KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and Applied Economics

Department of Economics

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

by

Kristof DE WITTE
Rui C. MARQUES

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 07.32
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

November 2007

DISCUSSION PAPER

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

November 27, 2007

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 (DEA), 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, robust 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 incentive regulation in the sense of regulatory and benchmark incentive schemes have a significant positive effect on efficiency. By suitably adapting the conditional efficiency measures to the bias corrected estimates, we incorporate environmental variables directly into the efficiency estimates. We firstly 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

*The authors acknowledge the many helpful comments of seminar participants at the University of Leuven (KUL), Université Catholique de Louvain (UCL), the EGPA in Madrid and the EWEPA in Lille. In particular, we would like to thank Laurens Cherchye, Wim Moesen, David Saal, Elbert Dijkgraaf, Jan van Helden, representatives of the Belgian and Portuguese drinking water sector and two anonymous referees for their very useful comments.

1 Introduction

The merits of competition are abundantly demonstrated in economic theory. However, a monopolistic configuration may 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 broad range of incentive schemes in order to reach this goal. The different institutional frameworks (e.g., divestiture, concession or yardstick competition) reflect the different regulatory and ideological views among societies. Especially within local public utilities, ideological views could 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 competition law.

In this article, we 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. To make abstraction of ideological conflicts, we are considering efficiency. Indeed, whatever the ideological background, no one can accept inefficiencies which are, basically, resources left over on the table. This article compares 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. To best our knowledge, this is the first paper applying international benchmarking in the water sector to the developed countries and trying to determine the best incentive scheme towards efficiency maximization.

In methodological terms, this paper follows the literature on Data Envelopment Analysis (DEA). This nonparametric 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 suffer from some serious inconveniences which are dealt with in this article. Firstly, the models, as 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 (i.e., with and without considering the exogenous environment). These outcomes are compared to the ones arising from the more traditional Tobit regressions with censored and truncated samples. Secondly, the deterministic frontier models are sensitive to outlying and atypical observations. Following Simar (2003), we apply the robust order-m efficiencies of Cazals *et al.* (2002) to detect the outlying observations in the sample. These results are compared with the more traditional outlier detection procedures of *peer count* (Charnes *et al.*, 1985), *super-efficiency* (Andersen and Petersen, 1993), *peer index* (Torgersen *et al.*, 1996) and the Wilson method (1993). Thirdly, when comparing the efficiency of entities in data sets with a heterogeneous size, Zhang and Bartels (1998) point out that the average efficiencies can not simply be compared. Therefore, we develop an approach based on the bandwidth of the Kernel estimates and employ the approach to stipulate the homogeneity of a country's drinking water sector. Finally, almost all two stage procedures suffer from the separability condition in that the exogenous environmental variables do not directly influence the estimated efficiency scores. Only recently, by the conditional efficiency estimates of Daraio and Simar (2005, 2007), this issue has been tackled. We suitably adapt the conditional efficiency analysis to the bias correction framework in order to obtain the *conditional bias corrected efficiency* estimates.

The paper is organized as follows. Section 2 describes 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. In addition, it develops and applies the conditional bias corrected efficiency estimates. Finally, Section 7 provides some concluding remarks.

2 The institutional framework in the water sector

Many approaches are suggested to solve the principle-agent problem (see, e.g., Laffont and Tirole, 1993). Although every government wants a secure drinking water provision at a price as low as possible, countries have different ideological views on the extent of state intervention in the economy which creates different incentive schemes. In this section, we compare the incentive schemes implemented 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* denotes the process of comparing the current performance of a utility with a reference performance. Therefore, it is only a tool to improve performances and not a regulatory method *per se*. The regulatory methods include the consequences and the effects of the use of benchmarking, e.g., employed in yardstick competition. In the water sector, the two existing types of yardstick competition (also referred to as competition by comparison) are 'price yardstick competition' and 'sunshine regulation' (Marques, 2006). The former 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 an inferior performance by a public discussion of the efficiency scores. Even if sunshine regulation is not triggered compulsorily (e.g., by a sector-specific regulator), the public display of the efficiency levels provides transparency in the sector and generates a competitive pressure which prevents the quiet life of Hicks (1935) and the X-inefficiency of Leibenstein (1966). 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 on the privatization of water services. The issue was driven by the Dutch 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 waste water treatment) was strongly opposed to the privatization idea. Therefore, the Dutch water companies tried to escape government regulation by using self-regulation and, in particular, by a voluntary benchmark organized by the Association of Dutch Drinking Water Companies (Vereniging voor Waterbedrijven in Nederland, VEWIN) in 1999, 2004 and 2007. The results are remarkable as in a sector with only very low technological change, the efficiency gains over the period 1997-2006 reached 23%. In addition, 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. Nowadays, 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. However, recent research indicates that the efficiency gains can not be attributed to merger economies, all the more to the incentive mechanisms (De Witte and Dijkgraaf, 2007).

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 entails 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 of relevance. 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 which is determined by comparing the performances of the water utilities (i.e., by benchmarking). 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.

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 USA older regulatory models. Thanks to the regulatory procedures close to the American ideas of transparency, enactment and accountability and to the typical UK performance incentives through benchmarking and yardstick competition, a unique incentive scheme based on strict regulation arose. Note that, Australia has been the pioneer of benchmarking in the water industry as even prior to the first American and English studies, Australia developed already research documents about benchmarking. 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 (National Competition Policy, 1998).

2.4 Portugal

In Portugal, except for Lisbon, the water service responsibility belonged until the 1990s exclusively to the municipalities. Private sector participation has only been allowed since 1993. In addition, the reform created the 'multimunicipal systems' which provide 'bulk' water to at least two municipalities and require a predominant investment by the State for reasons of national interest. The remaining organizational forms are called 'municipal systems', even though they could be managed by an association of municipalities. The regulatory reform 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, Empresa Portuguesa de Águas Livres (EPAL), is responsible for the water service of Lisbon,but it embodies an atypical situation in Portugal.

2.5 Belgium

Although Belgium is a federal country and the drinking water supply has been a regional policy since 1980 (i.e., competition of 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. By law, drinking water supply is the responsibility of the municipalities although four different organizational structures can be distinguished. Firstly, both in the Flemish and the Walloon region, municipalities have organized themselves into 'intercommunales'. *Intercommunales* are a typical Belgian structure which gives the organized municipalities corporate personality. Secondly, if the municipalities 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 and fourth organizational structure are *municipal water suppliers* and *municipal services.* These utilities, respectively, do have and do not have technical and financial autonomy. There are no structural incentives in the sector (De Witte, 2006).

3 International benchmarking by DEA

In this study, we will 'benchmark' the Dutch, English and Welsh, Australian, Portuguese and Belgian drinking water utilities against each other. To obtain a comparison of the current performance against a reference performance (and hence to benchmark), we assume a common frontier technology, allowing utilities from different countries to support the envelope. Alternatively, we can establish a national frontier production function in which only a country's own firms may be best practices. Since a whole range of methodologies exist to determine performance scores based on production estimates, it is important to identify the strengths and limitations of these techniques (for a survey, see Berg, 2006). We use Data Envelopment Analysis (DEA) to estimate the production frontier. We first focus on 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). In this article, we develop a fourth advantage of cross-country comparisons as we would like to examine objectively the effectiveness of incentive schemes. Therefore, in an international data set, 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 firm's management, we evaluate the effect on efficiency of a benchmarking and regulation incentive scheme.

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 differently. Even in national benchmark studies, interpretation of definitions and measurement of variables could differ. Secondly, exchange rate fluctuations are important when comparing monetary units. Thirdly, the unequal extent of outsourcing in the different countries influences the total number of employees (and the staff cost). Fourthly, some country specific differences are beyond the control of the firms' managers. Dissimilarities such as wage rates, taxes or rates of return on capital could induce different policy options (Jamasb and Pollitt, 2001). Finally, heterogeneity creates differences between countries which could be falsely taken as inefficiencies.

In this article, we try to take into account these concerns by focusing on four specific assumptions. Firstly, we adopt variables in quantities (e.g., the inputs staff and mains length) that are less susceptible to the lack of comparability. Secondly, the major differences among countries are related to taxation issues. However, thanks to the quantities variables, tax heterogeneity does not significantly influence the model. Indeed, higher water prices should reduce the consumption, although this is not empirically observed. Nevertheless, in second stage analysis, we include variables for water consumption per capita and relative wealth of consumers. Thirdly, also other heterogeneous factors which characterize the operational and institutional environment are integrated in the second stage analysis of the model. Fourthly, an important aspect not completely integrated in this article is the level of outsourcing. Although these data are not available, we contacted the professional associations of the water sector which confirmed that the degree of outsourcing is more or less the same among the countries in the analysis. Notice for example, in our sample, the Australian water services, which may seem very different from the others, correspond only to major cities of Australia, and thus, are more comparable to the other observations. Fifthly, we identify the outliers by several methods, so that atypical observations are eliminated from the sample and homogeneity of the data set increases.

3.2 Determining efficiency

From 1985 until the beginning of 2006, around 40 DEA applications to the water services were carried out. The case-studies which were made public amount to 30. The most frequently cited studies are referred to in Table 1 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. Out of these case-studies, 12 comprise both the water supply, the sewerage and the sewage treatment; 12 comprise only the water supply; 3 the sewerage and 3 the sewage treatment separately. The 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 operational expenditures ($OPEX$), the energy and the mains length. The leading outputs are the distributed (revenue) water volume, the number of customers and the network length, while the main 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 2 systematizes the inputs, outputs and explanatory factors which are used more than three times by at least more than one article.

Table 1: Main DEA studies

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 (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 (2000a, 2000b)	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%

3.3 Data Envelopment Analysis model

The DEA approach constructs the above mentioned nonparametric frontier as the piecewise linear combination of all efficient Decision Making Units (DMUs) in a sample. The generic DEA model was proposed by Charnes, Cooper and Rhodes (CCR) (1978). As their model assumed constant returns to scale (CRS), Banker *et al.* (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 2: 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 consuming a varying amount 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 the evaluated observation by the subscript 'o'. The DEA-CRS model with input-orientation, which searches for the minimum proportion of inputs to produce the same amount of outputs, is 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

$$\frac{\sum_{r=1}^{s} \mu_r y_{ri}}{\sum_{j=1}^{m} \nu_j x_{ji}} \leq 1, \qquad i = 1, 2, \ldots, n$$
$$\mu_r > 0, \nu_j > 0, \qquad for\ all\ r, j.$$

To obtain the technical efficiency score θ for each of the n DMUs, the linear programming problem needs to be repeated n times. 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 $\theta = 1$ while an inefficient DMU is denoted by $\theta < 1$.

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, by construction, joining separate data sets does not increase the efficiency scores of the individual DMUs in comparison with the separate analysis. This is an important aspect in international benchmark studies as the combination of national databases increases the number of observations.

4 Data and indicators

Choosing the input and output variables is the most important stage in any DEA assessment as the results are highly influenced by this choice (Section 4.1). Kittelsen (1993) proposes a statistical procedure to analyze the selection of the variables, which we apply to the choice of the orientation (i.e., input versus output) and the option of returns to scale (i.e., constant versus variable) (Section 4.2). Section 4.3 compares the homogeneity in efficiency among the countries in the data set. We partly tackle the intricate comparability of data 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 carefully examine the results and test them on outliers (Section 4.4). In international comparisons, it is appealing to estimate the technical efficiency of companies as the goal of maximizing technical efficiency is not in conflict with any other goals. Indeed, inefficiencies are, basically, resources left over on the table.

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 which collects information from each of the water companies. Most data tables contain information from the 1997-2005 period. The Australian data are obtained by means of 'Water Services Association of Australia facts', published annually (since 1996) by the Water Services Australian Association that compiles and audits the data. The data of the Portuguese water services are collected directly by the annual accounts and activity reports produced by the utilities. As some technical data were sometimes missing in the reports, the companies were contacted in order to provide them. Data on the Belgian water industry are compiled by Belgaqua, the Belgian umbrella organization, since 1993. In contrast to other countries, the most recent year available is 2004.

Descriptive statistics for the various countries are presented in Table 3. 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 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

Table 3: Descriptive 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 4: 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	121	2.4340	0.017
Wilcoxon signed-rank		2.8620	0.014

Table 5: 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	121	6.3090	0.0000
Wilcoxon signed rank		6.3150	0.0000

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 (GRP) in the second stage (see *infra*), we try to control for 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 from variations in estimating current construction and exchange rates. The outputs in the model reflect the main activities from the drinking water companies, i.e., the companies have to deliver water to their customers. We use the volume of delivered water as a first output indicator (in m^3), while the number of connections is applied as the second output variable.

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 data set, 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 of Kittelsen (1993) to decide on the orientation of the DEA model. 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 hypothesis 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 the ordinary paired t-test (in particular, two sided for the input-output test, while one sided for the CRS-VRS test), while the median efficiencies are compared by the nonparametric 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 test results are presented in Table 4. Firstly, the input-oriented model does significantly differ from the output-oriented model, so that a choice has to be made. As water utilities are obliged to supply all customers and as they cannot encourage the consumption (i.e., 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. Secondly, 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. The test results in Table 5 show that the CRS-model significantly differs from the VRS-model, so that a choice with respect to returns to scale has also to be made. We apply the VRS model as this assumption is less stringent and ensures that DMUs are only compared with 'similar' DMUs. In addition, the water utilities cannot change its size in the short-term.

4.3 Homogeneity in efficiency

By restricting the data set 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.

Table 6: 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

Hence, as in De Witte (2006), we 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 firms. If all companies in the data set are rather similar (i.e., 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 the average efficiencies cannot simply be compared. On average, the technical efficiency score of a DMU will decrease as the sample size increases. To equalize the size of the data sets, we resample the efficiency scores by a procedure which comprises three steps. Firstly, for every country, we compute the input-oriented DEA-VRS efficiency scores as described in Section 4.2. Secondly, by the use of the code 'FEAR' (Wilson, 2005), we determine the optimal bandwidth for a nonparametric Kernel function of the efficiency estimates. Following Simar and Wilson (2006), we use the unbiased cross-validation criterion which minimizes the estimate of the mean-integrated square error. Although the Kernel density estimate also depends on the size of the data set, the sample size issue is only of minor importance (cf., the empirical rule approximating the cross-validation principle equals $h = 1.06 \min(s_{2m}, r_{2m}/1.34)(2m)^{-1/5}$ where $2m$ denotes the reflected data, s_{2m} the standard deviation of $2m$ and r_{2m} the interquartile range). The bandwidth of every country is presented in Table 6. Finally, we resample the original DEA-VRS efficiency scores to obtain s values drawn from a Kernel estimate of the bounded density of the efficiency estimates (i.e., the estimates are bounded above at one). We set s equal to 44, the size of the largest data set. The average resampled efficiencies are presented in the third column of Table 6. Remark that the sample size bias could have been avoided by the use of the robust order-m efficiency scores of Cazals *et al.* (2002) as well (for a simulated example, see De Witte and Marques, 2007a).

It turns out that the efficiency in the Belgian drinking water sector is the most homogeneous, 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 utilities. 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.

4.4 Outlier detection

Influential data affect the efficiency results of a significant number of other DMUs. 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, 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.

4.4.1 Theoretical background

A major drawback of DEA lies in its deterministic nature, in that the frontier model assumes

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

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. Outlying observations could arise from measurement errors, noise and influential observations (e.g., atypical data) or observations with favorable values on a specific variable. Although our data are obtained from national regulators and sector organizations, measurement errors could arise from the different definitions operated. Therefore, outlier detection procedures are employed in this and the next section. Due to the specific characteristics of each outlier procedure, several techniques to identify outliers have to be evaluated (De Witte and Marques, 2007b).

Firstly, a simple outlier determination procedure is the computation of the '*peer count index*' (Charnes *et al.*, 1985). This involves the computation of the number of times an efficient DMU is peer of an inefficient DMU. Both higher and lower values point to the presence of outliers.

Secondly, Andersen and Petersen (1993) compute the *super-efficiencies* which calculate 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, so that the efficiency can be higher than 1. Observations with high values are suspected to be outliers.

Thirdly, 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} denotes 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.

Fourthly, 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.

Finally, Simar (2003) uses the robust order-m efficiencies of Cazals *et al.* (2002). 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. 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.

Detecting outliers

The results of the peer index, super-efficiency, peer count and Wilson are presented in Table 7, while order-m efficiencies are shown in Table 8. The former four 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). As the order-m results 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

Table 7: 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 8: 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

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, from the sample of 122 observations we eliminate the 15 outlying DMUs as determined by the more traditional procedures, so that we obtain a data set 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, while this section tackles 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 estimated efficiency score, $\hat{\theta}(x, y)$, is an upward-biased estimator of the true efficiency score $\theta(x, y)$ (for an extensive discussion, see Simar and Wilson, 2006). The difference is visualized in Figure 1. The bootstrap (with B bootstrap replications) mimics this estimation and creates a pseudo frontier from which it provides estimates of the sampling distributions of the bias term $\hat{\theta}(x, y) - \theta(x, y)$. These bootstrap ideas are presented in Figures 2 and 3. For practical reasons we invert the efficiency scores:

$$\hat{\delta}(x, y) = \frac{1}{\hat{\theta}(x, y)}. \tag{5}$$

Indeed, as $\hat{\delta}(x, y) \geq 1$ for all $(x, y) \in \Psi$, we only have to deal with one boundary condition for $\hat{\delta}$, not two as in the case of $\hat{\theta}$. Although the literature describes several approaches to simulate a bootstrap sample χ_n^* (from the original sample χ_n), only the homogeneous smoothed bootstrap is here introduced. This approach assumes the distribution of the efficiency scores to be homogeneous over the input-output space (compare with a homoskedasticity assumption in linear regression models), which allows us to base the bootstrap on the sample estimates

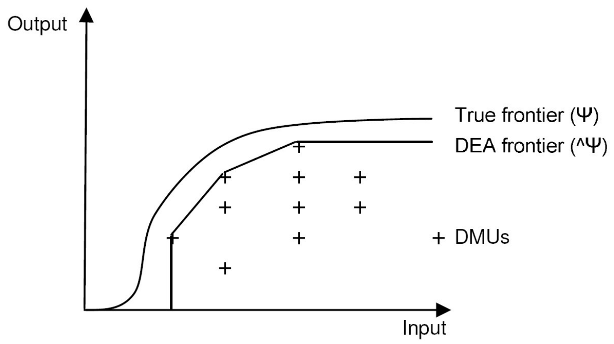

Figure 1: The true and the DEA frontier

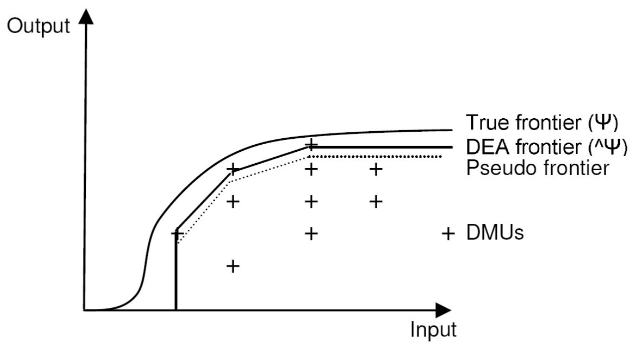

Figure 2: Bootstrap idea

$\hat{\delta}_i(x_i, y_i)$. The eight steps of the bootstrap algorithm are presented in Simar and Wilson (2006).

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

$$\widehat{BIAS}_B(\hat{\delta}(x, y)) = B^{-1} \sum_{b=1}^{B} \hat{\delta}^*(x, y) - \hat{\delta}(x, y). \tag{6}$$

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

$$\hat{\hat{\delta}}(x, y) = \hat{\delta}(x, y) - \widehat{BIAS}_B(\hat{\delta}(x, y)). \tag{7}$$

Confidence intervals are obtained by means of the percentile method.

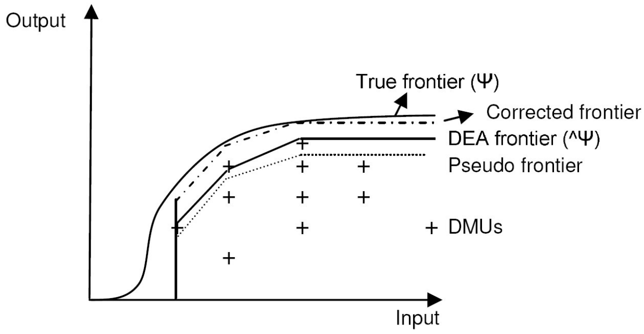

Figure 3: Bootstrap idea (2)

Table 9: Average first stage bootstrap results

	Uncorr. eff. ($\hat{\delta}$)	Bias-corr. ($\hat{\hat{\delta}}$)	$\widehat{Bias}$	Variance ($\hat{\sigma}^2$)	Lower bound	Upper bound	Weighted bias-corr.
the Netherlands	1.2158	1.3956	-0.1797	0.0092	1.2343	1.5692	1.3307
England and Wales	1.3583	1.5502	-0.1919	0.0117	1.3792	1.7475	1.4124
Australia	1.4528	1.6653	-0.2125	0.0120	1.4758	1.8575	1.4801
Portugal	1.7540	1.9093	-0.1553	0.0072	1.7776	2.0710	1.7776
Belgium	1.5998	1.8026	-0.2028	0.0111	1.6245	1.9940	2.2693
Portugal - public	1.7607	1.9081	-0.1474	0.0062	1.7843	2.0584	1.7450
Portugal - private	1.7415	1.9117	-0.1702	0.0089	1.7651	2.0946	1.8836
All countries	1.5628	1.7438	-0.1809	0.0096	1.5856	1.9230	1.5243

5.2 First stage results

Having cleared the data set from outlying observations, we proceed to the first stage analysis where the input-oriented DEA-VRS model is applied. Average results per country are presented in Table 9 (detailed results available upon request). The average efficiency amounts to 1.56. This indicates that an average DMU could decrease its inputs by 35.9% (i.e., 1-1/1.56) while keeping its outputs constant, if it performed as efficiently 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 DMU with an efficiency score higher than 1 is relatively inefficient with respect to its benchmarks.

Recognizing the presence of random noise in the data, we account for noise in the data (although we can not *remove* this effect) by estimating DEA corrected efficiencies with their 95%-confidence intervals. The noise resembles the missing data, the imperfect quality of the data (even after auditing by regulators) and the atypical results. By a homogeneous bootstrap, we generate 200 bootstrap samples (the estimated bandwidth yields $h = 0.2469$ and $B = 200$). The average bias-corrected estimates are displayed in Table 9. We clearly notice the upward-bias in the original estimates (i.e., the unconditional estimates are more efficient). In addition, we present the bootstrap estimates of 95%-confidence intervals for which the average width amounts to 0.337. There is only little difference in the confidence interval bandwidth among the five countries.

Due to the upward-bias in the original estimates and due to the bootstrap correction in the confidence intervals, the original uncorrected 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. In general, it is important to note that due to the overlap among the confidence intervals, making a relative comparison among the firms is an intricate issue. In addition, 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. Indeed, as the true efficiency of a DMU cannot exceed 100%, and we measure the 95%-confidence intervals, this is a correct observation.

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). Here, we opt to weight the efficiency scores by the number of connections as this is a measure for the number of people who are affected by the relative (in)efficiency of a company. The average weighted efficiency scores are presented in the last column of Table 9 and reveal that 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.

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 important for evaluating the cost of regulation. If in an input-oriented model an environmental variable z is unfavorable on efficiency, the variable can be considered 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. Section 6.1 addresses the more conventional Tobit regression and the use of bootstrapping in a second stage analysis. Section 6.2 applies these theories to the drinking water utilities. In Section 6.3, 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:

$$\hat{\delta}_i = z_i\beta + \epsilon_i \tag{8}$$

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, by construction, the efficiency estimates $\hat{\delta}_i$ $(= \frac{1}{\hat{\theta}_i})$ have a lower limit of 1 which creates a concentration of observations at this single value. This leads to a censored sample. 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. 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 $(\hat{\delta}_i)$. Secondly, also 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 $\hat{\delta}_i$, valid confidence interval estimates for the parameters in the second-stage regression.

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 Section 6.3, we discuss and implement procedures which avoid these assumptions.

6.2 Second stage results

Although many elements in the physical, social as well as the institutional environment highly influence the cost level of the drinking water utilities, they lie outside the control of the firms' managers. However, due to the lack of (uniform) data, we make simplifications on the exogenous variables.

A first physical variable included in the second stage model is the percentage of leakage. This variable captures the geographical relief (i.e., more hilly landscape requires more pressure on the network of pipes which could cause more easily leakage) and the extent of maintenance (i.e., 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 on 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 on 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 on efficiency (i.e., in accordance to the literature we expect to find diseconomies of scale; De Witte and Dijkgraaf, 2007). 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 on efficiency.

In this subsection, we first evaluate the importance of the environmental influences. 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, therefore, 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, so 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 10.

Although in 3 out of 11 estimates the bootstrapping algorithm discovers the opposite sign of the Tobit regressions, in only 2 cases the Tobit estimates are not covered by the 95%-confidence intervals of the bootstrapped variables (see Table 11). To get an idea on how strongly the different estimates are related, we present the correlation coefficients in Table 12. 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 second stage results in Table 10 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 on 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 on 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 on efficiency of the variable delivery in one municipality suggests that the water utilities in the sample studied do not have scale economies, an observation in line with the literature. 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 regulation on efficiency. Finally, the effect of benchmarking on efficiency is positive and always with significance. This tool to improve performance turns out to be very appropriate.

Table 10: 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 11: 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.

Table 12: 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)

6.3 Taking into account environmental variables

The above mentioned separability assumption presumes that the environmental variables do not directly influence the efficiency scores, so that only *ex post* the influence of environmental variables on efficiency can be measured. To avoid this assumption, we suitably adapt the conditional efficiency measures of Daraio and Simar (2005, 2007) to the statistical inference framework of Simar and Wilson (1998, 2000). The integration of these two frameworks has some attractive features as (1) avoiding the separability condition, (2) creating a fully nonparametric model (i.e., it does not rely on any *a priori* assumption on the functional form of the production set) and (3) allowing to explore the influence of the environmental variables.

The conditional efficiency measure of Daraio and Simar (2005) introduces environmental variables z in the efficiency scores of the evaluated DMU by conditioning on the environmental variable z_o. They propose to use a Kernel with compact support (i.e., $K(u) = 0$ if $|u| > 1$) to smooth the variable z and to determine the appropriate bandwidth by the use of the cross-validation principle. The procedure selects all observations (x, y) in the neighborhood of z_o :

$$\chi_n^{DS} = \{i \mid (x_{i,}y_i) \in \chi_n; |z_o - z_i| \leqslant h\}\,. \tag{9}$$

In a multidimensional framework, we first decorrelate the environmental variables by the use of a Mahalanobis transformation (see, e.g., Mardia *et al.*, 1979) and afterwards perform a sequential (Epanechnikov) Kernel estimation.

The Daraio-Simar procedure is suitably adapted to the Simar-Wilson model by, firstly, computing for each evaluated observation (x_o, y_o) the appropriate reference set and, secondly, determining the noise corrected efficiency score by the bootstrap procedure, so that we obtain a *conditional bias corrected efficiency* estimate $\hat{\hat{\delta}}(x, y \mid z)$. As an exploratory tool to visualize the effect of the environmental variable, Daraio and Simar (2005, 2007) suggest to nonparametrically regress the ratio $\hat{\hat{\delta}}(x, y \mid z) \ / \ \hat{\hat{\delta}}(x, y)$ (i.e., the conditional bias corrected efficiency estimates to the unconditional estimates) against z_i. In the input-oriented model, a decreasing regression indicates a favorable effect on efficiency of z (i.e., if not accounted for, the efficiency score will goes up with z), while an increasing regression specifies an unfavorable effect on efficiency (i.e., behaving as a substitutive input). In a multivariate framework, we nonparametrically regress the ratio of the fully conditioned variables to the partially conditioned efficiency scores against the conditioned variable.

In this article, we are specially interested in the effects of incentive regulation to the efficiency while taking into account the exogenous influences. In particular, by employing the conditional efficiency measures, we control for those exogenous characteristics which were significant in the DEA-VRS bootstrap method (i.e., controlling for leakage, industry delivery, GRP, consumption per capita, economies of scope, economies of scale, corporatization, regulation and benchmarking). Therefore, we can interpret the conditional bias corrected efficiency scores as if the utilities are facing the same physical, social and institutional constraints and benefits. In comparison to the first stage results, the conditional efficiency estimates, as presented in Table 13, reduce the efficiency of the English and Welsh, the Dutch, the Australian and the private Portuguese utilities (although, the estimated bias is similar). The companies obtained higher first stage estimates thanks to favorable environmental influences. Next, we estimate the conditional efficiencies without taking into account the effect of incentive regulation (i.e., the benchmarking and regulation dummy). As efficiency is reduced in all countries, we expect to find a favorable influence from incentive schemes on efficiency. This is tested by the exploratory graph in Figure 4 where the ratio of the fully conditioned estimates (i.e., the full set of exogenous characteristics including incentive regulation) to the partially conditioned estimates (i.e., without incentive regulation) is regressed against the presence of incentive regulation. The negative monotonic first order effect indicates that incentive regulation is favorable on efficiency, thus, utilities which are facing incentive regulation, in the sense of benchmarking or regulation, are producing more efficiently.

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 incentive regulation. 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.

Table 13: Conditional bias corrected efficiencies

	z = physical, social and institutional				z = physical and social			
	Bias-corr. ($\hat{\hat{\delta}}$)	$\widehat{Bias}$	Lower bound	Upper bound	Bias-corr. ($\hat{\hat{\delta}}$)	$\widehat{Bias}$	Lower bound	Upper bound
the Netherlands	1.820	-0.170	1.651	2.152	2.033	-0.175	1.859	2.402
England and Wales	2.120	-0.214	1.906	2.448	2.148	-0.218	1.930	2.456
Australia	1.613	-0.161	1.452	1.847	1.732	-0.160	1.572	1.969
Portugal	1.696	-0.147	1.549	1.913	1.702	-0.152	1.552	1.912
Belgium	1.838	-0.258	1.580	2.148	1.831	-0.251	1.581	2.129
Portugal - public	1.724	-0.142	1.581	1.932	1.721	-0.146	1.584	1.933
Portugal - private	1.645	-0.156	1.489	1.875	1.665	-0.165	1.493	1.873
All countries	1.805	-0.185	1.621	2.073	1.845	-0.187	1.660	2.108

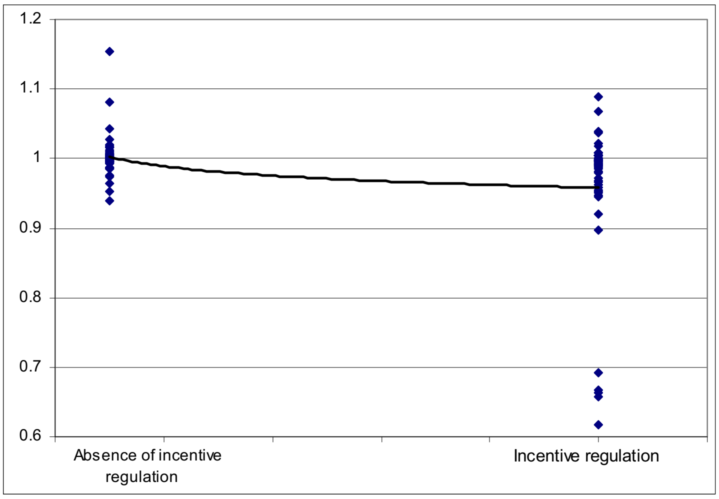

Figure 4: Impact of incentive regulation

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 incentive schemes implemented in the Netherlands, England and Wales, Australia, Portugal and Belgium. Our results show large differences in bias and noise corrected first stage inefficiencies. On average, the benchmarked Dutch drinking water companies are performing better (average efficiency score of 1.40) than the privatized English and Welsh utilities (1.55). However, the strict regulatory model of Australia (1.66), the municipal provision in Belgium (1.80) and especially the Portuguese municipal provision with private sector participation (1.90) 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 equalized the sizes of the data sets. 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 physical, social 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 physical, social and institutional environmental factors in the efficiency scores by suitably adapting the conditional efficiency measures of Daraio and Simar (2005) to the bias correction model of Simar and Wilson (1998, 2000). The conditional bias corrected efficiency estimates 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. We noticed that the Dutch, English and Welsh, Australian and private Portuguese utilities are working in a favorable environment. In addition, our results provide significant evidence for the positive effects of incentive schemes on 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. 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.

References

[1] ACT (1995), ACTEW benchmarked, Auditor General's Office, *Australia Capital Territory (ACT)* 7, p. 15.

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

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

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

[5] Andersen, P. and Petersen, N. (1993), A procedure for ranking efficient units in data envelopment analysis, *Management Science* 39 (10), 1261-1264.

[6] Andrews, D. and Pregibon, D. (1978), Finding the outliers that matter, *Journal of the Royal Statistical Society* 40 (1), 85-93.

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

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

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

[10] Berg, S. (2006), Survey of benchmarking methodologies in the water sector, *PURC Working Paper, University of Florida*, 1–17.

[11] Bosworth, D., Stoneman, P. and 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.

[12] Byrnes, P. (1986), Ownership and efficiency in the water supply industry: an application of the nonparametric programming approach on efficiency measurement, *University Microfilms International*, p. 325.

[13] Byrnes, P., Grosskopf, S. and Hayes, K. (1986), Efficiency and ownership: further evidence, *The Review of Economics and Statistics* 68 (2), 337-341.

[14] Cazals, C., Florens, J. and Simar, L. (2002), Nonparametric frontier estimation: a robust approach, *Journal of Econometrics* 106 (1), 1-25.

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

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

[17] 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), 75-85.

[18] Daraio, C. and Simar, L. (2005), Introducing environmental variables in nonparametric frontier models: a probabilistic approach, *Journal of Productivity Analysis*, 24 (1)n 93-121.

[19] Daraio, C. and Simar, L. (2007), *Advanced robust and nonparametric methods in efficiency analysis*, Series: Studies in Productivity and Efficiency, Springer.

[20] De Witte, K. (2006), Efficiëntieprikkels in de drinkwatersector, *Economisch Statistische Berichten* 91 (4485), 201-204.

[21] De Witte, K. and Dijkgraaf, E. (2007), Mean and bold? On seperating merger economies from structural efficiency gains in the water sector, *mimeo.*

[22] De Witte, K. and Marques, R. (2007a), Environmental Factors in Frontier Models, a Metafrontier Approach to the Drinking Water sector. *mimeo.*

[23] De Witte, K. and Marques, R. (2007b), Detecting Influential Observations in Frontier Models, a Robust Approach. *mimeo.*

[24] Dijkgraaf, E., de Jong, R., van de Mortel, E.G., Nentjes, A., Varkevisser, M. and Wiersma, D. (1997), Mogelijkheden tot marktwerking in de Nederlandse watersector, *Onderzoeksreeks directie marktwerking, Ministerie van Economische Zaken, The Hague,* 1-93.

[25] Efron, B. (1979), Bootstrap methods: another look at the jackknife, *The Annals of Statistics* 7 (1), 1-26.

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

[27] Farrell, M. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society* 120 (3), 253-290.

[28] Hicks, J. (1935), Annual survey of economic theory: the theory of monopoly, *Econometrica* 3 (1), 1-20.

[29] Jamasb, T. and Pollitt, M. (2001), Benchmarking and regulation: international electricity experience, *Utilities Policy* 9 (3), 107-130.

[30] Kittelsen, S. (1993), Stepwise DEA: choosing variables for measuring technical efficiency in Norwegian electricity distribution, *University of Oslo, Department of Economics,* p. 37.

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

[32] 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*, 81-112.

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

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

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

[36] 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.

[37] Lobina, E. and Hall, D. (2001), UK water privatisation - a briefing, *Public Services International Research Unit, PSIRU*, 1-27.

[38] London Economics (1995), Benchmarking the economic performance of Australian urban water utilities, *Urban Water Research Association of Australia (UWRAA), Research Report* 93, p. 66.

[39] Mardia, K., Kent, J. and Bibby, J. (1979), *Multivariate analysis*, Academic Press New York.

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

[41] 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.

[42] 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.

[43] National Competition Policy (1998), *Compendium of National Competition Policy Agreements*. Second Edition, National Competition Policy, Melbourne.

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

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

[46] Simar, L. and Wilson, P. (1998), Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models, *Management Science* 44 (1), 49-61.

[47] Simar, L. and Wilson, P. (2000), Statistical inference in nonparametric frontier models: the state of the art, *Journal of Productivity Analysis* 13 (1), 49-78.

[48] Simar, L. and Wilson, P. (2006), Statistical inference in nonparametric frontier models: recent developments 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.

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

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

[51] 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), 436-453.

[52] Torgersen, A., Forsund, F. and Kittelsen, S. (1996), Slack-adjusted efficiency measures and ranking of efficient units, *Journal of Productivity Analysis* 7 (4), 379-398.

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

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

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

[56] Wood, D., Barrar, P. and 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.

[57] 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, p. 40.

[58] Zhang, Y. and 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), 187-204.