting outliers

The results of the peer index, super-efficiency and peer count index are presented in table 8. These methods can be considered as the most traditional outlier detection procedures. There seems to be a high consensus among the three methods, although, concerning the super-efficiency

we can only label Dwr Cymry, Lisboa and Thames for sure as suspected outliers (the other observations have a lower super-efficiency value). The three methods identify the same 5 Belgian, 4 Dutch, 3 Australian, 1 English, 1 Welsh and 1 Portuguese firms as possible outliers. These outlying observations differ from the Wilson (1993) analysis in which we equalized i to 12. As the separation is relatively large for i=1,5 and 8, we regard the corresponding observations as outliers (details are available upon request). The suspected outliers are presented in the last column of table 8. As the order-m results, presented in table 9, are influenced by the value of m we compute the order-m efficiency score for different values of m. Following the example by Simar (2003) we use 200 Monte-Carlo replications in computing the estimates. As it is difficult to decide on an appropriate value from which on an observation should be determined as an outlier (i.e. what is considerably larger than one?), we consider the 15 most outlying observations as outliers. On average, there are 6 DMUs which are stipulated as outliers by all 4 methods.
As the results of the peer index, sensitivity analysis and peer count index are closely related, we consider these procedures as more robust. In the remaining of this article, we eliminate from the sample of 122 observations the 15 outlying DMUs as determined by the more traditional procedures. Hence, we obtain a dataset of 107 observations.

5 First stage analysis

5.1 Bootstrap method

The deterministic nature of DEA creates several problems. Above, we dealt with the aspect of influential observations. In this section, we will tackle the problem of noise in the data. Although the applied researcher can only estimate the *observed* production frontier by the use of DEA, the literature interprets the estimates as the *true* frontier. Simar and Wilson (1998, 2000) make a clear distinction between the *true* (e.g. $\theta(x, y)$) and the *estimated* concepts (e.g. $\hat{\theta}(x, y)$). The DEA efficiency estimates are prone to uncertainty due to sampling variation. By the use of a bootstrap methodology, Simar and Wilson allow to carry out traditional statistical inference in DEA.
The bootstrap procedure, as invented by Efron (1979), is useful if the sampling properties of estimators are difficult to obtain analytically. The bootstrap approximates the sampling distribution by reproducing the data generating process (DGP). This is the statistical model which describes the process that yields the observed data in the sample. The DGP follows the principle that, restricted to the relations between inputs and outputs, the stochastic elements in the productive process are totally encompassed by the random inputs efficiency measures (hence, we do not assume measurement errors). This makes the DEA estimators biased by construction as the estimate of the production set $\hat{\Psi}$ is part of the real attainable set Ψ: $\hat{\Psi} \subseteq \Psi$. Therefore, the efficiency score, $\hat{\theta}(x, y)$, is an upward-biased estimator of the true efficiency score $\theta(x, y)$ (see Simar and Wilson (2006) for an extensive discussion). The difference is visualized in figure 1. The bootstrap mimics this estimation and creates a pseudo frontier from which it provides estimates of the sampling distributions of the bias term $\hat{\theta}(x, y) - \theta(x, y)$ (see figure 2).
For practical reasons we invert the efficiency scores:

$$\hat{\delta}(x, y) = \frac{1}{\hat{\theta}(x, y)} \tag{6}$$

Indeed, as $\hat{\delta}(x, y) \geq 1$ for all $(x, y)\epsilon\Psi$, we only have to deal with one boundary condition for $\hat{\delta}$, not two as in the case of $\hat{\theta}$.
The literature contains several approaches to simulate a bootstrap sample χ_n^*. Kneip et al. (2003) describe the properties of the naive bootstrap, the subsampling of DEA scores and a smoothing technique. We will briefly introduce the homogeneous smoothed bootstrap. In this approach, we assume the distribution of the efficiency scores to be homogeneous over the input-output space (compare with a homoskedasticity assumption in linear regression models). This

allows us to base the bootstrap on the sample estimates $\hat{\delta}_i(x_i, y_i)$. The bootstrap algorithm, in accordance with Simar and Wilson (2006), follows eight steps. In the first step, we compute from the original dataset χ_n for every DMU the DEA efficiencies $\hat{\delta}_i$. In later steps, we will apply a standard Kernel density estimator. This estimator belongs to the class of non-parametric density estimators which do not have a fixed structure and use all the observations to reach an estimate. As the quality of the Kernel estimate depends on the appropriateness of the bandwidth h, in a second step we select the optimal value of h. We will only take into account the non-efficient observations ($\hat{\delta}_i \neq 1$) in order to adjust the discrete distribution to mimic a continuous underlying distribution of efficiencies. Several approaches could be used to estimate the bandwidth. As in the estimation of the homogeneity in efficiency of a country's drinking water sector, we use the unbiased cross-validation method to estimate h (see Simar and Wilson, 2006). In a third phase, we generate naive bootstrap pseudo-data $\beta_1^*, \ldots, \beta_n^*$ from a set $\mathcal{D}_{2n} = \{\hat{\delta}_1, \ldots, \hat{\delta}_n, (2 - \hat{\delta}_1), \ldots, (2 - \hat{\delta}_n)\}$. The naive bootstrap draws independently, uniformly and with replacement pseudo-data from the dataset $\mathcal{D}_{2n}$. In step four, the sample of pseudo-efficiencies as drawn in step three are perturbed by the use of draws ϵ_i^* from a Kernel function such that $\beta_i^{**} = \beta_i^* + h\epsilon_i^*$. After correcting some bias in β_i^{**} (which is typical in Kernel techniques), in phase five we compute the bootstrap efficiencies δ_i^* as $\delta_i^* = 2 - \beta_i^{**}$ if $\beta_i^{**} < 1$ or $\delta_i^* = \beta_i^{**}$ otherwise. In the next stage, we create a bootstrap sample χ_n^* of input-output levels for each DMU where the new input-level is defined by $x_i^* = \delta_i^* \hat{\delta}_i^{-1} x_i$. Hence, the bootstrap outputs remain the same, but the input levels are obtained by first projecting the original input level to the efficient level and next to an inefficient level. In step seven, the bootstrap efficiency $\hat{\delta}^*(x, y)$ is computed by DEA using the set of n bootstrap input-output levels in χ_n^*. Finally, the steps 3 to 7 are repeated B times, so that a set of B values of efficiency estimates can be computed for each DMU: $\{\hat{\delta}_b^*(x, y) | b = 1, \ldots, B\}$. Hall (1986) recommends a B minimal value equal to 1000. We show the different frontiers in figures 2 and 3.

Having defined the bootstrap efficiencies, we can construct a bias-corrected estimator of $\delta(x, y)$. Therefore, in a first phase, we estimate the bootstrap bias of $\hat{\delta}(x, y)$:

$$\widehat{BIAS}_B(\hat{\delta}(x, y)) = B^{-1} \sum_{b=1}^{B} \hat{\delta}^*(x, y) - \hat{\delta}(x, y) \tag{7}$$

The first term on the right hand side corresponds to the average of the bootstrap efficiency result and the second term to the original DEA estimate. In a second phase, the bias-corrected estimator is computed as

$$\hat{\hat{\delta}}(x, y) = \hat{\delta}(x, y) - \widehat{BIAS}_B(\hat{\delta}(x, y)) \tag{8}$$

This bias correction introduces an additional source of noise as the bias is again an estimate. Therefore, we have to check whether the mean-square error (MSE) of $\hat{\hat{\delta}}(x, y)$ will not be too large. This is the case if the MSE of $\hat{\hat{\delta}}(x, y)$ is larger than the MSE of $\hat{\delta}(x, y)$. As a rule of thumb, the sample variance σ^2 of the bootstrap values $\hat{\delta}^*(x, y)$ should respect (Simar and Wilson, 2000):

$$\sigma^2 = B^{-1} \sum_{b=1}^{B} \left[\hat{\delta}^*(x, y) - B^{-1} \sum_{b=1}^{B} \hat{\delta}^*(x, y) \right]^2 < \frac{1}{3} \left[\widehat{BIAS}_B(\hat{\delta}(x, y)) \right]^2 \tag{9}$$

Efron (1982) suggests that a relationship between the bias estimate and the standard error higher than 0.25 is meaningful and consequently the bootstrap results should not be rejected. Acknowledging the bootstrap efficiencies empirical distribution function $\hat{\delta}_{ib}^*$ (with $b = 1, \ldots, B$) and after the bias correction, confidence intervals can be obtained, for example, by means of the percentile method. We follow the procedure as outlined by Simar and Wilson (2000). After sorting the B estimates of the bias term $\hat{\delta}^*(x, y) - \hat{\delta}(x, y)$ in increasing order and deleting the $(\frac{\alpha}{2} \times 100)$-percent of the observations at both ends of the list, we relabel the endpoints of the

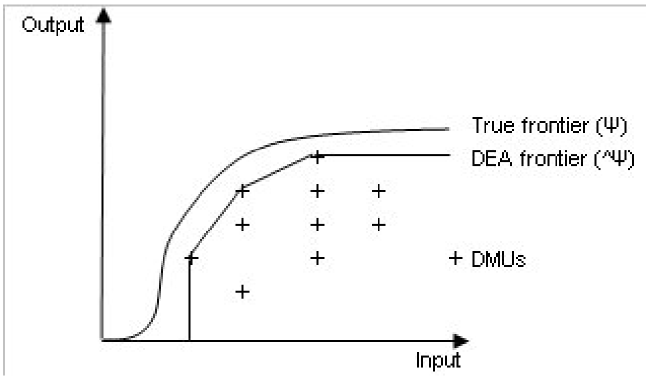

Figure 1: The true and the DEA frontier

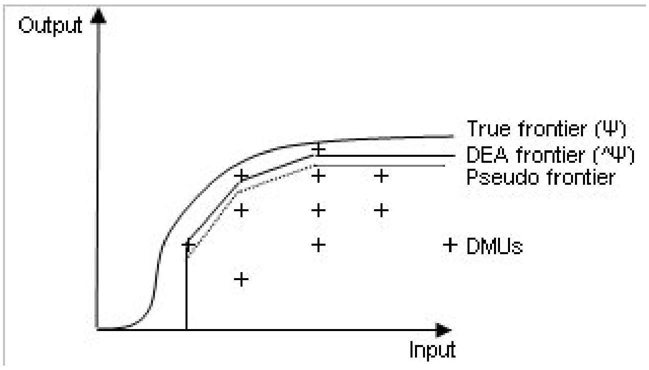

Figure 2: Bootstrap idea

new row as the estimated borders of the confidence interval ($\hat{c}_{\alpha/2}$ and $\hat{c}_{1-\alpha/2}$). We can indicate the estimated $(1-\alpha) \times 100$-procent confidence interval for $\theta(x, y)$ by:

$$\frac{1}{\hat{\delta}(x,y) - \hat{c}_{\alpha/2}} \leq \theta(x,y) \leq \frac{1}{\hat{\delta}(x,y) - \hat{c}_{1-\alpha/2}} \tag{10}$$

We repeat this algorithm n times to produce for every observation of the sample a confidence interval of the form (10).

5.2 First stage results

Having cleared the dataset from outlying observations, we can proceed to the first stage analysis. In this step, we will use an input-oriented DEA-VRS-model with input and output variables as defined in section (4.2). The results, presented in table 10, are computed by the DEA software of Zhu (2003). The average efficiency amounts to 69.8%. This indicates that an average DMU

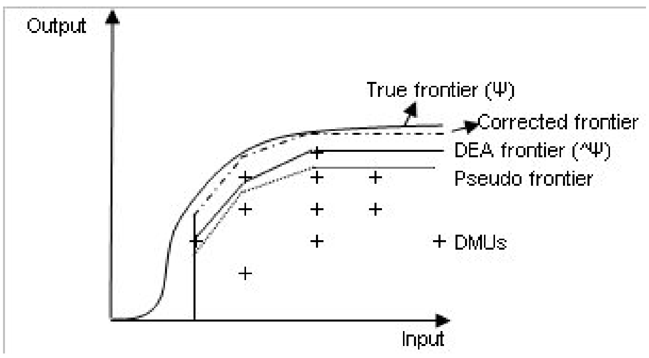

Figure 3: Bootstrap idea (2)

Table 10: First stage results

DMU	DMU Name	Efficiency	Benchmarks							
1	Actew	0.619	0.086	DMU 03	0.209	DMU 28	0.510	DMU 29	0.196	DMU 58
2	Barwon	0.488	0.104	DMU 03	0.061	DMU 28	0.625	DMU 29	0.210	DMU 58
3	Gippsland	1.000	1.000	DMU 03						
4	Highlands	0.372	0.049	DMU 03	0.026	DMU 12	0.861	DMU 29	0.064	DMU 58
5	Coliban	1.000	1.000	DMU 05						
6	Gold Coast	0.910	0.154	DMU 03	0.179	DMU 28	0.330	DMU 29	0.337	DMU 58
7	Gosford	0.699	0.790	DMU 29	0.078	DMU 58	0.098	DMU 66	0.034	DMU 69
8	Goulburn	0.598	0.309	DMU 03	0.001	DMU 28	0.629	DMU 29	0.060	DMU 58
9	Hunter	0.667	0.399	DMU 03	0.009	DMU 28	0.220	DMU 29	0.372	DMU 58
10	Power Water	0.713	0.056	DMU 03	0.287	DMU 28	0.657	DMU 29		
11	SA Water	0.716	0.016	DMU 13	0.515	DMU 28	0.236	DMU 58	0.234	DMU 62
12	South East	1.000	1.000	DMU 12						
13	Sidney	1.000	1.000	DMU 13						
14	Water Corporation	0.673	0.289	DMU 13	0.470	DMU 28	0.165	DMU 58	0.077	DMU 69
15	Wggr	0.753	0.290	DMU 12	0.174	DMU 21	0.536	DMU 29		
16	Wmd	0.612	0.232	DMU 12	0.089	DMU 21	0.679	DMU 29		
17	Vitens	1.000	1.000	DMU 17						
18	Pwn	0.933	0.603	DMU 12	0.170	DMU 21	0.227	DMU 62		
19	Oasen	0.910	0.536	DMU 21	0.418	DMU 35	0.016	DMU 58	0.030	DMU 66
20	Hydron-Fl	0.764	0.151	DMU 12	0.058	DMU 21	0.792	DMU 29		
21	Hydron-MN	1.000	1.000	DMU 21						
22	Brabant Water	0.917	0.533	DMU 12	0.183	DMU 17	0.284	DMU 62		
23	WML	0.709	0.047	DMU 12	0.877	DMU 21	0.076	DMU 29		
24	A.I.E.	0.783	0.247	DMU 29	0.732	DMU 35	0.001	DMU 58	0.020	DMU 66
25	Aquasambre	0.598	0.078	DMU 21	0.669	DMU 35	0.253	DMU 66		
26	A.I.E.C.	0.443	0.013	DMU 12	0.987	DMU 41				
27	A.I.E.M.	0.458	0.008	DMU 21	0.949	DMU 29	0.043	DMU 35		
28	A.W.W.	1.000	1.000	DMU 28						
29	Bastogne	1.000	1.000	DMU 29						
30	Bullingen	1.000	1.000	DMU 30						
31	C.I.E.V.	0.880	0.579	DMU 29	0.335	DMU 30	0.086	DMU 41		
32	Cile	0.673	0.283	DMU 21	0.356	DMU 35	0.361	DMU 66		
33	I.E.C.B.W.	0.527	0.063	DMU 12	0.022	DMU 21	0.915	DMU 29		
34	Idemls	0.582	0.087	DMU 21	0.789	DMU 35	0.124	DMU 66		
35	I.E.G.M.	1.000	1.000	DMU 35						
36	Inasep	0.414	0.007	DMU 12	0.031	DMU 21	0.961	DMU 29		
37	Pidpa	0.455	0.182	DMU 12	0.622	DMU 21	0.196	DMU 29		
38	Wavre	0.986	0.002	DMU 12	0.010	DMU 21	0.988	DMU 29		
39	St.-Vith	0.712	1.000	DMU 29						
40	Swde	0.359	0.689	DMU 21	0.279	DMU 62	0.033	DMU 66		
41	Theux	1.000	1.000	DMU 41						
42	Tmvw	0.611	0.679	DMU 21	0.133	DMU 35	0.030	DMU 58	0.158	DMU 66
43	V.M.W.	0.525	0.301	DMU 12	0.125	DMU 17	0.573	DMU 62		
44	Anglian	0.675	0.171	DMU 13	0.356	DMU 54	0.473	DMU 62		
45	Bristol	0.775	0.027	DMU 21	0.099	DMU 35	0.832	DMU 58	0.042	DMU 66
46	Bournemouth	0.774	0.746	DMU 03	0.080	DMU 12	0.175	DMU 29		
47	Cambridge	0.617	0.040	DMU 12	0.006	DMU 21	0.794	DMU 29	0.159	DMU 58
48	Dee Valley	0.581	0.722	DMU 29	0.167	DMU 58	0.050	DMU 66	0.062	DMU 69
49	Essex Suffolk	0.931	0.132	DMU 13	0.326	DMU 62	0.542	DMU 66		
50	Folkestone	0.748	0.009	DMU 28	0.858	DMU 29	0.116	DMU 58	0.017	DMU 69
51	Mid Kent	0.625	0.070	DMU 28	0.513	DMU 29	0.412	DMU 58	0.005	DMU 69
52	South East water	0.743	0.103	DMU 03	0.499	DMU 12	0.337	DMU 58	0.061	DMU 62
53	Northumbrian	0.740	0.233	DMU 13	0.521	DMU 62	0.246	DMU 66		
54	United Utilities	1.000	1.000	DMU 54						
55	Portsmouth	0.975	0.011	DMU 21	0.415	DMU 35	0.500	DMU 58	0.074	DMU 66
56	Sutton Surrey	0.826	0.003	DMU 21	0.404	DMU 35	0.435	DMU 58	0.158	DMU 66
57	Southern	0.830	0.519	DMU 13	0.481	DMU 66				
58	South Staffs	1.000	1.000	DMU 58						
59	Severn Trent	1.000	1.000	DMU 59						
60	South West	0.410	0.737	DMU 03	0.151	DMU 13	0.112	DMU 62		
61	Tendring Hundred	0.814	0.048	DMU 21	0.894	DMU 35	0.030	DMU 58	0.028	DMU 66
62	Three Valleys	1.000	1.000	DMU 62						
63	Wessex	0.438	0.202	DMU 13	0.006	DMU 62	0.792	DMU 66		
64	Yorkshire	0.981	0.476	DMU 13	0.321	DMU 17	0.202	DMU 54		
65	Sintra	0.613	0.006	DMU 13	0.994	DMU 66				
66	SMAS Oeiras	1.000	1.000	DMU 66						
67	SMAS Loures	0.662	0.099	DMU 29	0.009	DMU 58	0.775	DMU 66	0.117	DMU 69
68	Aguas Gaia	0.696	0.086	DMU 21	0.572	DMU 35	0.342	DMU 66		
69	SMAS Porto	1.000	1.000	DMU 69						
70	Vimagua	0.576	0.206	DMU 29	0.661	DMU 35	0.026	DMU 58	0.107	DMU 66
71	Almada	0.724	0.362	DMU 29	0.076	DMU 66	0.561	DMU 69		
72	Coimbra	0.530	0.675	DMU 29	0.065	DMU 58	0.078	DMU 66	0.182	DMU 69
73	Xira	0.836	0.637	DMU 29	0.237	DMU 66	0.126	DMU 69		
74	Maia	0.722	0.101	DMU 29	0.684	DMU 35	0.215	DMU 66		
75	Leiria	0.502	0.020	DMU 21	0.870	DMU 35	0.021	DMU 58	0.090	DMU 66
76	Viseu	0.401	0.735	DMU 29	0.163	DMU 35	0.027	DMU 58	0.075	DMU 66
77	Castelo	0.491	0.006	DMU 21	0.917	DMU 35	0.076	DMU 66		
78	Aveiro	0.422	0.829	DMU 29	0.118	DMU 66	0.052	DMU 69		
79	Torres Vedras	0.345	0.005	DMU 21	0.909	DMU 35	0.086	DMU 66		

could decrease its inputs by 30.2% while keeping its outputs constant, if it would perform as efficient as its benchmark(s). A benchmark or best practice is a company which performs technically efficient and hence makes part of the DEA frontier. Out of the 107 observations, there are 17 efficient DMUs (15.8%). These companies originate from Belgium (5), England and Wales (4), Australia (4), the Netherlands (2) and Portugal (2). A histogram showing the estimates is presented in figure 4. A DMU with an efficiency score less than 1 is relatively inefficient with respect to its benchmarks. We represent the benchmark(s) of every DMU in table 10 where the

Table 10: First stage results - continued

DMU	DMU Name	Efficiency	Benchmarks							
80	Santarem	0.371	0.864	DMU 29	0.019	DMU 58	0.052	DMU 66	0.064	DMU 69
81	Castelo Branco	0.685	0.001	DMU 21	0.918	DMU 35	0.006	DMU 58	0.076	DMU 66
82	Faro	0.514	0.728	DMU 29	0.195	DMU 35	0.028	DMU 58	0.049	DMU 66
83	Alcobaca	0.611	0.439	DMU 29	0.508	DMU 35	0.015	DMU 58	0.038	DMU 66
84	Caldas Rainha	0.827	0.000	DMU 21	0.946	DMU 35	0.054	DMU 66		
85	Covilha	0.463	0.629	DMU 29	0.269	DMU 35	0.102	DMU 66		
86	Portimao	0.873	0.792	DMU 29	0.145	DMU 66	0.063	DMU 69		
87	Guarda	0.826	0.893	DMU 29	0.073	DMU 66	0.034	DMU 69		
88	Abrantes	0.544	0.890	DMU 29	0.049	DMU 66	0.060	DMU 69		
89	Vila Real	0.434	0.002	DMU 21	0.973	DMU 35	0.001	DMU 58	0.025	DMU 66
90	Cantanhede	0.581	0.000	DMU 28	0.951	DMU 29	0.010	DMU 58	0.040	DMU 69
91	Beja	0.655	0.907	DMU 29	0.093	DMU 69				
92	Esposende	0.401	0.921	DMU 29	0.024	DMU 35	0.055	DMU 66		
93	Cascais	0.648	0.536	DMU 29	0.070	DMU 58	0.256	DMU 66	0.138	DMU 69
94	Gondomar	0.627	0.006	DMU 21	0.653	DMU 35	0.341	DMU 66		
95	Braga	0.513	0.386	DMU 29	0.233	DMU 35	0.381	DMU 66		
96	Maria Feira	0.495	0.398	DMU 29	0.581	DMU 35	0.011	DMU 58	0.010	DMU 66
97	Sado	0.669	0.710	DMU 29	0.019	DMU 58	0.211	DMU 66	0.059	DMU 69
98	Indaqua Santo	0.663	0.949	DMU 29	0.021	DMU 66	0.030	DMU 69		
99	Valongo	0.618	0.312	DMU 29	0.563	DMU 35	0.124	DMU 66		
100	Paredes	0.715	0.009	DMU 28	0.959	DMU 29	0.032	DMU 69		
101	Figueira	0.560	0.013	DMU 21	0.917	DMU 35	0.071	DMU 66		
102	Mafra	0.457	0.016	DMU 21	0.974	DMU 35	0.005	DMU 58	0.006	DMU 66
103	Fafe	0.622	0.756	DMU 29	0.213	DMU 35	0.001	DMU 58	0.030	DMU 66
104	Ourem	0.390	0.028	DMU 12	0.689	DMU 29	0.283	DMU 41		
105	Batalha	0.612	0.994	DMU 29	0.001	DMU 58	0.001	DMU 66	0.003	DMU 69
106	Alcanena	0.649	0.990	DMU 29	0.001	DMU 58	0.006	DMU 66	0.002	DMU 69
107	Carrazeda	0.598	1.000	DMU 29						

Table 11: First stage best practices

Observation	nr.	benchmarks
Bastogne	29	55
South East	66	53
South Staffs	58	39
I.E.G.M.	35	32
Hydron-MN	21	31
A.W.W.	28	27
SMAS Porto	69	22
South East	12	17
Three Valleys	62	11
Gippsland	3	10
Sidney	13	10
Vitens	17	3
Theux	41	3
United Utilities	54	2
Bullingen	30	1

weight in front of each best practice denotes its relative importance (the λ in the dual DEA model, see Cooper et al. (2004)). As some best practices are more influential than others, we present the frequency that an efficient DMU is benchmark for another company in table 11.

Above, we discussed the problem to find an appropriate weighting scheme. In this and the remaining sections, we opt to weight the efficiency scores by the number of connections. This is a measure for the number of people who are affected by the relative (in)efficiency of a company. The average weighted and non-weighted efficiency scores are presented in table 12. All countries except for Belgium gain from the weighting scheme, which indicates that the larger utilities are relatively more efficient in the Netherlands, England, Wales, Australia and Portugal, while the smaller utilities are relatively more efficient in Belgium.

Table 12: First stage averages

	non-weighted	weighted
the Netherlands	0.844	0.910
England	0.785	0.861
Australia	0.747	0.855
Portugal	0.608	0.668
Belgium	0.700	0.532
Portugal - public	0.618	0.693
Portugal - private	0.589	0.590

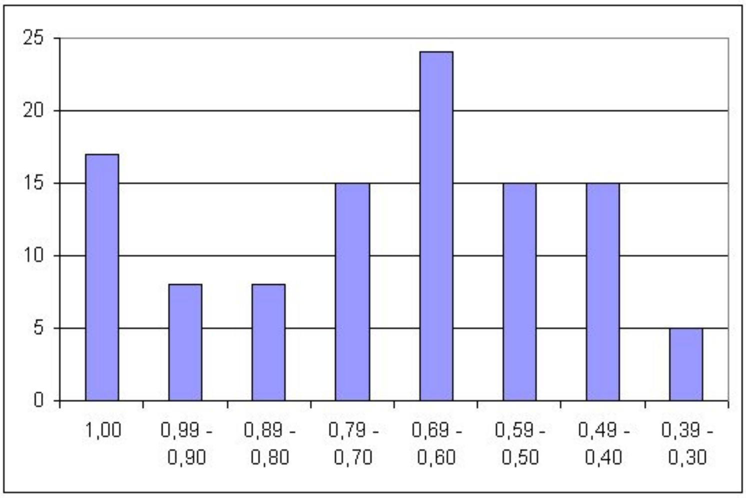

Figure 4: Histogram estimate

If we recognize that our data are subject to random noise, we can correct the DEA efficiencies for bias and estimate confidence intervals for them. The noise provides the missing data, the imperfect quality of the data (even if it is audited by regulators) and the atypical results. By the use of the package FEAR in R, we generate bootstrap samples by a homogeneous bootstrap as described above (Wilson, 2005). The estimation of the bandwidth yields $h = 0.2469$. The reciprocals of the original estimates as well as their bias-corrected counterpart are displayed in table 13. Except for five utilities (Hydron-FL, Almada, Batalha, Alcanena and Carrazeda), according to the rule in equation 9, the bias is sufficiently large relative to the variance. This means that the bias-corrected efficiency estimates should be preferred to the original DEA estimates. We clearly notice the upward-bias in the original estimates. The last two columns of table 13 show the bootstrap estimates of 95-percent confidence intervals. The average width of the confidence intervals amounts 0.337. The widest intervals correspond to South West (0.649) and Aiec (0.636). Caldas Rainha (0.097) and Iegm (0.116) have the smallest 95-% confidence intervals. There is little difference in the confidence interval bandwidth of the five countries.

Due to the upward-bias in the original estimates and due to the bootstrap correction in the confidence intervals, the original estimates lie for every observation outside, but close to, the lower-bound of the confidence interval. However, the bias-corrected estimates lie for every observation inside the confidence intervals. Due to the overlap among the confidence intervals, making relative comparisons among the firms is an intricate issue. Moreover, as the original DEA-estimates are biased, they cannot be interpreted as a ranking device. Notice that we do not observe confidence intervals with a lower bound of 100% or below. As the true efficiency of a DMU cannot exceed 100%, and we measure the 95% confidence intervals, this is a correct observation.

Obviously, the average weighted and unweighted efficiency scores, as presented in table 14, are lower than their upward-biased counterparts. The Dutch water utilities are performing most efficiently, in a weighted as well as in an unweighted scheme. They are closely followed by the privatized English and Welsh firms. The Portuguese and especially the Belgian firms lag behind. The next section develops a second stage analysis which tries to explain the efficiency scores by the use of environmental variables.

Table 13: First stage bootstrapping results

DMU	Eff. score ($\hat{\delta}$)	bias-corrected ($\hat{\hat{\delta}}$)	$\widehat{Bias}$	Variance ($\hat{\sigma}^2$)	Lower bound	Upper bound
Actew	1.6156	1.7588	-0.1432	0.0049	1.6385	1.9089
Barwon	2.0473	2.2052	-0.1579	0.0065	2.0711	2.3847
Gippsland	1.0000	1.2228	-0.2228	0.0096	1.0154	1.3656
Highlands	2.6913	3.0052	-0.3139	0.0200	2.7387	3.2922
Coliban	1.0000	1.2441	-0.2441	0.0151	1.0127	1.4372
Gold Coast	1.0983	1.2002	-0.1019	0.0021	1.1160	1.3028
Gosford	1.4311	1.5197	-0.0886	0.0013	1.4542	1.5946
Goulburn	1.6716	1.9101	-0.2385	0.0140	1.6981	2.1232
Hunter	1.4994	1.6815	-0.1821	0.0070	1.5217	1.8592
Power Water	1.4018	1.6424	-0.2406	0.0166	1.4217	1.8690
SA Water	1.3970	1.5851	-0.1881	0.0096	1.4205	1.7924
South East	1.0000	1.2905	-0.2905	0.0223	1.0197	1.5305
Sidney	1.0000	1.2921	-0.2921	0.0199	1.0210	1.5297
Water Corporation	1.4857	1.7559	-0.2702	0.0196	1.5114	2.0146
Wggr	1.3277	1.5099	-0.1822	0.0082	1.3495	1.6856
Wmd	1.6341	1.8403	-0.2062	0.0125	1.6601	2.0609
Vitens	1.0000	1.3144	-0.3144	0.0281	1.0178	1.5792
Pwn	1.0720	1.2183	-0.1463	0.0042	1.0918	1.3505
Oasen	1.0994	1.1983	-0.0989	0.0028	1.1101	1.3106
Hydron-Fl	1.3095	1.4571	-0.1476	0.0073	1.3241	1.6324
Hydron-MN	1.0000	1.1467	-0.1467	0.0037	1.0195	1.2673
Brabant Water	1.0904	1.2844	-0.1940	0.0085	1.1105	1.4684
Wml	1.4095	1.5906	-0.1811	0.0077	1.4250	1.7683
Aie	1.2850	1.3559	-0.0709	0.0011	1.3036	1.4313
Aquasambre	1.6753	1.7815	-0.1062	0.0026	1.6927	1.8880
Aiec	2.2574	2.6160	-0.3586	0.0271	2.3064	2.9419
Aiem	2.1819	2.4331	-0.2512	0.0158	2.2123	2.6935
Aww	1.0000	1.2786	-0.2786	0.0197	1.0188	1.4854
Bastogne	1.0000	1.2187	-0.2187	0.0096	1.0156	1.3819
Bullingen	1.0000	1.2203	-0.2203	0.0121	1.0173	1.4269
Ciev	1.1367	1.3307	-0.1940	0.0092	1.1594	1.5310
Cile	1.4852	1.6075	-0.1223	0.0030	1.5070	1.7274
Iecbw	1.8969	2.0918	-0.1949	0.0110	1.9280	2.3115
Idemls	1.7190	1.8113	-0.0923	0.0018	1.7381	1.9051
Iegm	1.0000	1.0650	-0.0650	0.0009	1.0154	1.1318
Inasep	2.4141	2.6239	-0.2098	0.0081	2.4563	2.8152
Pidpa	2.2001	2.4715	-0.2714	0.0176	2.2296	2.7410
Wavre	1.0138	1.1401	-0.1263	0.0033	1.0275	1.2515
St-Vith	1.4051	1.6220	-0.2169	0.0156	1.4163	1.8790
Swde	2.7840	3.1548	-0.3708	0.0252	2.8332	3.4554
Theux	1.0000	1.2370	-0.2370	0.0132	1.0154	1.4353
Tmvw	1.6355	1.7888	-0.1533	0.0067	1.6573	1.9651
Vmw	1.9056	2.2037	-0.2981	0.0193	1.9402	2.4810
Anglian	1.4814	1.7354	-0.2540	0.0176	1.5096	2.0279
Bristol	1.2896	1.4574	-0.1678	0.0078	1.3000	1.6360
Bournemouth	1.2925	1.5246	-0.2321	0.0137	1.3116	1.7212
Cambridge	1.6219	1.8050	-0.1831	0.0073	1.6489	1.9871
Dee Valley	1.7223	1.8386	-0.1163	0.0030	1.7499	1.9591
Essex Suffolk	1.0739	1.1913	-0.1174	0.0041	1.0906	1.3249
Folkestone	1.3364	1.4334	-0.0970	0.0021	1.3583	1.5431
Mid Kent	1.6007	1.7335	-0.1328	0.0049	1.6223	1.8915
South East water	1.3456	1.6006	-0.2550	0.0216	1.3676	1.8834
Northumbrian	1.3519	1.5354	-0.1835	0.0102	1.3701	1.7391
United Utilities	1.0000	1.3128	-0.3128	0.0321	1.0140	1.6274
Portsmouth	1.0256	1.1219	-0.0963	0.0027	1.0335	1.2255
Sutton Surrey	1.2112	1.3176	-0.1064	0.0031	1.2251	1.4299
Southern	1.2050	1.3496	-0.1446	0.0068	1.2205	1.5238
South Staffs	1.0000	1.1556	-0.1556	0.0052	1.0157	1.3074
Severn Trent	1.0000	1.3183	-0.3183	0.0312	1.0202	1.6264
South West	2.4374	2.7662	-0.3288	0.0276	2.4701	3.1190
TendringHundred	1.2280	1.2969	-0.0689	0.0009	1.2449	1.3620
Three Valleys	1.0000	1.2359	-0.2359	0.0113	1.0202	1.4249
Wessex	2.2819	2.6055	-0.3236	0.0213	2.3327	2.9041
Yorkshire	1.0188	1.2195	-0.2007	0.0114	1.0383	1.4339
Sintra	1.6310	1.9052	-0.2742	0.0247	1.6500	2.2418
Smas Oeiras	1.0000	1.2107	-0.2107	0.0097	1.0198	1.4105
Smas Loures	1.5115	1.7725	-0.2610	0.0208	1.5346	2.0729
Águas Gaia	1.4362	1.5442	-0.1080	0.0026	1.4549	1.6531
Smas Porto	1.0000	1.2781	-0.2781	0.0194	1.0177	1.4986
Vimágua	1.7352	1.8232	-0.0880	0.0013	1.7585	1.9019
Almada	1.3817	1.5597	-0.1780	0.0111	1.4052	1.8015
Coimbra	1.8857	2.0181	-0.1324	0.0041	1.9123	2.1596
Xira	1.1963	1.3556	-0.1593	0.0054	1.2188	1.5131
Maia	1.3860	1.4850	-0.0990	0.0027	1.3999	1.5959
Leiria	1.9913	2.0834	-0.0921	0.0015	2.0167	2.1708
Viseu	2.4936	2.6401	-0.1465	0.0033	2.5295	2.7575
Castelo	2.0389	2.1318	-0.0929	0.0018	2.0577	2.2244
Aveiro	2.3676	2.5591	-0.1915	0.0096	2.3946	2.7745
Torres Vedras	2.8981	3.0284	-0.1303	0.0040	2.9239	3.1663

Table 13: First stage bootstrapping results - continued

DMU	Eff. score ($\hat{\delta}$)	bias-corrected ($\hat{\hat{\delta}}$)	$\widehat{Bias}$	Variance ($\hat{\sigma}^2$)	Lower bound	Upper bound
Santarem	2.6928	2.8922	-0.1994	0.0074	2.7363	3.0701
Castelo Branco	1.4615	1.5284	-0.0669	0.0009	1.4772	1.5969
Faro	1.9466	2.0622	-0.1156	0.0022	1.9756	2.1607
Alcobaca	1.6378	1.7202	-0.0824	0.0013	1.6590	1.8008
Caldas Rainha	1.2103	1.2660	-0.0557	0.0006	1.2249	1.3222
Covilha	2.1588	2.3080	-0.1492	0.0043	2.1861	2.4443
Portimao	1.1459	1.2575	-0.1116	0.0028	1.1614	1.3766
Guarda	1.2104	1.3082	-0.0978	0.0022	1.2282	1.4116
Abrantes	1.8384	1.9803	-0.1419	0.0054	1.8572	2.1427
Vila Real	2.3064	2.4205	-0.1141	0.0027	2.3354	2.5412
Cantanhede	1.7214	1.8793	-0.1579	0.0061	1.7493	2.0362
Beja	1.5255	1.6746	-0.1491	0.0047	1.5470	1.8214
Esposende	2.4913	2.7337	-0.2424	0.0120	2.5295	2.9675
Cascais	1.5435	1.6755	-0.1320	0.0041	1.5666	1.8130
Gondomar	1.5943	1.7362	-0.1419	0.0061	1.6099	1.9122
Braga	1.9500	2.1649	-0.2149	0.0128	1.9701	2.4046
Maria Feira	2.0214	2.1468	-0.1254	0.0035	2.0524	2.2804
Sado	1.4940	1.6192	-0.1252	0.0040	1.5183	1.7612
Indaqua Santo	1.5076	1.6488	-0.1412	0.0048	1.5328	1.8086
Valongo	1.6175	1.7191	-0.1016	0.0020	1.6393	1.8146
Paredes	1.3993	1.6030	-0.2037	0.0080	1.4244	1.7709
Figueira	1.7877	1.8657	-0.0780	0.0013	1.8071	1.9444
Mafra	2.1878	2.3146	-0.1268	0.0030	2.2201	2.4303
Fafe	1.6079	1.7458	-0.1379	0.0050	1.6349	1.8961
Ourém	2.5652	2.8977	-0.3325	0.0215	2.6019	3.1836
Batalha	1.6330	1.8870	-0.2540	0.0223	1.6540	2.1713
Alcanena	1.5407	1.7559	-0.2152	0.0185	1.5584	2.0208
Carrazeda	1.6724	1.8953	-0.2229	0.0172	1.6863	2.2074

Table 14: First stage bias-corrected averages

	reciprocal ($\hat{\delta}$)		efficiency score ($\hat{\theta}$)	
	non-weighted	weighted	non-weighted	weighted
the Netherlands	1.396	1.331	0.732	0.761
England	1.550	1.412	0.682	0.728
Australia	1.665	1.480	0.639	0.696
Portugal	1.909	1.778	0.553	0.593
Belgium	1.803	2.269	0.614	0.470
Portugal - public	1.908	1.745	0.563	0.609
Portugal - private	1.912	1.884	0.536	0.543

6 Second stage analysis

The efficiency of drinking water utilities is prone to environmental factors which are not under control of the firms' managers. Nevertheless, insight in these factors is very important for evaluating the cost of regulation. If in an input-oriented model an environmental variable z is unfavorable to efficiency, one can consider the variable as an additional and undesired output variable. The 'production' of this undesired output decreases the efficiency as it absorbs inputs. A favorable environmental variable can be considered as a substitutive input which could save the use of other inputs in the production process. A first subsection addresses the more conventional Tobit regression and the use of bootstrapping in a second stage analysis. A second subsection applies these theories to the drinking water utilities. In a third subsection, the efficiency scores are corrected by taking into account the environmental variables.

6.1 Theoretical framework

To explain the efficiency of DMUs, researchers have frequently employed a regression model on the DEA-efficiency scores:

$$\hat{\delta}_i = z_i\beta + \epsilon_i \tag{11}$$

where z_i is a (row) vector of firm-specific variables which is expected to influence the efficiency of DMU_i. β denotes a vector of parameters to be estimated together with some statistical noise ϵ_i. The ordinary least squares (OLS) method will lead to a biased estimate as it assumes a normal and homoscedastic distribution of the error term and the dependent variable. However, the efficiency estimates $\hat{\delta}_i (= \frac{1}{\hat{\theta}_i})$ have by construction a lower limit of 1 which creates a concentration of observations at this single value. This leads to a censored sample. In the literature, Tobit models are usually considered to provide a solution whenever there is a mass of observations at a limiting value.
Simar and Wilson (2007) consider the justification for the use of a censored Tobit regression as '*nonsense*'. As $\hat{\delta}_i \geq 1$, they argue that this involves a truncated rather than a censored error term. Both censoring and truncation involve a loss of information about the dependent variable, but where censoring assumes the observation of all right-hand side variables, truncation supposes an information loss on both sides (left and right-hand side) of the regression (see appendix Simar and Wilson (2007) for an extensive discussion). Therefore, β and σ should be estimated by the use of maximum likelihood. Nevertheless, the standard inference is intricate due to three problems. First, in small samples $\hat{\delta}_i$ is highly influenced by the position of the estimated frontier. As in linear regression models, this causes correlation among the estimates ($\hat{\delta}_i$). Likewise in small samples, as the input and output variables which determine the DEA-efficiency are correlated with the environmental variables, the error term ϵ_i will be correlated with z_i. These first two issues disappear asymptotically. A third and more serious problem is, as mentioned above, the bias of the DEA-efficiency score $\hat{\delta}_i$ towards 1. Simar and Wilson (2007) recommend a double-bootstrap procedure to produce, with bias-corrected estimates of $\hat{\hat{\delta}}_i$, valid confidence interval estimates for the parameters in the second-stage regression. The following 7 steps, as developed by Simar and Wilson (2007), are needed:

1. Compute for every DMU the estimated DEA-efficiency score $\hat{\delta}_i$.
2. Calculate $\hat{\beta}$ and $\hat{\sigma}_\epsilon$ by the use of maximum likelihood from the left normal truncated regression in equation (11) using the observations for which $\hat{\delta}_i > 1$.
3. Obtain L_1 bootstrap estimates for each $\hat{\delta}_i$ by looping L_1 times the following four steps:
   (a) For $i = 1, \ldots, n$ draw ϵ_i^* from $N(0, \hat{\sigma}_\epsilon^2)$ with left-truncation at $(1 - \hat{\beta}' z_i)$.
   (b) Compute $\hat{\delta}_i^* = z_i\hat{\beta} + \epsilon_i$.

(c) Modify the output variables such that $y_i^* = \frac{\hat{\delta}_i}{\hat{\delta}_i^*} y_i$.

(d) Recompute the DEA-algorithm, where (x_i, y_i) is replaced by (x_i, y_i^*) to obtain $\hat{\delta}_i^*$.

4. Compute by the use of the bootstrap estimates of the previous step the bias-corrected estimator $\hat{\hat{\delta}}_i = \hat{\delta}_i - \widehat{BIAS}(\hat{\delta}_i)$ where the estimated bias is computed as in equation (7).

5. Again, estimate the truncated regression of $\hat{\hat{\delta}}_i$ on z_i by maximum likelihood. Denote the estimates by $(\hat{\hat{\beta}}, \hat{\hat{\sigma}})$.

6. Obtain L_2 bootstrap estimates for $\hat{\beta}$ and $\hat{\sigma}_\epsilon$ by looping L_2 times the following three steps:

   (a) For $i = 1, \ldots, n$ draw ϵ_i^{**} from $N(0, \hat{\hat{\sigma}}_\epsilon^2)$ with left-truncation at $(1 - \hat{\hat{\beta}}' z_i)$.

   (b) Compute $\delta_i^{**} = z_i \hat{\hat{\beta}} + \epsilon_i^{**}$.

   (c) Compute $(\hat{\hat{\beta}}^*, \hat{\hat{\sigma}}^*)$ by a maximum likelihood estimation of the truncated regression of $\hat{\delta}_i^{**}$ on z_i.

7. Construct the estimated confidence intervals for β and σ_ϵ by using the bootstrap values of step 6. Test hypotheses such as the probability that an estimate $\hat{\beta}_i < 0$ by considering the relative frequency of nonnegative $\hat{\hat{\beta}}_i^*$ bootstrap estimates.

The first bootstrap in the algorithm is used to compute the bias-corrected efficiency scores, while the second bootstrap estimates the effect of the environmental variables on the bias-corrected efficiencies. However, as in the conventional Tobit regressions, the environmental variables z_i in the double-bootstrap procedure do not influence the boundary of Ψ. This is due to a separability condition: by assumption the variables of Z lie in a space apart from the production space for inputs and outputs Ψ. A second drawback of the described double-bootstrap procedure is the reliance on some parametric assumptions such as a linear model and a truncated normal error term. In the following subsection, we discuss and implement procedures which avoid these assumptions.

6.2 Second stage results

Many elements in the production process of drinking water utilities are outside the control of the firms' managers. Especially the social, physical and institutional environment are not taken into account in the first stage analysis. The institutional environment is determined by the government (or its regulatory agencies). The government is able to influence the drinking water companies directly by, for example, enforcing strict quality requirements, or more indirectly by introducing incentive schemes, such as benchmarking or economic regulation (e.g. price cap formula). The physical environment captures differences in, for example, the geographic relief, soil structure, climate (and especially rainfall and temperature) or the relative abundance of pure drinking water. The social environment includes the relative wealth of the customers or attitude towards (excessive) water consumption. Although many elements in the social, physical as well as the institutional environment highly influence the cost level of the drinking water utilities, due to the lack of (uniform) data, we have to make considerable simplifications.

A first physical variable included in the second stage model is the percentage of leakage. This variable captures the geographical relief (as a more hilly landscape requires more pressure on the network of pipes which could cause more easily leakage) and the extent of maintenance (more leakages correspond to less expenses with maintenance). If the influence of the geographical circumstances outweighs the neglect of maintenance, we expect a negative influence on efficiency. In the opposite case, we anticipate a positive effect in efficiency. A second physical factor is the percentage of groundwater extraction. The utilities that abstract more groundwater are supposed to be more efficient, since the production cost is much lower than the counterparts that abstract superficial water or import water from other utilities. The proportion of water delivered to industrial customers relative to domestic users is the third, and last, physical variable. It is expected that efficiency will change positively with a higher percentage of industrial customers. The first social environmental variable, gross regional product (GRP), captures the relative wealth of the customers, the difference in skill distribution (see above) and approximates the average productivity of a region. GRP is measured in per capita purchasing power parity. Water consumption per capita, the second and last social environmental factor, measures demand side management. We incorporate five institutional dummy variables in the second stage analysis. The first captures the scope of activities: we assign a dummy variable if the utility's only activity is providing drinking water. Evidence from the literature suggests that drinking water services have economies of scope and therefore they are more efficient when they are responsible also for other activities as a result of the savings obtained with the existing synergies. Corporatization, as a second institutional factor, is supposed to have a positive effect in efficiency thanks to harder budget constraints. Corporatization is the application by public entities of rules and mechanisms of the private sector, which enable the public entities to practise a private management. The third institutional variable denotes the water delivery in one (or maximum three) municipalities. This indication of scale economies is expected to have a negative effect in efficiency. Finally, we include dummy variables for utilities which have a regulator or use a kind of benchmarking. We did not assign Portugal with a dummy for benchmarking as it introduced its benchmarking only in 2005. We expect that these two variables have a positive effect in efficiency.

In this subsection, we first evaluate the importance of the environmental influences. As there is still a lot of discussion in the literature about the appropriate way of measuring second stage effects, we present results both from Tobit regressions of the original values obtained as well as from the bootstrapping algorithm. Although Tobit estimates could be possibly biased, it is not clear that bootstrap estimates are necessarily more reliable (Simar and Wilson (2007) provide only Monte-Carlo evidence). Comparable results of both estimation techniques will add robustness and confidence to the estimates. We estimate the Tobit regression of the original as well as of the bias-corrected efficiency estimates in both a censored and a truncated sample. The bias-corrected efficiency estimates are those obtained in the bootstrap analysis of section (5.2). Note that, in order to avoid two boundaries, the depend variable ($\hat{\delta}$) is larger or equal to one, such that a positive sign denotes a negative influence on the efficiency (i.e. a favorable environmental factor), while a negative sign denotes a positive influence (i.e. an unfavorable environmental factor). The results are presented in table 15.

Although in 3 out of 11 estimates the boostrapping algorithm discovers the opposite sign of the Tobit regressions, the Tobit estimates are not covered by the 95%-confidence intervals of the bootstrapped variables in only 2 cases (see table 16). To get an idea on how strongly the different estimates are related, we measure the correlation coefficient. The Pearson correlation coefficient estimates the strength and direction of a *linear* relationship between the variables, while the non-parametric Spearman's rank correlation coefficient uses ranks of the data instead of the actual observed data. The results of the correlation coefficients are reported in table 17. According to the Pearson measure, the different estimation techniques are closely related. This contrasts with the Spearman's rank correlation which reveals a close correlation between the Tobit regressions, however, no significant correlation between the Tobit and bootstrap estimates. The question, which we do not solve here, is to know whether the bootstrap estimates are superior.

The second stage results in table 15 indicate that firms which spend less resources on maintenance, and hence have a higher percentage of leakage, wrongly appear as more efficient. The positive and significant Tobit results on groundwater use subvert the postulate that the use of (cheaper) groundwater increases the efficiency. It is highlighted that the groundwater abstraction in some countries is often associated with the size of utilities. For example, in Portugal only the small companies have the abstracted water as source. Yet, more likely than providing an indication for economies of scale (as we capture this effect later on), the estimation on groundwater use could indicate that only the most efficient companies are capable to purify the most costly surface water. The estimations are inconclusive on whether industrial customers encourage the utilities to produce most efficiently. The truncated DEA-VRS and bootstrap second stage method depict a negative effect on this variable. The social explanatory factor GRP reveals the expected positive influence on efficiency. The negative influence in efficiency of consumption per capita indicates that the policies of demand side management are filling up the wished. Hence, the companies increase the efficiency by cost reductions rather than by increasing the water sale. Concerning the first of the institutional variables, utilities with activities only in drinking water provision show a positive signal. This evidence counters the literature in that water services seem not to have economies of scope. Although the Tobit regressions yield the expected positive effect of corporatization in efficiency, these estimates are not significantly different from 0. The significant negative effect of corporatization by the bootstrap estimates could be linked with the fact that corporatization makes the companies comprise all costs, leading them to seem wrongly inefficient. The positive effect in efficiency of the variable delivery in one municipality suggests that the water utilities in the sample studied have no scale economies. The values obtained are always significant except for the truncated sample with DEA-VRS efficiencies. The results of the regulator (existent or non-existent) are not much conclusive, although they are not significant in all cases, except for the bootstrap. The latter reveals, in correspondence with the literature, a positive effect of the regulation in efficiency. Finally, the effect of benchmarking in efficiency is positive and always with significance. This tool to improve performance turns out to be very appropriate.

6.3 Taking into account environmental variables

The above mentioned separability assumption assumes that the environmental variables do not directly influence the efficiency scores. Hence, one can only *ex post* measure the influence of environmental variables to efficiency. In this paragraph, we present and apply some approaches to incorporate the environmental variables directly in the efficiency measurement.

A first model is based on the *one-stage* or the *all-in-one approach*. In this approach, we include the environmental variables directly in the DEA-problem along with the traditional input and output variables. Therefore, we have to decide *a priori* whether a variable will be classified as an input (i.e. if the variable is favorable to efficiency), or as an output (i.e. if the variable is damaging to efficiency). For a constant efficiency, an additional input enables a DMU to produce more outputs, while an additional output variable requires the use of more inputs. A major drawback of this estimation technique is its free disposability assumption (as does the traditional DEA model). Hence, in an input-oriented model, one assumes that all the inputs are reducible, or similarly, in an output-oriented model all outputs are expansible (see e.g. Fried et al., 1999). Of course, by definition environmental variables are not freely reducible nor expansible.

In the one-stage analysis, we include the significant continuous bootstrap variables in addition to the traditional variables. We know the influence on efficiency of the bootstrap estimates from previous estimations. As input variables for the DEA model we use besides the number of employees (as proxy for labor expenses) and the length of mains (as a proxy for capital expenses), the percentage of leakage and the gross regional product. The applied output variables are beside the number of customers and the volume of delivered drinking water, the proportion of delivery to industrial customers and the consumption per capita. The input-oriented DEA-VRS efficiency scores as well as the input-oriented DEA-VRS bias-corrected efficiency estimates are reported in the second and third column of table 18. As DEA allows the possibility of specialization, an additional input or output variable cannot decrease the efficiency estimate of a DMU. However, thanks to the specialization, some utilities increase their efficiency considerably (e.g. Swde, Torres Vedras or Ourém). This suggest that some utilities could wrongly appear as efficient by one specific low input variable (e.g. a low regional product) or one specific high output factor (e.g. a high proportion of industrial customers). Furthermore, as the efficiency of the DMUs only increases (or in worst case scenario remains unchanged), the one-stage approach seems impractical in correcting efficiency for environmental variables.

A second approach to reshape the boundary of the attainable set uses *the residuals* of the Tobit regression. As a residual factor, the residuals capture the share of technical efficiency which remains after controlling for exogenous influences. A positive residual, ϵ_i, indicates managerial inefficiency, while a negative residual reflects managerial efficiency. As the corrected efficiency score ($\hat{\delta}_i''$) should possess a one-sided distribution, following De Witte en Moesen (2006), we sort the residuals of the Tobit regression in order of magnitude (from small to large) and compute:

$$\hat{\delta}_i'' = \epsilon_i + (1 - \frac{1}{w} \sum_{j=1}^{w} \epsilon_j) \tag{12}$$

Utilities which show after correcting for environmental factors a managerial efficiency (i.e. $\epsilon_i < 0$) will decrease their efficiency score $\hat{\delta}_i''$ (and hence increase their efficiency) relative to the 'average inefficiency' (i.e. the term in brackets). The relative strictness of the correction depends on the number of utilities w included in the 'average inefficiency' term. As this term is computed with ordered residuals, a larger w denotes relatively more tolerance towards inefficiencies. One can perform a sensitivity analysis to measure the influence of w (De Witte and Moesen, 2006).
In applying this second approach, we use the residuals of the bias-corrected censored Tobit-model as these are best correlated with the bootstrap estimates (see table 17). In order to interpret in equation (12) the term in brackets as the average inefficiency, we set w equal to 53 (half of our dataset). The corrected efficiency scores which turn out to be lower than 1 are rounded off. The results are presented in table 18.
In a first analysis, we equalize the environmental influences for all companies by taking the residuals of the bias-corrected DEA efficiencies in a censored sample regressed on the above described social, physical and institutional environmental variables, and applying equation (12). The efficiency scores which are presented in the fourth column of table 18 can be interpreted as if the utilities were facing exactly the same social, physical and institutional constraints and benefits. There are 15 utilities in the sample which could previously take advantage of their favorable environmental factors, and therefore have a reduced efficiency score when taking into account the environment (remark that this number as well as all the values in this analysis depend on the value of w). With the exogenous influences equalized, the variation left between the DMUs can mainly be attributed to managerial influences (as disregarded environmental factors are primarily captured by the intercept of the regression term). The country averages, at the bottom of table 18, show that the Portuguese and Belgian drinking water utilities gain most from the correction for environmental variables. In particular, in the following analysis we examine whether this could be attributed to the equalization of incentives.
In order to capture the effects of the regulatory and benchmark incentive schemes on the average efficiency of analyzed countries, we re-estimate the previous censored Tobit regression without the dummy variables 'regulation' and 'benchmarking'. The obtained scores in column (5) of table 18 reflect efficiencies as would the utilities work in exactly the same environment but with different incentive schemes. The country averages at the bottom of the table reveal the effectiveness of the Dutch benchmarking scheme (as the Dutch companies are performing more efficient if benchmarking is taken into account) and the power of the English and Welsh, Australian and (private) Portuguese regulatory models. As there is no clear incentive structure for the Belgian and Portuguese public utilities, their average efficiency falls in comparison to the equalized situation. The Belgian and Portuguese authorities could ameliorate the performances of their drinking water sector by introducing a clear incentive scheme.

Table 15: Second stage results

Dependent variable Sample assumption	DEA-VRS eff. censored	DEA-VRS eff. truncated	bias-corr VRS censored	bias-corr VRS truncated	DEA-VRS bootstrap
Intercept	3.1285 ***	2.0693 ***	2.7042 ***	3.1409 ***	4.2216 ***
	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)
Leakage (%)	-0.01580 **	-0.000895	-0.007403	-0.01101	-0.02258 ***
	(0.020)	(0.934)	(0.209)	(0.245)	(0.000)
Groundwater extraction (%)	0.002825 **	0.002178	0.002231 *	0.003477 *	-0.0001359
	(0.030)	(0.277)	(0.059)	(0.068)	(0.150)
Industry water / household delivery	-0.2313 *	0.3487	-0.1026	-0.1772	0.02396 ***
	(0.079)	(0.313)	(0.269)	(0.328)	(0.000)
Gross regional product (PPP/capita)	-4.16 E-5 ***	-1.52 E-5	-2.15 E-5 *	-3.74 E-5 *	-6.879 E-5 ***
	(0.004)	(0.546)	(0.092)	(0.085)	(0.000)
Consumption per capita	4.56 E-5 **	5.55 E-5 **	5.08 E-5 **	6.08 E-5 **	5.716 E-5 ***
	(0.042)	(0.039)	(0.015)	(0.027)	(0.000)
Water unique activity (=1)	-0.2461 **	-0.1545	-0.2087 *	-0.3362 *	-0.2644 ***
	(0.049)	(0.448)	(0.065)	(0.073)	(0.000)
Corporatization (=1)	-0.09583	-0.6898	-0.07701	-0.2759	1.2254 ***
	(0.703)	(0.188)	(0.735)	(0.515)	(0.000)
Delivery in one municipality (=1)	-0.2973 *	-0.3443	-0.2943 **	-0.4677 **	-1.3448 ***
	(0.062)	(0.183)	(0.041)	(0.049)	(0.000)
Regulator (=1)	0.2620	0.6866	0.2056	0.4674	-0.9637 ***
	(0.212)	(0.162)	(0.274)	(0.224)	(0.000)
Benchmarking (=1)	-0.7091 ***	-0.7529 **	-0.6198 ***	-0.9314 ***	-0.1198 ***
	(0.002)	(0.035)	(0.002)	(0.005)	(0.000)
SE of regression	0.4424	0.4319	0.4643	0.4692	1.1498

Note: n=107; p-values in brackets; *** denotes significance at 1% level, ** at 5% and * at 10%

Table 16: Bootstrapping estimates - confidence intervals

	bootstrap estimate	95% conf. int. lower bound	95% conf. int. upper bound	Tobit estimates
intercept	4.2216 ***	1.9184	6.6944	in conf. inter.
leakage (%)	-0.02258 ***	-0.06406	0.01550	in conf. inter.
groundwater extraction (%)	-0.0001359	-0.007760	0.007383	in conf. inter.
industry water / household delivery	0.02396 ***	0.01353	0.03404	no in conf. inter.
gross regional product	-6.879 E-5 ***	-1.579 E-4	7.980 E-6	in conf. inter.
consumption per capita	5.716 E-5 ***	-6.598 E-5	1.714 E-4	in conf. inter.
water unique activity (=1)	-0.2644 ***	-1.0319	0.4602	in conf. inter.
corporatization (=1)	1.2254 ***	-0.3341	2.6574	in conf. inter.
delivery in one municipality (=1)	-1.3448 ***	-2.4574	-0.3601	some in conf. inter.
regulator (=1)	-0.9637 ***	-2.1884	0.1741	not in conf. inter.
benchmarking (=1)	-0.1198 ***	-1.3720	1.1548	in conf. inter.

Together with the second stage results of section (6.2), this analysis provides significant evidence for the positive effects of incentives schemes on efficiency. The analysis even demonstrates that in absence of clear and structural incentives the average efficiency of the utilities even falls in comparison with utilities which are encouraged by incentives. The natural monopoly in the drinking water sector leads to the *quiet life* of Hicks (1935) and *X-inefficiency* of Leibenstein (1966). The presence of benchmarking (in the sense of sunshine regulation or yardstick competition) is a key element which replaces competition *in* the market or competition *for* the market by competition by comparison.

7 Conclusion

This paper has explored the effect of incentive schemes in the drinking water sector. Different ideological views on the extent of state intervention in the economy create various incentive structures. We have compared the implemented incentive schemes in the Netherlands, England and Wales, Australia, Portugal and Belgium. Our results show large differences in first stage inefficiencies. On average, the benchmarked Dutch drinking water companies are performing

Table 17: Correlation coefficients among second stage estimates

Pearson \ Spearman	DEA eff. cens.	DEA eff. trunc.	bias-corr DEA cens.	bias-corr DEA trunc.	DEA bootstrap
DEA eff. Cens.	1.000	0.818 (**)	1.000 (**)	0.991 (**)	0.391
DEA eff. trunc.	0.912 (**)	1.000	0.818 (**)	0.882 (**)	0.327
bias-corr cens.	0.999 (**)	0.918 (**)	1.000	0.991 (**)	0.391
bias-corr trunc.	0.994 (**)	0.946 (**)	0.994 (**)	1.000	0.382
DEA bootstrap	0.871 (**)	0.668 (**)	0.878 (**)	0.835 (**)	1.000

Note: n=11; ** denotes significance at 1% level (two-tailed) and * at 5% level (two-tailed)

better (average efficiency score of 0.84) than the privatized English and Welsh utilities (0.79). However, the strict regulatory model of Australia (0.75), the municipal provision in Belgium (0.70) and especially the Portuguese municipal provision with private sector participation (0.61) are lagging behind.

We have interpreted the average 'national' efficiency score of a country as a measure for the homogeneity in efficiency of a country's drinking water sector. Since the number of utilities in the different national samples differ, by resampling we have equalized the sizes of the datasets. It turns out that the efficiency of the Belgian and Dutch drinking water sectors are the most homogeneous. In those two countries, policy makers should relatively easily find agreement among the utilities to adopt new laws.

The second stage procedures examine to which extent the inefficiencies could be attributed to (un)favorable social, physical and institutional environmental factors. Therefore, we have employed censored and truncated Tobit models and a double-bootstrap procedure. The results detect the negative effect on efficiency of the proportion of industrial customers and groundwater extraction, the consumption per capita and the effect of a corporate structure. The portion of leakage, the gross regional product, only supplying drinking water, the delivery in only one municipality and the regulatory and benchmark incentive schemes yield a positive effect on efficiency.

Finally, we have incorporated the social, physical and institutional environmental factors in the efficiency scores. The obtained scores reflect efficiencies as would the utilities work in exactly the same environment. With the exogenous influences equalized, the variation left between the DMUs can mainly be attributed to managerial influences. Here again, the Dutch, English and Welsh utilities perform more efficiently. In order to capture the effects of the regulatory and benchmark incentive schemes on the average efficiency of the analyzed countries, we have estimated efficiencies as the utilities would work in exactly the same environment but with different incentive schemes. The results provide significant evidence for the positive effects of incentives schemes to efficiency. The analysis demonstrates that in absence of clear and structural incentives the average efficiency of the utilities even falls in comparison with utilities which are encouraged by incentives.

Table 18: Correction for environmental variables

Method	(1) First stage	(2) One stage	(3) One stage	(4) Residuals	(5) Residuals
Utility	DEA-VRS eff. (2 in/outp.) ($\hat{\delta}$)	DEA eff. ($\hat{\delta}'$)	bias-corr. ($\hat{\hat{\delta}}'$)	all influences equalized ($\hat{\delta}''$)	Incentive scheme ($\hat{\delta}''$)
Actew	1.616	1.299	1.344	1.664	1.639
Barwon	2.047	1.346	1.394	1.958	1.949
Gippsland	1.000	1.000	1.083	1.207	1.202
Highlands	2.691	1.366	1.410	2.586	2.552
Coliban	1.000	1.000	1.079	1.000	1.000
Gold Coast	1.098	1.021	1.050	1.000	1.000
Gosford	1.431	1.000	1.081	1.289	1.255
Goulburn	1.672	1.068	1.112	1.736	1.722
Hunter	1.499	1.328	1.370	1.406	1.385
Power Water	1.402	1.225	1.273	1.306	1.271
SA Water	1.397	1.259	1.295	1.254	1.238
South East	1.000	1.000	1.082	1.166	1.121
Sidney	1.000	1.000	1.081	1.220	1.219
Water Corporation	1.486	1.187	1.230	1.491	1.481
Wggr	1.328	1.000	1.087	1.423	1.404
Wmd	1.634	1.098	1.137	1.471	1.426
Vitens	1.000	1.000	1.085	1.246	1.228
Pwn	1.072	1.000	1.080	1.209	1.191
Oasen	1.099	1.000	1.058	1.058	1.039
Hydron-Fl	1.310	1.000	1.080	1.000	1.000
Hydron-MN	1.000	1.000	1.082	1.000	1.000
Brabant Water	1.090	1.000	1.082	1.055	1.028
Wml	1.410	1.023	1.057	1.266	1.227
Aie	1.285	1.201	1.237	1.000	1.061
Aquasambre	1.675	1.390	1.428	1.293	1.373
Aiec	2.257	1.000	1.069	2.123	2.221
Aiem	2.182	1.192	1.227	1.874	1.955
Aww	1.000	1.000	1.081	1.270	1.353
Bastogne	1.000	1.000	1.078	1.000	1.000
Bullingen	1.000	1.000	1.077	1.000	1.000
Ciev	1.137	1.000	1.057	1.000	1.000
Cile	1.485	1.280	1.316	1.130	1.226
Iecbw	1.897	1.080	1.111	1.615	1.712
Idemls	1.719	1.283	1.316	1.342	1.422
Iegm	1.000	1.000	1.079	1.000	1.000
Inasep	2.414	1.313	1.347	2.104	2.185
Pidpa	2.200	1.626	1.676	2.228	2.348
Wavre	1.014	1.003	1.038	1.000	1.000
St-Vith	1.405	1.000	1.066	1.181	1.244
Swde	2.784	1.000	1.051	2.575	2.639
Theux	1.000	1.000	1.083	1.000	1.000
Tmvw	1.636	1.370	1.415	1.439	1.520
Vmw	1.906	1.243	1.280	1.722	1.816
Anglian	1.481	1.102	1.138	1.403	1.366
Bristol	1.290	1.000	1.081	1.212	1.137
Bournemouth	1.293	1.000	1.083	1.174	1.132
Cambridge	1.622	1.264	1.306	1.456	1.387
Dee Valley	1.722	1.247	1.284	1.635	1.571
Essex Suffolk	1.074	1.016	1.045	1.000	1.000
Folkestone	1.336	1.095	1.128	1.121	1.055
Mid Kent	1.601	1.318	1.356	1.410	1.345
South East water	1.346	1.203	1.248	1.369	1.311
Northumbrian	1.352	1.000	1.039	1.108	1.063
United Utilities	1.000	1.000	1.083	1.000	1.000
Portsmouth	1.026	1.000	1.033	1.000	1.000
Sutton Surrey	1.211	1.107	1.136	1.063	1.010
Southern	1.205	1.164	1.199	1.000	1.000
South Staffs	1.000	1.000	1.052	1.071	1.016
Severn Trent	1.000	1.000	1.081	1.127	1.087
South West	2.437	1.000	1.082	1.417	1.428
TendringHundred	1.228	1.013	1.043	1.000	1.000
Three Valleys	1.000	1.000	1.079	1.000	1.000
Wessex	2.282	1.263	1.307	2.270	2.233
Yorkshire	1.019	1.000	1.050	1.000	1.000
Sintra	1.631	1.157	1.187	1.496	1.557
Smas Oeiras	1.000	1.000	1.080	1.000	1.000
Smas Loures	1.512	1.199	1.237	1.362	1.411
Águas Gaia	1.436	1.001	1.023	1.000	1.042
Smas Porto	1.000	1.000	1.081	1.000	1.000
Vimágua	1.735	1.014	1.032	1.244	1.312
Almada	1.382	1.199	1.238	1.009	1.084

Table 18: Correction for environmental variables - continued

Method	(1) First stage	(2) One stage	(3) One stage	(4) Residuals	(5) Residuals
Utility	DEA-VRS eff. (2 in/outp.) ($\hat{\delta}$)	DEA eff. ($\hat{\delta}'$)	bias-corr. ($\hat{\hat{\delta}}'$)	all influences equalized ($\hat{\delta}''$)	Incentive scheme ($\hat{\delta}''$)
Coimbra	1.886	1.000	1.016	1.442	1.486
Xira	1.196	1.090	1.134	1.000	1.048
Maia	1.386	1.000	1.048	1.000	1.000
Leiria	1.991	1.000	1.008	1.404	1.454
Viseu	2.494	1.000	1.041	2.087	2.146
Castelo	2.039	1.014	1.033	1.491	1.570
Aveiro	2.368	1.000	1.008	1.937	1.979
Torres Vedras	2.898	1.184	1.217	2.636	2.731
Santarém	2.693	1.460	1.476	2.462	2.492
Castelo Branco	1.462	1.000	1.022	1.000	1.000
Faro	1.947	1.158	1.190	1.567	1.627
Alcobaça	1.638	1.000	1.011	1.008	1.058
Caldas Rainha	1.210	1.000	1.058	1.000	1.000
Covilhã	2.159	1.000	1.005	1.638	1.675
Portimão	1.146	1.000	1.081	1.000	1.000
Guarda	1.210	1.000	1.052	1.000	1.000
Abrantes	1.838	1.000	1.050	1.391	1.450
Vila Real	2.306	1.016	1.022	1.816	1.862
Cantanhede	1.721	1.000	1.080	1.193	1.252
Beja	1.526	1.024	1.049	1.078	1.113
Esposende	2.491	1.016	1.023	2.145	2.195
Cascais	1.544	1.307	1.350	1.445	1.353
Gondomar	1.594	1.012	1.028	1.301	1.172
Braga	1.950	1.011	1.025	1.756	1.646
Maria Feira	2.021	1.016	1.027	1.694	1.561
Sado	1.494	1.327	1.367	1.219	1.125
Indaqua Santo	1.508	1.010	1.041	1.354	1.220
Valongo	1.618	1.016	1.026	1.292	1.177
Paredes	1.399	1.000	1.083	1.117	1.000
Figueira	1.788	1.000	1.011	1.413	1.296
Mafra	2.188	1.000	1.035	2.198	2.111
Fafe	1.608	1.016	1.037	1.393	1.216
Ourém	2.565	1.000	1.044	2.546	2.404
Batalha	1.633	1.000	1.072	1.451	1.294
Alcanena	1.541	1.025	1.063	1.512	1.399
Carrazeda	1.672	1.000	1.072	1.505	1.400
The Netherlands	1.216	1.014	1.083	1.192	1.171
England and Wales	1.358	1.085	1.136	1.230	1.197
Australia	1.453	1.150	1.206	1.449	1.431
Portugal public	1.761	1.055	1.089	1.407	1.448
Portugal private	1.741	1.049	1.085	1.546	1.425
Belgium	1.600	1.149	1.202	1.445	1.504

Table 19: Data appendix

Abbreviation	Company	Country	employees (fte)	mains (km)	production (m^3)	connections
Actew	ACTEW Corporation	AU	315	3013	52275000	136000
Barwon	Barwon Water	AU	254	3361	39433000	123000
Brisbane	Brisbane City Council	AU	644	6273	255009000	420000
Gippsland	Central Gippsland Water	AU	145	1933	62702000	56000
Highlands	Central Highlands Water	AU	114	2117	16080000	55000
CityWest	CityWest Water Limited	AU	161	4004	108818000	307000
Coliban	Coliban Water	AU	42	1993	24886000	62000
Gold Coast	Gold Coast Water	AU	259	2995	70247000	205000
Gosford	Gosford City Council	AU	116	940	15154000	65000
Goulburn	Goulburn Valley Water	AU	118	1712	27704000	51000
Hunter	Hunter Water Corporation	AU	288	4480	71586000	211000
Power Water	Power and Water Authority	AU	233	1220	35142000	43000
SA Water	S. A. Water Corporation	AU	888	8773	165550000	492000
South East	South East Water Limited	AU	278	8336	153869000	594000
Sidney	Sidney Water Corporation	AU	2275	20669	526367000	1684617
Water Corp.	Water Corporation	AU	1481	12045	225481000	649000
Yarra	Yarra Valley Water Limited	AU	279	8787	168192000	632000
Wggr	Waterbedrijf Groningen	NE	188	4766	46000000	272000
Wmd	Waterleidingmaatschappij Drenthe	NE	161	4216	30000000	191000
Vitens	Vitens	NE	921	38012	249000000	1644000
Pwn	Waterleidingbedrijf Noord-Holland	NE	493	10038	87000000	717000
Wlb	Waterleidingbedrijf Amsterdam	NE	520	2708	88000000	477000
Dzh	Duinwaterbedrijf Zuid-Holland	NE	456	4430	75000000	582000
Evides	Evides	NE	402	12314	175000000	907000
Oasen	Oasen	NE	221	4014	45000000	320000
Hydron-Fl	Hydron-Flevoland	NE	87	2249	20000000	126000
Hydron-Mn	Hydron-Midden Nederland	NE	329	6389	76000000	557000
Brabant Water	Brabant Water	NE	668	16864	163000000	948000
Twl	Tilburgse Waterleiding-Maatschappij	NE	61	811	13000000	93000
Wml	Waterleiding Maatschappij Limburg	NE	426	8466	71000000	517000
Aie	Association Intercommunale pour l'Energie et l'Eau	BE (W)	32	313	3558513	20406
Aquasambre	Aquasambre	BE (W)	215	1490	15493907	99607
Aiec	Association Interc. des Eaux du Condroz	BE (W)	19	700	1636496	11439
Aiem	Association Interc. des Eaux de la Molignee	BE (W)	22	372	1420601	10324
Aww	Antwerpse Water Werken	BE (F)	535	2392	107154283	146485
Bastogne	Bastogne	BE (W)	7	116	1371151	5432
Bullingen	Bullingen	BE (W)	5	135	433960	2177
Ciev	Intercommunale des Eaux de la Vallée de la Thyle	BE (W)	7	143	104613	4222
Chimay	Chimay	BE (W)	5	72	1214499	2563
Cile	Compagnie Intercommunale Liégeoise des Eaux	BE (W)	337	3287	33382818	225680
Iecbw	Interc. Gestion et Reali. d'Etudes Techn. et Econ.	BE (W)	59	1463	9205055	54570
Idemls	Interc. Distr. d'Eau de Mons, La Louviere et Soignies	BE (W)	152	1490	11443180	85292
Iegm	L'Interc. d'Etude et de Gestion de Mouscron	BE (W)	21	260	3272505	20427
Inasep	Intercommunale Namuroise de Services Publics	BE (W)	46	900	3429016	27009
Iwva	Interc. Waterleidingmaat. van Veurne-Ambacht	BE (F)	46	101	5458590	28015
Calamine	La Calamine	BE (W)	6	66	743688	4075
Hulpe	La Hulpe	BE (W)	5	55	419834	2590
Pidpa	Prov. En Interc. Drinkwatermaat. Provincie Antwerpen	BE (F)	565	12132	67816151	455815
Wavre	Régie de l'Eau de Wavre	BE (W)	11	200	2455630	12291
St-Vith	Stadtwerke St-Vith	BE (W)	10	163	597357	3028
Swde	Société wallonne des eaux	BE (W)	1470	23432	118428033	713248
Theux	Theux	BE (W)	5	156	697246	4056
Tmvw	Tussengem. Maat. Vlaanderen vr Waterbedeling	BE (F)	477	7658	60752681	422343
Vmw	Vlaamse Maatschappij voor Watervoorziening	BE (F)	1517	29509	147381406	1052128
Waimes	Waimes	BE (W)	4	74	240000	1567
Anglian	Anglian Water	E-W	3221	36762	420071200	1820223
Bristol	Bristol Water	E-W	401	6553	104890050	439905
Bournemouth	Bournemouth and West Hampshire Water	E-W	170	2749	59272350	17258
Cambridge	Cambridge Water	E-W	119	2262	27115850	111214
Dee Valley	Dee Valley Water	E-W	182	1976	25480650	105173
Essex Suffolk	Essex and Suffolk Water	E-W	891	8446	170097300	692972
Folkestone	Folkstone and Dover Water	E-W	77	1080	16921400	66365
Mid Kent	Mid Kent Water	E-W	295	4209	58469350	219626
South East	South East water	E-W	448	9683	143397550	541598
Northumbrian	Northumbrian Water	E-W	1566	16879	259387250	1039909
United Utilities	United Utilities	E-W	3630	40741	702975400	2757159
Portsmouth	Portsmouth Water	E-W	215	3236	65404350	276916
Sutton Surrey	Sutton and East Surrey Water	E-W	259	3385	59531500	253849
Southern	Southern Water	E-W	2045	13424	214386400	955029
South Staffs	South Staffordshire water	E-W	343	5825	121614350	499724
Severn Trent	Severn Trent	E-W	5083	45949	698245000	3017000
South West	South West Water	E-W	1383	14991	160902950	6503
TendringHun	Tendring Hundred Water	E-W	67	907	11128850	64705
Thames	Thames Water	E-W	4616	31416	1026259550	3255831
Three Valleys	Three Valleys Water	E-W	1041	14315	313845250	1163429
Dwr Cymru	Dwr Cymru	E-W	134	27112	317320050	1347600
Wessex	Wessex Water	E-W	1696	11294	132495000	480863
Yorkshire	Yorkshire Water	E-W	2154	31217	472977950	1888567
Lisboa	Lisboa Epal	P(pub)	850	1690	255443624	336401
Sintra	Sintra	P(pub)	750	1600	33186571	177235
SMAS Oeiras	Smas Oeiras e Amadora	P(pub)	350	867	32252859	168509
SMAS Loures	Smas de Loures	P(pub)	490	1250	31127137	153530
Águas Gaia	Águas de Gaia	P(pub)	230	1432	19307187	117428
SMAS Porto	Smas Porto	P(pub)	421	769	41763910	152565
Vimágua	Vimágua	P(pub)	107	766	9082542	45733
Almada	Smas de Almada	P(pub)	367	746	17921000	100472

Table 19: Data appendix - continued

Abbreviation	Company	Country	employees (fte)	mains (km)	production (m^3)	connections
Coimbra	Águas de Coimbra	P(pub)	247	1250	18919692	76959
Xira	Smas Vila Franca de Xira	P(pub)	168	450	13311234	62596
Maia	Smas da Maia	P(pub)	125	521	8333554	50702
Leiria	Smas Leiria	P(pub)	126	1100	9769147	54357
Viseu	Smas Viseu	P(pub)	110	870	7226000	33393
Castelo	Sm Viana do Castelo	P(pub)	98	705	5951068	35172
Aveiro	Sma Aveiro	P(pub)	164	566	2548868	32433
Torres Vedras	Smas Torres Vedras	P(pub)	147	996	4717976	35868
Santarém	Santarém	P(pub)	156	824	7873325	32821
Castelo Branco	Castelo Branco	P(pub)	70	500	6192758	34732
Faro	Sm Faro	P(pub)	70	665	6638337	30252
Alcobaça	Alcobaça	P(pub)	53	500	5358240	26849
Caldas Rainha	Caldas da Rainha	P(pub)	47	355	4484392	28482
Covilhã	Covilhã	P(pub)	99	500	4370473	26152
Portimão	Portimão	P(pub)	95	305	7383371	38377
Guarda	Smas Guarda	P(pub)	56	234	3007488	22404
Abrantes	Sm Abrantes	P(pub)	90	354	2891603	22377
Vila Real	Vila Real	P(pub)	69	670	4208502	25430
Cantanhede	Cantanhede	P(pub)	46	340	4150334	16071
Beja	Emas Beja	P(pub)	75	270	3801963	19173
Esposende	Esposende	P(pub)	65	400	2515000	14711
Cascais	Águas de Cascais	P(pri)	271	1230	23256297	102028
Gondomar	Águas de Gondomar	P(pri)	215	800	13062868	73943
Braga	Agere Braga	P(pri)	275	850	13202356	71098
Maria Feira	Santa Maria da Feira	P(pri)	45	550	4145078	21354
Sado	Águas do Sado	P(pri)	165	632	12598703	58090
Indaqua Santo	Indaqua Santo/Tirso	P(pri)	40	228	2061671	13240
Valongo	Águas de Valongo	P(pri)	93	470	6083557	34148
Paredes	Águas de Paredes	P(pri)	35	220	3610228	5876
Figueira	Águas da Figueira	P(pri)	86	680	6075617	37661
Mafra	Cge Mafra	P(pri)	64	848	5158589	32079
Fafe	Fafe	P(pri)	33	278	2781588	13849
Ourém	Ourém	P(pri)	36	918	4082997	21549
Batalha	Batalha	P(pri)	15	206	1688802	6730
Alcanena	Alcanena	P(pri)	16	200	1805231	7413
Carrazeda	Águas de Carrazeda	P(pri)	20	194	568815	4839

References

ACT (1995). ACTEW benchmarked. *Auditor General's Office, Australia Capital Territory (ACT)*, 7, pp 15.

Aida, K., Cooper, W., Pastor, J. & Sueyoshi, T. (1998). Evaluating water supply services in Japan with RAM: a range-adjusted measure of inefficiency. *Omega*, 26(2), pp 207–232.

Ancarani, A. (2000). Evolution of water firms performance in providing water services in Sicily. *In EIASM International Conference on Accounting, Auditing & Management in Public Sectors Reform: Proceedings, Saragoca (Spain)*, p. 14.

Ancarani, A. and Gapaldo, G. and Raffa, M. (2000). A framework for evaluating quality level of standardised public services: an application to Italian water supply service. *In EIASM 10th Workshop on Quality Management in Services: Designing System Chain Processes for High Performance Services: Proceedings. Birmingham (UK), 17-19 May*, p. 20.

Andersen, P. & Petersen, N. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), pp 1261–1264.

Andrews, D. & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society.*, 40(1), pp 85–93.

Anwandter, L. & Ozuna, T. (2002). Can public sector reforms improve the efficiency of public water utilities?. *Environment and Development Economics*, 7(04), pp 687–700.

Bakker, K. (2003). From public to private to... mutual? Restructuring water supply governance in England and Wales. *Geoforum*, 34(3), pp 359–374.

Banker, R. and Charnes, A. and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, pp 1078–1092.

Banker, R., Charnes, A., Cooper, W., Swarts, J. & Thomas, D. (1989). An introduction to data envelopment analysis with some of its models and their uses. *Research in Governmental and Nonprofit Accounting*, 5, pp 125–163.

Beguin, C. & Simar, L. (2004). Analysis of the expenses linked to hospital stays: how to detect outliers. *Health Care Management Science*, 7(2), pp 89–96.

Bosworth, D., Stoneman, P. & Thanassoulis, E. (1996). The measurement of comparative total efficiency in the sewerage and water industry: an exploratory study. *Birmingham: Office of Water Services (Ofwat)*, p. 29.

Byrnes, P. (1986). Ownership and efficiency in the water supply industry: an application of the nonparametric programming approach to efficiency measurement. *University Microfilms International*, 8622963, pp 325.

Byrnes, P., Grosskopf, S. & Hayes, K. (1986). Efficiency and ownership: further evidence. *The Review of Economics and Statistics*, 68(2), pp 337–341.

Caves, D., Christensen, L. & Diewert, W. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *The Economic Journal*, 92(365), pp 73–86.

Cazals, C., Florens, J. & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1), pp 1–25.

Charnes, A. and Cooper W. and Rhodes E. (1978). Measuring efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp 428–449.

Charnes, A., Cooper, W., Golany, B., Seiford, L. & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1), pp 91–107.

Coelli, T., Prasada, R. & Battese, G. (2001). *An introduction to efficiency and productivity analysis*. second edition edn. Boston (Mass.): Kluwer.

Cooper, W., Li, S., Seiford, L. & Zhu, J. (2004). *Handbook on data envelopment analysis*. Kluwer Academic Publishers, Boston.

Cubbin, J. and Tzanidakis, G. (1998). Regression versus data envelopment analysis for efficiency measurement: an application to the England and Wales regulated water industry. *Utilities Policy*, 7(1), pp 75–85.

De Witte, K. (2006). Efficiëntieprikkels in de drinkwatersector. *Economisch Statistische Berichten*, 91(4485), pp 201–204.

De Witte, K. and Moesen, W. (2006). Sizing the government. *Working Paper, Centre for Economic Studies, KU Leuven, Belgium*, p. 23.

Dijkgraaf, E., de Jong, R., van de Mortel, E., Nentjes, A., Varkevisser, M. & Wiersma, D. (1997). Mogelijkheden tot marktwerking in de Nederlandse watersector. *Onderzoeksreeks directie marktwerking, Ministerie van Economische Zaken, The Hague*, pp 1–93.

Dijkgraaf, E., van der Geest, S. and Varkevisser, M. (2005). Efficiëntie boven water. *Economisch Statistische Berichten*, 4451, pp 34–35.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1), pp 1–26.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics Philadelphia.

Eggleton, I. (1994). A benchmarking methodology for the Australian water industry. *Urban Water Research Association of Australia*, 77, pp 10.

Escosa (2004). Inquiry into urban water pricing process. *Essential services commission of south australia*, p. 15.

Estache, A., Rossi, M. & Ruzzier, C. (2004). The case for international coordination of electricity regulation: evidence from the measurement of efficiency in South America. *Journal of Regulatory Economics*, 25(3), pp 271–295.

Färe, R. and Grosskopf, S. and Lovell, K. (1985). *The measurement of efficiency of production*. Kluwer-Nijhoff.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3), pp 253–290.

Førsund, F. & Hjalmarsson, L. (1979). Generalised farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *The Economic Journal*, 89(354), pp 294–315.

Fried, O., Schmidt, S. & Yaisawarng, S. (1999). Incorporating the operating environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, 12(3), pp 249–267.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, 14(4), pp 1453–1462.

Hicks, J. (1935). Annual survey of economic theory: the theory of monopoly. *Econometrica*, 3(1), pp 1–20.

IPART (2000). Developer charges from October 2000. *Independent pricing and regulatory tribunal, Sidney, Australia*, p. 47.

Jamasb, T. & Pollitt, M. (2001). Benchmarking and regulation: international electricity experience. *Utilities Policy*, 9(3), pp 107–130.

Keirsebilck, D. and Gellynck, X. (2006). *Socio-economisch belang van de watersector in Vlaanderen.* TNAV, Thematisch Netwerk Afvalwaterzuiveringstechnologie Vlaanderen.

Kittelsen, S. (1993). Stepwise DEA: choosing variables for measuring technical efficiency in Norwegian electricity distribution. , p. 37.

Kneip, A., Simar, L. & Wilson, P. (2003). Asymptotics for DEA estimators in nonparametric frontier models. , p. 37.

KS (2003). Konkurrenceredegorelse. *Konkurrencestyrelsen (KS), Copenhagen (Danmark)*, p. 451.

Kuks, S. (2001). *The privatisation debate on water services in the Netherlands. An examination of the public duty of the Dutch water sector and the implications of market forces and water chain cooperation.* in: Holzwarth, F., Andreas Kraemer, R. (Ed.), Umweltaspekte einer Privatisierung der Wasserwirtschaft in Deutschland (pp.81-112). Berlin: Ecologic.

Laffont, J. & Tirole, J. (1993). *A Theory of incentives in procurement and regulation.* MIT Press.

Lambert, D., Dichev, D. & Raffiee, K. (1993). Ownership and sources of inefficiency in the provision of water services. *Water Resources Research*, 29(6), pp 1573–1578.

Leibenstein, H. (1966). Allocative efficiency vs. 'X-efficiency'. *The American Economic Review*, 56(3), pp 392–415.

Liang, T. (2003). Empirical relative efficiency analysis of major Australian water service providers using DEA. *Global Developments in Water Industry Performance Benchmarking: Proceedings, Perth, Western Australia (Australia)*, p. 11.

Lobina, E. and Hall, D. (2001). *UK water privatisation - a briefing.* Public Services International Research Unit, PSIRU.

London Economics (1995). Benchmarking the economic performance of Australian urban water authorities. *Urban Water Research Association of Australia (UWRAA), Research Report*, 93, pp 66.

Manning, I. & Molyneux, E. (1993). Measurement of total factor productivity in major water utilities. *Urban Water Research Association of Australia, Melbourne*, p. 10.

Marques, R. (2005). Regulação de serviços públicos. *Edições Sílabo, Lisbon*, p. 404.

Marques, R. (2006). A yardstick competition model for Portuguese water and sewerage services regulation. *Utilities Policy*, 14(3), pp 175–184.

Marques, R. and Monteiro, A. (2003). Measuring productivity and efficiency of water and sewerage services - a case study in Portugal. *Global Developments in Water Industry Performance Benchmarking: Proceedings. Perth, Western Australia (Australia), 29 September to 2 October*, p. 18.

Marques, R. and Monteiro, A. (2004). Measuring efficiency and productivity of water and sanitation services - a benchmarking exercise in Portugal. *Conference Joint Management 2004: Proceedings. Phoenix, Arizona (USA), 13-16 March*, p. 18.

Marques, R. and Monteiro, A. (2005). Technical efficiency of the water and sewerage services. An empirical study. *3th International Conference Efficient Use and Management of Water for Urban Supply: Proceedings. Santiago (Chile): International Water Association (IWA), 14-18 March*, p. 8.

Mobley, L. & Magnussen, J. (1998). An international comparison of hospital efficiency: does institutional environment matter?. *Applied Economics*, 30(8), pp 1089–1100.

Norman, M. & Stoker, B. (1991). *Data Envelopment Analysis: the assessment of performance.* John Wiley & Sons, Inc. New York, NY, USA.

SA Government (2004). Statement transparency. *South Australian Government*, p. 74.

Sawkins, J. (2001). The development of competition in the English and Welsh water and sewerage industry. *Fiscal Studies*, 22(2), pp 189–215.

Seiford, L. and Thrall, R. (1990). Recent developments in DEA, the mathematical programming approach to frontier analysis. *Journal of Econometrics*, 46, pp 7–38.

Sexton, T., Silkman, R. & Hogan, R. (1986). Measuring efficiency: an assessment of data envelopment analysis, new directions for program evaluations. *Measuring Efficiency: An Assessment of Data Envelopment Analysis. Publication No 32 in the series New Directions for Program Evaluations*, p. 25.

Shephard, W. (1970). *The theory of cost and production functions.* Princeton Univeristy Press, New Jersey.

Simar, L. (2003). Detecting outliers in frontier models: a simple approach. *Journal of Productivity Analysis*, 20(3), pp 391–424.

Simar, L. and Wilson, P. (2006). *Statistical inference in nonparametric frontier models: recent developements and perspectives.* in The measurement of productive efficiency, 2nd edition, chapter 4, ed. by H. Fried, C.A.K. Lovell, and S.S. Schmidt, Oxford: Oxford University Press, forthcoming.

Simar, L. and Wilson, P. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), pp 31–64.

Simar, L. & Wilson, P. (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science*, 44(1), pp 49–61.

Simar, L. & Wilson, P. (2000). Statistical inference in nonparametric frontier models: the state of the art. *Journal of Productivity Analysis*, 13(1), pp 49–78.

Thanassoulis, E. (1997). The use of data envelopment analysis in the regulation of UK water utilities: water distribution. *Warwick Business School Research Paper*, 254, pp 34.

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

Thanassoulis, E. (2000b). The use of data envelopment analysis in the regulation of UK water utilities: water distribution. *European Journal of Operational Research*, 126(2), pp 436–453.

Torgersen, A., Førsund, F. & Kittelsen, S. (1996). Slack-adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis*, 7(4), pp 379–398.

Tupper, H. & Resende, M. (2004). Efficiency and regulatory issues in the Brazilian water and sewage sector: an empirical study. *Utilities Policy*, 12, pp 29–40.

Vlaams Parlement (2001). Voorstel van resolutie: betreffende de drinkwatervoorziening in Vlaanderen. *Stukken van het Vlaams Parlement*, 669(2), pp 1–79.

Waterspiegel (2005). Best practices over Europa verspreiden, pilot internationale benchmark van start. *Opinieblad van Vewin*, 8(6), pp 12–14.

Waterspiegel (2006). Efficiency waterbedrijven weer gestegen. *Opinieblad van Vewin*, 9(4), pp 9.

Williamson, B. & Mumssen, Y. (2000). Economic regulation of network industries. *A Report for the New Zealand Treasury, Prepared by National Economic Research Associates: London.*, 00/5, pp 55.

Wilson, P. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business and Economic Statistics*, 11(3), pp 319–323.

Wilson, P. (2005). FEAR 1.0: A software package for frontier efficiency analysis with R. *working paper, University of Texas, Austin, Texas*, p. 13.

Wood, D., Barrar, P. & Kodwani, D. (1997). Benchmarking water utilities' performance using data envelopment analysis. *Fifth European Workshop on Efficiency and Productivity Analysis, Copenhagen (Denmark), 9-11 October*, p. 15.

Woodbury, K. and Dollery, B. (2003). Efficiency measurement in Australian local government: the case of NSW municipal water services. *New South Wales (Australia): University of New England, Working paper*, 13, pp 40.

Zhang, Y. & Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis*, 9(3), pp 187–204.

Zhu, J. (2003). *Quantitative models for performance evaluation and benchmarking*. Kluwer Academic Boston.