Capturing the Environment, a Metafrontier Approach to 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 14, 2008

Abstract

Environmental factors add complexity to the comparison between specific activities or entire entities. Decision making units with an inferior performance are tempted to invoke that their organization is 'different' from the others in the data set. By reinterpreting and extending the metafrontier literature, we propose an all-embracing concept to fully capture the operational environment. We suggest the 'Group Specific Technical Efficiency' as a new measure to assess the overall efficiency of a utility while allowing for environmental differences. A real-world example of drinking water utiles out of 5 different countries illustrates the concept.

JEL Classification: C14, C61, D24, L95

Keywords: Free Disposal Hull, Efficiency Measurement, Environmental Variables, Metafrontier, Water Industry

1 Introduction

The performance comparison of an entity with a reference entity has been widely applied for both managerial and academic purposes. There are several examples in different fields, such as health, utilities (water, waste, energy, etc.), defence, education, justice, either in

*We thank A. Brochado, L. Cherchye, W. Moesen and the seminar participants at Université Catholique de Louvain (UCL) and the Tenth European Workshop on Efficiency and Productivity Analysis for helpful suggestions.

†Corresponding author

the public administration or in the private sector. When comparing only two activities or entities, it is relatively easy to pin-point the exact environmental factor which causes the inferior performance. When comparing a bundle of activities or entire organizations, it is more intricate to identify the precise factors causing the poor efficiency. It is thus easier for managers to argue that, due to environmental factors, their organization is 'different' from the other entities in the data set. Although these concerns are a drawback highlighted by the benchmarking literature, frequently a benchmarking initiative is the only incentive to trigger efficiency and innovation in a natural monopolistic sector as competition *in* or *for* the market is impractical or not desirable.

In this article we concentrate on the measurement of efficiency (i.e. to which extent resources are converted into products) by the use of deterministic frontier models. Methodologies such as the non-parametric Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH) are popular among scholars and practitioners. However, despite the fact that hundreds of papers apply these methodologies, only a few of them try to take into account the operational environment, most of them simply neglect the exogenous influences. Nevertheless, exogenous environmental factors could influence the efficiency scores to a large extent. Indeed, favorable environmental variables behave as a substitutive input, while unfavorable factors absorb inputs to compensate the disadvantageous conditions.

The literature suggests various methodologies to integrate the operational environment (i.e., the heterogeneity) into the DEA analysis, however, none of them is consensual. Following Fried *et al.* (1999), the methodologies can be divided into several groups. Firstly, the frontier separation approach divides the observations in groups according to the environmental characteristics. Secondly, the all-in-one approach takes the exogenous variables immediately as an additional input (if a favorable effect) or as an additional output (if an unfavorable effect). Thirdly, the frequently employed two-stage model employs a truncated Tobit regression to estimate the direction of the influence of the environmental effect. However, Simar and Wilson (2007) show some serious doubts with respect to the Tobit estimates. Fourthly, the multi-stage models take into account slacks. Each of these techniques has its own advantages and drawbacks which causes (small) changes in the results. However, an important common disadvantage of these approaches is the necessity to specifiy *a priori* the exogenous influences. In addition, none of these methodologies is successful in fully capturing the operational environment. Evaluated entities could easily invoke other (also implicit) exogenous factors in addition to the *a priori* specified environmental variables. In many empirical evaluations, the efficiency of the DMUs could be largely influenced by 'untouchable' variables (e.g. corporate atmosphere) which are difficult to capture in numerical data, or by influences which are on its own of a minor importance (e.g. weather conditions) but in interaction with other variables have a significant impact on the observation's efficiency (e.g. social and cultural aspects).

Reinterpreting and extending the work of Battese and Rao (2002), Battese *et al.* (2004)

and O'Donnell *et al.* (2007), we develop a methodology which fully tries to capture the heterogeneity by the use of a 'Metafrontier'-framework. The latter concept, firstly, evaluates each observation relatively to the own group best practice frontier (where the units of the group are assumed to have the same environmental characteristics) and, secondly, to the overall-metafrontier constituting from the best practices of the different groups. The comparison of these two efficiency scores delivers the 'Group Specific Technical Efficiency' (GTE) which measures the overall efficiency of an entity while fully incorporating the explicit and implicit environmental characteristics. Therefore, the developed metafrontier framework is relatively easy to handle for practitioners, well explainable to stakeholders and straightforward to adapt to different models and situations.

We adapt the metafrontier framework to the robust order-m efficiencies as developed by Cazals *et al.* (2002). In the second section of the paper, we show by simulation that the robust efficiency scores could solve the current problem in the metafrontier literature of different group sizes. Indeed, as Zhang and Bartels (1998) point out, when comparing the average efficiencies of samples with different size, the results will be biased.

Finally, we apply the theory to an international data set of 122 utilities from 5 countries. Corrected for the heterogeneity, we find that the benchmarked Dutch utilities, the English and Welsh utilities regulated by yardstick competition and the regulated Australian utilities are performing better than the Belgian and Portuguese drinking water companies.

The research is organised as follows. In Section 2, we discuss the measurement of efficiency by the use of deterministic and robust frontier models. Section 3 develops the metafrontier concept to incorporate the operational environment. In Section 4, the model is applied to the drinking water sector. Section 5 concludes the article.

2 Frontier models

2.1 Deterministic frontier models

Decision Making Units (DMUs) transform multiple inputs into heterogenous outputs with a varying success. Conditional on the technology, each of the n firms absorb p inputs x to create q outputs y. The set of all these firms, called technology set, is characterized by

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}. \tag{1}$$

One way to estimate the relative efficiency of a DMU is to assume that the best performing units constitute a frontier which represents the best in class technology. The boundary of the technology set Ψ is represented by

$$\delta\Psi = \{(x, y) \in \Psi \mid (\theta x, y) \notin \Psi, \forall 0 < \theta < 1, (x, \lambda y) \notin \Psi, \forall \lambda > 1\}. \tag{2}$$

Viz-a-viz the firms along the production frontier, DMUs which are using relatively more resources to produce the given outputs in an input-oriented model, or DMUs which are producing relatively less outputs given their used inputs in an output-oriented model are considered as relatively inefficient. These firms are operating in the interior of the frontier Ψ. For a DMU located at $(x, y) \in \mathbb{R}_+^{p+q}$ one can measure the input and output efficiency as, respectively,

$$\theta(x, y) = \inf\{\theta \mid (\theta x, y) \in \Psi\}, \tag{3}$$

$$\lambda(x, y) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\}. \tag{4}$$

A procedure to measure the relative inefficiency scores θ and λ is to apply non-parametric techniques such as Data Envelopment Analysis (DEA) or parametric techniques as Stochastic Frontier Analysis (SFA). In this article, we concentrate on a less restrictive non-parametric estimator than DEA, i.e. the Free Disposal Hull (FDH) estimator of Deprins *et al.* (1984). Both DEA and FDH estimate the technology set Ψ by the smallest set $\hat{\Psi}$ that envelops the observed data. But whereas DEA uses constant (CRS) or variable returns to scale (VRS), respectively, the convex cone or convex hull of the FDH estimator, FDH relies only on a free disposability assumption (i.e. if $(x, y) \in \Psi$, then for any (x', y') such that $x' \geq x$ and $y' \leq y$, $(x', y') \in \Psi$). The DEA convexity assumption is not always valid as there may be returns to scale or non-divisible inputs and outputs. The FDH estimator is given by

$$\hat{\Psi}_{FDH} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq y_i, x \geq x_i, i = 1, \ldots, n\}. \tag{5}$$

The efficiency scores $\hat{\theta}_{FDH}$ can be measured relatively to this frontier via

$$\hat{\theta}_{FDH}(x, y) = \min_{i \mid y \leq y_i} \max_{j=1,\ldots,p} \left(\frac{x_i^j}{x^j} \right). \tag{6}$$

Alternatively, the input-oriented FDH efficiency scores can be obtained by solving the mixed integer linear programming problem

$$\hat{\theta}_{FDH}(x, y) = \min\{\theta \mid y \leq \sum_{i=1}^{n} \gamma_i y_i; \theta x \geq \sum_{i=1}^{n} \gamma_i x_i \ ,$$

$$\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \in \{0, 1\}; i = 1, \ldots, n\}. \tag{7}$$

2.2 Robust frontier models

Another, less deterministic, procedure are the robust order-m efficiencies as suggested by Cazals *et al.* (2002). This non-parametric estimator of the technology set Ψ is related to the FDH estimator but instead of constructing a full frontier as FDH does, it creates a partial frontier which envelops only m (≥ 1) data points. The order-m efficiency score is defined as

$$\theta_m = E\left[\min_{i=1,\ldots,m} \left\{ \max_{j=1,\ldots,p} \left(\frac{x_i^j}{x^j} \right) \right\} \mid y_i \geq y \right] \tag{8}$$

where the p-dimensional random variables $x_1, \ldots, x_m$ are drawn randomly and repeatedly from the conditional distribution of X given $y_i \geq y$. The estimator is based on the 'expected minimum input function' which allows to compare the efficiency of an observation with that of m potential units that have a production larger or equal to y. As it does not include all observations, it is less sensitive to outliers, extreme values or noise in the data. Although the estimator converges to the FDH estimate when m increases, this is only an asymptotic result.

Correcting for sample size by order-m efficiencies

As Zhang and Bartels (1998) indicate, a dissimilarity in sample sizes makes the comparison of average efficiency scores impossible. To solve this problem, we propose to use the order-m efficiencies of Cazals *et al.* (2002) in the spirit of the bootstrap ideas of Zhang and Bartels (1998). The latter suggest to focus on the group with the largest data set ($n_1 > n_2$) and after drawing n_2 random firms without replacement from the largest group's sample, to carry out an efficiency measurement (e.g. by FDH) to find the efficiency for these n_2 firms. To make a valid comparison between the mean efficiencies of group 1 and 2, this has to be replicated a number of times (bootstrapping). Indeed, the idea of the partial frontier in the order-m approach, as introduced above, is similar to this bootstrapping procedures. In contrast to Zhang and Bartels (1998), we set m equal to the smallest data set as this fits better in the metafrontier framework (see infra).

To show the merits of this approach, we simulate a data set and compute for different sample sizes the input-oriented FDH and the order-m efficiencies. We create 100 observations which use one input to produce two outputs. Both input and outputs are created from random numbers between 0 and 1: $x = rand_1 * 15$, $y_1 = rand_2 * rand_3 * 10$, $y_2 = 6.5 * (rand_1 * rand_2)/rand_3$. There is no significant correlation between the three variables. The estimation results, as presented in Table 1, reveal the decrease in average FDH-efficiency. The decline in terms of percentage relatively to the average FDH-efficiency estimates of a sample of 10 observations is drawn in Figure 1. As a striking contrast, the average order-m efficiencies (with $m = 10$ and $B = 100$) initially decline, but from a sample size of 50 on, they remain more or less constant on values above one which shows the decreased influence of outlying observations. Remark that both the ranks of the observations and the values of the efficiency scores are highly correlated for different values of m (a correlation of 0.95). The main advantage of a lower trimming value m is the reduced sensibility to outlying observations in the sample.

2.3 Taking into account the operational environment

The ability of efficiently transforming the resources into products does not only depend on the technical efficiency of the DMUs but also on the operational environment that characterizes

Table 1: Difference between average FDH efficiencies and order-m efficiencies

m = 10	FDH Average efficiency	order-m (B=100) Average efficiency	Average st. deviation
n = 10	0.8523	0.9499	0.0179
n = 20	0.8146	0.8886	0.0131
n = 30	0.6969	0.8668	0.0294
n = 40	0.6592	0.8479	0.0353
n = 50	0.6897	1.0538	0.0531
n = 60	0.6323	1.0358	0.0573
n = 70	0.6409	1.0255	0.0515
n = 80	0.6353	1.0871	0.0518
n = 90	0.6300	1.0286	0.0465
n = 100	0.6337	1.0609	0.0473

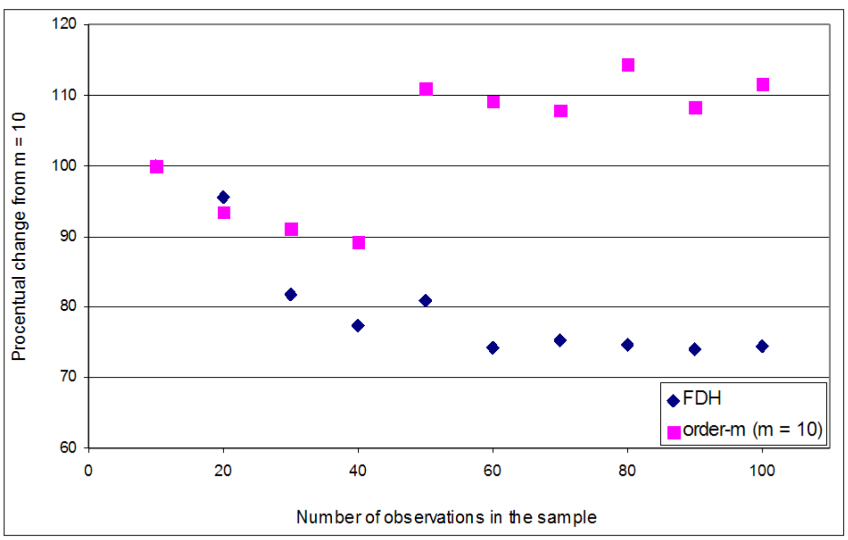

Figure 1: Difference between FDH and order-m

them. Therefore, the results of the benchmarking frontier models have a limited value if the operational environment where the DMUs perform is not taken into account (see Daraio and Simar, 2007; Fried *et al.*, 1999].

For example, for the water utilities it is argued that the ownership has an impact on the DMU efficiency. Other factors are the consumers' density, the percentage of non-domestic volume required or the water quality at its source. Hence, if the environmental variables are not adequately taken into account, some DMUs can be considered as efficient when they are inefficient or vice-versa, which is misleading. The operational environment, defined here, comprises all the explanatory factors that interfere, to a larger or lesser extent, with the DMU performance. There is not a precise definition of explanatory factors in the literature. They can be contextual or not. The environment (annual rainfall and topography) are context factors but the level of outsourcing is not, although the managers can only influence the latter factor in the long-term. However, these explanatory factors should not be confounded with non-discretionary (or non-controllable) inputs or outputs that are part of the productive process. To illustrate this in a water utilities context, the input mains length, representative of the input capital, is not controllable, at least in the short run, by the DMUs, so it cannot be classified as discretionary (controllable), even though it remains an input. From a different perspective, the peak factor is clearly an exogenous factor, in spite of having a great importance in the productive efficiency.

The classification of the variables into inputs, outputs and explanatory factors is somewhat complex, but fundamental in non-parametric studies. Lovell (2003) suggests that all the non-discretionary variables (non-controllable) should be considered in a second stage. This issue is discussed later. Unhappily in the water sector few studies try to encompass the influence of the explanatory factors in efficiency. Byrnes (1985), Woodbury and Dollery (2004), Resende and Tupper (2004) and De Witte and Marques (2007) are some of the exceptions.

However, in the comparison of entities, it is intricate and sometimes (almost) impossible to identify and measure the environmental factors which create the heterogeneity among the entities. In a similar benchmarking process, observations can easily claim that they are totally different and do not fit in the benchmark. Therefore, in the next section, we develop a framework which tries to fully compare the entities of several groups without the *a priori* selection and determination of exogenous characteristics.

3 A Metafrontier Approach

3.1 Metafrontiers

By reinterpreting and extending the ideas of Battese and Rao (2002), Battese *et al.* (2004) O'Donnell *et al.* (2007), we develop an attempt to fully (with explicit and implicit determined environmental variables) take into account the operational environment by the use of what they call ‘Metafrontiers’. Although not explicitly referred to, the idea of metafrontiers is used in other studies as well, see e.g. Morita (2003) and Portela and Thanassoulis (2001). In this section we analyze and enrich the literature on metafrontiers.

In a world with K groups, each having their specific state of technology and environmental factors, a metafrontier is defined as the boundary of the unrestricted technology set. Hence, the metafrontier envelops each of the separate group frontiers. For each of the K groups, the production process is constrained by the state of technology which transforms for each of the n_k observations in group k the p inputs x_k into q outputs y_k (for $k = 1, \ldots, K$). The group specific input efficiency is measured relatively to the n_k DMUs in the group sample such that

$$\theta^k(x_k, y_k) = \inf\{\theta^k \mid (\theta^k x_k, y_k) \in \Psi^k\} \tag{9}$$

where the technology set Ψ^k for group k is defined as

$$\Psi^k = \{(x_k, y_k) \in \mathbb{R}_+^{p+q} \mid x_k \text{ can produce } y_k\}. \tag{10}$$

If technology is freely interchangeable and thus if the k different groups have potential access to the same technology, we can apply the previously explained group frontier analysis relatively to the metafrontier. By pooling the observations of the K subgroups, the DMUs are evaluated with respect to the same standards. In this sense, the metafrontier represents an over-arching metatechnology where the technology set is defined by

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\} \tag{11}$$

where x and y denote, respectively, the input and output vector of an observation of any of the k groups. The (pooled) sample size is the main difference between the meta-concept and the group-concepts.

We present a graphical analysis of metafrontiers for the input-oriented FDH framework (an extension to the output-oriented framework is trivial). A simple example in the case of a single input and output technology is illustrated in Figure 2. Z represents two coinciding DMUs, Z_1 and Z_3, respectively of group 1 and group 3. As group frontier 1 envelops group frontier 3 in this particular interval, the best performing utilities in group 1 are able to produce the same amount of outputs with less inputs than the best performing observations of group 3. Therefore, utilities of group 1 are part of the metafrontier. This superiority of group 1 on the other groups could be attributed to a more advanced technology or to more favorable environmental factors, out of control of the firm’s managers. Measured relatively to the group frontier, an inefficient observation Z_3 performs relatively more efficiently than an inefficient observation Z_1:

$$\text{Efficiency of } Z_1 \text{ relatively to its group frontier} = TE_1^k = \frac{OA}{OZ} = 0.3 \tag{12}$$

$$\text{Efficiency of } Z_3 \text{ relatively to its group frontier} = TE_3^k = \frac{OB}{OZ} = 0.6 \tag{13}$$

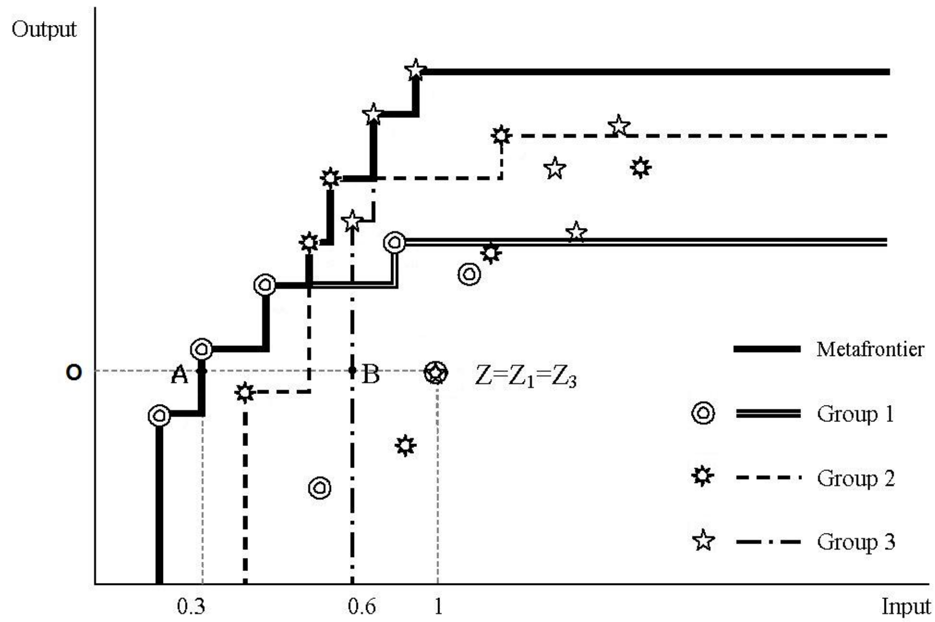

Figure 2: Graphical analysis of metafrontiers

This implies that using the technology available in group 1, the output vector of Z can be produced by using only 30 per cent of the input vector, while using the technology of group 3, only 60 per cent of the inputs would be necessary in comparison to the best practices. The efficiency measured relatively to the metafrontier, TE^*, is the same for both DMUs (i.e. TE^* equals 0.30). As an interesting exercise, we could compare the meta and the group technology. A first concept to measure this gap is the *technology gap ratio* (TGR) introduced by Battese and Rao (2002). It reveals, in terms of percentage (1-TGR), the gap between the maximum input reduction possible under the group technology and the metatechnology. Hence, it shows the potential payoff from copying the best practice technology.

$$TGR_1 = \frac{TE_1^*}{TE_1^k} = \frac{\frac{OA}{OZ}}{\frac{OA}{OZ}} = \frac{0.3}{0.3} = 1.0 \qquad TGR_3 = \frac{TE_3^*}{TE_3^k} = \frac{\frac{OB}{OZ}}{\frac{OA}{OZ}} = \frac{0.3}{0.6} = 0.5 \tag{14}$$

3.2 Group Specific Technical Efficiency

The TGR only indicates the potential input reduction of copying the best practice (meta) technology. In this subsection, we introduce a composite statistic which is complementary to the TGR and which tries to rank the observations while fully taking into account the operational environment. Relying on two assumptions, the developed indicator uses the efficiency estimates relatively both to the metafrontier and to the group frontier. Firstly, we assume, and test in the next subsection, that DMUs of the same group face homogenous environmental factors. This is a reasonable assumption as the assignment of observations to a particular group is mostly clear-cut and thus the operational environment within a group will be well comparable (e.g. utilities from the same region, sector or company). Secondly, we assume that, although the DMUs of different groups could be exposed to different environmental factors, the pooled observations perform relatively to the metafrontier as efficient as they are doing relatively to their group frontier. Hence, on the one hand, DMUs which are a benchmark in the group sample will not additionally be penalized in their efficiency relatively to the metafrontier. On the other hand, DMUs that are already inefficient in the homogeneous group sample, will be additionally penalized in their inefficiency in the meta sample. The newly obtained variable, which we label *group specific technical efficiency* (GTE^*), computes the degree of inefficiency if the group inefficiency were unchanged and if the utilities faced an equalized technology. It is computed relatively to the metafrontier as

$$GTE_i^* = TE_i^* \times TE_i^k. \tag{15}$$

If we compare in our graphical example DMUs of group 1 with observations of group 3, the former could have both a superior technology and more favorable environmental factors. Since favorable environmental factors behave as substitutive inputs, DMUs of group 1 need relatively less resources to produce the same proportion of outputs, which yields higher efficiency scores. Utilities of group 3 in comparison to utilities of group 1, besides having an inferior technology, could face more unfavorable environmental factors. As these behave as substitutive outputs, they absorb resources, causing lower efficiency estimates. Following our first assumption, we assume that regarding to the group frontier all DMUs of group k are working in the same environment and with the same technology. Hence, observation Z_1 is relatively to the other observations in its group very inefficient (0.3), which contrasts to Z_3 that is relatively to the other DMUs in the group only 'somewhat' inefficient (0.6). Relative to the metafrontier, all DMUs in group 3 have a low efficiency score due to the unfavorable environmental factors.

On the one hand, the group frontier analysis can not be used to detect differences in efficiency between the groups since each group has its own benchmarks. On the other hand, although a metafrontier analysis is able to detect the group differences, it pools all the observations but does not take into account differences in environmental factors. The GTE^* fills this gap by computing efficiency estimates relatively to the pooled sample and by taking into account the operational environment. Assuming in the graphical example that Z_1 and Z_3 transform their resources with the same extent of inefficiency in the meta approach as in the group approach, the GTE^* ranks Z_3 higher than Z_1:

$$GTE_1^* = TE_1^* \times TE_1^k = 0.3 \times 0.3 = 0.09 \qquad GTE_3^* = TE_3^* \times TE_3^k = 0.3 \times 0.6 = 0.18$$

By the use of the graphical example in Figure 2, we can derive a geometric interpretation of the GTE^*:

$$GTE_{Z_3}^* = \frac{OB}{OZ} \cdot \frac{OA}{OZ} = \frac{OZ - BZ}{OZ} \cdot \frac{OZ - AZ}{OZ}$$

$$= 1 - \frac{BZ}{OZ} - \frac{AZ}{OZ} + \frac{AZ.BZ}{OZ^2}. \tag{16}$$

We could interpret the GTE^* as we would give every DMU the benefit of the doubt (i.e. efficiency score of 1), but penalize for its inefficiency relatively to the group frontier ($\frac{BZ}{OZ}$) and to the metafrontier ($\frac{AZ}{OZ}$) while correcting for the interaction between the group and the meta-efficiency ($\frac{AZ.BZ}{OZ^2}$).

Up to now, the metafrontier literature has been using both parametric and non-parametric estimators, in particular DEA and SFA, to estimate the relative efficiency of an observation. However, this can easily be extended to other models as well. In this paper, we concentrate on a specific non-parametric estimation technique to estimate the group efficiencies since the use of DEA to compare several groups with different group sizes could create biased results (see supra). By our best knowledge, this particular issue is still neglected in the metafrontier literature. Many authors are using DEA to compare group frontiers with a different number of observations. Hence, it is impossible to compare the obtained results correctly as the relative efficiency score of the DMUs depends on the number of observations in the group sample. Also FDH faces this drawback. By the use of order-m frontiers, we propose an alternative approach.

4 Metafrontiers to the drinking water sector

4.1 Data and model specification

Due to sunk capital costs in the mains and in production plants, the drinking water sector can be labeled as a natural monopolistic network industry. The fixed and largely sunk costs for water distribution represent up to 70 per cent of the total drinking water price (compare to the 40 per cent in the energy sector). Transportation difficulties further reinforce the natural monopoly as a transport over 100 kilometers increases the wholesale cost of water by about 50 percent (compared to 5% for electricity and 2.5% for gas). Another inhibitor of competition in the water sector is the percentage of consumption of the different consumers, which is more penalizing for the water sector than in other network industries. Industrial consumers, for example, guarantee 50% of the revenue of the electricity sector where only 13% of the water sector. Besides, the water sector depends mainly on domestic consumers (75%), compared with the 30% for the electricity sector. Among others, these facts show that a liberalization of the water sector would unlikely result in the same benefits as in other network industries. However, as a natural monopoly creates X-inefficiencies, quiet life and excess profits, governments should establish regulatory bodies to promote efficiency (doing the things right) and effectiveness (doing the right things) in the drinking water sector. This section attempts to perform a benchmark study by comparing 5 countries by the metafrontier approach.

We consider in our research the water utilities from England and Wales, Australia, the

Netherlands, Belgium and Portugal. In 1989 England and Wales water utilities were privatised, which was the major landmark that has occurred in water sector worldwide until now. This bold reform makes the State watchdog role more important and visible and, maybe for this reason, the best value of the water sector has been its regulation which, by means of a price cap regulatory method and by the use and abuse of benchmarking, has formed a virtual market in the water sector in these countries. Water services are vertically integrated and some of them provide wastewater along with water. In the Netherlands the water services are provided by public limited companies separately from other activities and are vertically integrated. The main features of the sector are, besides its corporatization, the merging that took place in the last decades reducing the two hundred operators that existed in the 1960s to the current ten and the self-regulation of the sector by the Association of Dutch Water Utilities (Vewin). This particular kind of 'regulator' has improved the performance of the sector, avoiding its privatisation by the application of a voluntary balanced scorecard benchmarking scheme among its members. Water utilities in Australia are similar to the Dutch in their governance. They are almost always public and are also corporatized, working under commercial principles. However, in Australia there is strict regulation. The competency of the water services belong to the States /Territories and a Federal Law obliges all monopolies, irrespective of the ownership, to be regulated by independent regulatory authorities. Water utilities often provide other services like wastewater, gas or waste. In Belgium the water sector institutional framework changes with the region. Unlike the Walloon region, where regulation recently exists, the water services provided by the municipalities or their associations are absent of regulation. There are a large number of players, although some of them with large size. Most of the water utilities work under a non-commercial principle. The rule is the separation of the activities (water and wastewater) and sometimes the production from the distribution. There are no incentive schemes in the Belgian water sector and only recently has the benchmarking tool started to be applied in the Walloon region. In Portugal the responsibility for the water services belongs to the municipalities. There are about 300 players where approximately 10 % are private concessions. The private water utilities are currently regulated by the Institute for the Regulation of Water and Waste (IRAR) with quality service supervision functions very relevant. The remaining utilities work in a deregulated environment and the majority in a non-commercial way. Water services are usually provided together with other activities like wastewater and they went through the unbundling process some years ago with separated entities for the production and the distribution.

The data set is obtained from the sector organizations Office of Water Services (OFWAT, England and Wales), Association of Dutch Water Utilities (VEWIN, Netherlands), Water Services Association of Australia (WSAA, Australia), Belgaqua (Belgium) and Portuguese Water Association (APDA, Portugal). All data originate from 2005, except for Belgium (2004). The model specification is the most important step in any efficiency analysis as it heavily influences the results. We opted, following the literature, for a simple model inspired

Table 2: Homogeneity assumption of countries

	leakage (%)	industrial water (%)	groundwater (%)	consumption/capita	regional product
Intercept	0.309 (***)	0.214 (***)	30.370 (***)	134.229	15687.440 (***)
Netherlands	-0.269 (***)	0.161 (***)	63.880 (***)	16.668	8784.003 (***)
England and Wales	-0.126 (***)	0.049 (*)	18.915 (*)	1308.377 (**)	5948.606 (***)
Australia	-0.190 (***)	0.073 (**)	35.339 (***)	229.670	10037.420 (***)
Belgium	-0.051 (**)	-0.046 (**)	33.492 (***)	13.221	5216.158 (***)
R^2	0.596	0.237	0.232	0.059	0.623
F-statistic	37.576 (***)	7.937 (***)	7.690 (***)	1.602	42.144 (***)

Note: n=107; *** denotes significance at 1% level, ** at 5% and * at 10%

on a production function. Drinking water utilities need capital, proxied by the length of mains, and labor, estimated by the number of personnel in full time equivalents, to produce the outputs which are the delivery of drinking water to a number of connections.

Among the different countries, environmental factors such as the geographic features, relative wealth of the consumers, quality of the ground and surface water or age of the infrastructure have a significant influence on the relative efficiency of the drinking water utilities. By the use of OLS, we test whether utilities located in the same country face more or less the same environmental factors. This corresponds to the first assumption of the previous section in which we argue that utilities of the same group face homogenous environmental factors. The results relative to the Portuguese utilities are presented in Table 2. It reveals that environmental variables are mainly country specific as for each of the environmental variables (except for consumption per capita) we can attribute a significant part of the variation in the variable to country dummies. Remark further that we are able to explain to a large extent (up to 62%) the variation in the environmental variables by only including country dummies. This strengthens our assumptions that utilities measured relatively to the group frontier are working in more or less the same environment. This makes sense as the differences in social, physical and institutional environment are minor within one country or region.

4.2 Group and metafrontiers

As argued before, the order-m approach is especially useful in the comparison of average efficiency scores of groups with different sample sizes. In the spirit of Zhang and Bartels (1998) we set the trimming value m equal to 12, the smallest group size (i.e. for the Netherlands). While assuming similar environmental factors for utilities working in the same country, we present the average efficiencies and the standard deviation of the estimates in Table 3. As the maximum efficiency score reveals, some countries have super-efficient observations. For example, consider the Dutch utility with an efficiency score of 1.251 indicating that with a proportionate reduction of the inputs, this DMU uses 25% less inputs than the expected minimum input level of 12 other Dutch firms drawn from the population and producing more than this DMU's output.

By merging the 5 group samples, we measure the meta-efficiency of the utilities. Again we employ the input-oriented order-m approach as it constructs the partial frontier only with observations with an equal or larger output vector. Sample size issues are avoided by keeping the trimming value m equal to 12, such that the metafrontier will be similar to the group frontier (however, remark that other observations will constitute the partial frontier and that in the strict sense the dissimilarity in sample size is not an issue in the measurement of efficiency relatively to the metafrontier). We equilibrate the increased number of observations in the data set (i.e. 104 instead of e.g. 22) by increasing fivefold the number of Monte-Carlo replications used in computing the order-m estimates (i.e. $B = 500$). As expected, the results in Table 3 show a lower average efficiency in all countries, which can intuitively be explained by the increased possibility of facing a more efficient observation in an enlarged data set. It should be noticed that the order-m efficiency scores of the metafrontier are not necessarily smaller than order-m efficiency scores of the group frontier due to the constraint $Y \geq y_o$.

The comparison of the efficiency measures relatively to the group and metafrontier shows up in the TGR which is computed for every DMU. The country averages, shown in Table 4, reveal that on average the DMUs are working on 82% of the best-practice technology available in one of the 5 countries. Especially the English and Welsh utilities are working closely to the metafrontier, while the difference between the group technology and the metatechnology is with 25.5% the largest for Portugal. It is interesting to note that in all countries, except for Portugal, the group frontiers are tangent to the metafrontier (as the TGR is equal or larger to one in each of these four regions) and thus some observations of these countries constitute the metafrontier.

In the complementary GTE^* exercise, we firstly assume that utilities relatively to the group frontier face similar environmental factors, and secondly, that relatively to the metafrontier, while using a different technology, the utilities work as efficient as they are doing relatively to the group frontiers. These two assumptions result in the GTE^*-ratio and allow us to rank the utilities while fully taking into account the environmental variables. Frequency distributions for the TGR and the GTE^* are presented in Figure 3. As shown in Table 4 the English and Welsh drinking water companies are performing most efficiently, followed by the Dutch and Australian utilities. In contrast to Portugal and Belgium these three countries apply regulatory schemes to stimulate the drinking water sector to produce more efficiently, so it could be expected that regulatory incentives effectively increase the efficiency of the sector. However, it is worthwhile to further investigate this issue. In addition, it is important to highlight that the inferior performances showed by the Portuguese and Belgian utilities are not controlled, or at least only partially controlled, by the utilities themselves and therefore their managers are not responsible for the scores presented. On the other hand, the results depict the usefulness of the metafrontier concept, since it constitutes a good base to take

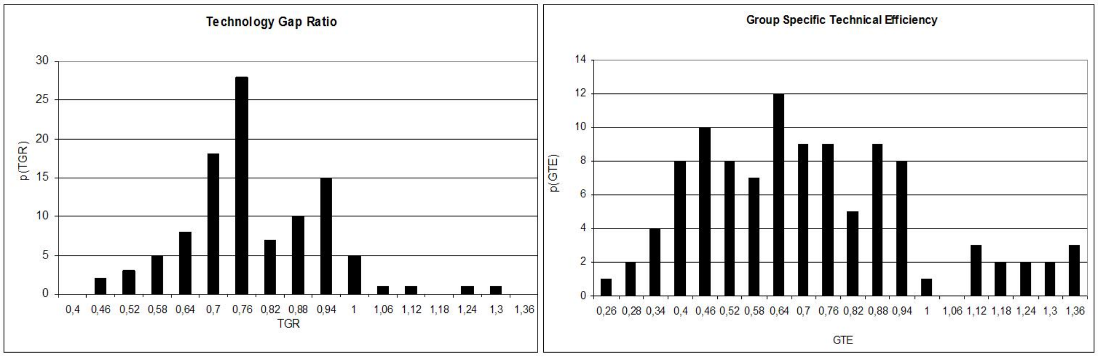

Figure 3: Frequency distributions of TGR and GTE

Table 3: Efficiency relatively to the group and metafrontier

	Relatively to the group frontier				to the metafrontier	
	Average efficiency	Standard deviation	Maximum	Minimum	Average efficiency	Standard deviation
Netherlands	0.980	0.098	1.251	0.864	0.811	0.144
England and Wales	0.943	0.151	1.234	0.657	0.908	0.200
Australia	0.948	0.078	1.090	0.772	0.806	0.135
Belgium	0.931	0.071	1.000	0.797	0.736	0.281
Portugal	0.922	0.172	1.400	0.526	0.686	0.152
Portugal - public	0.990	0.166	1.400	0.713	0.714	0.167
Portugal - private	0.934	0.064	0.973	0.526	0.630	0.098

decisions by the autorities. For example, in this case, the Portuguese and Belgian politicians can look closely at England and Wales, the Netherlands and Australia and mimic or find help in their regulatory framework to take their own decisions.

5 Conclusion

This paper reinterpreted and extended the metafrontier-concept to a framework to account for exogenous environmental characteristics. By comparing the group frontier and the overall

Table 4: TGR and GTE

	Technology Gap Ratio				Group Specific Technical Efficiency			
	Average efficiency	Standard deviation	Maximum	Minimum	Average efficiency	Standard deviation	Maximum	Minimum
Netherlands	0.827	0.114	1.011	0.623	0.804	0.220	1.393	0.492
England and Wales	0.957	0.111	1.275	0.716	0.882	0.319	1.619	0.332
Australia	0.847	0.103	1.015	0.693	0.772	0.176	0.983	0.478
Belgium	0.778	0.251	1.339	0.481	0.700	0.308	1.339	0.305
Portugal	0.745	0.088	0.956	0.461	0.653	0.258	1.370	0.211
Portugal - public	0.736	0.104	0.956	0.461	0.714	0.282	1.370	0.365
Portugal - private	0.763	0.042	0.856	0.698	0.531	0.147	0.682	0.211

metafrontier formed by the best practices of the several groups, the environement corrected efficiency of entities can be computed and explained easily. In addition, the approach is applicable to several models, including the robust order-m approach for which, by a simulated example, we show that it disregards the sample size bias. To judge on the relative efficiency of the DMUs while fully considering the environmental factors, the 'Group Specific Technical Efficiency' measure is proposed and applied to drinking water utilities from 5 different countries. This research indicated that some of the delay detected in Portugal and Belgium is derived from a technological gap relatively to other countries, mainly England and Wales. On the one hand, these results justify the poor performance in this sector in these countries that is not only responsibility of the managers and, on the other hand, they clearly indicate the way they have to follow with regard to the water sector governance. It allows for the conclusion that the use of incentive schemes like regulation and benchmarking is convenient.

References

[1] Battese, G., Rao, D., 2002. Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business* 1 (2), 87-93.

[2] Battese, G., Rao, D., O'Donnell, C., 2004. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis* 21 (1), 91-103.

[3] Byrnes, P., 1985. Ownership and efficiency in the water supply industry: An application of the nonparametric programming approach to efficiency measurement. PhD Dissertation. Carbondale (USA): Southern Illinois University, 1-325.

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

[5] Daraio, C., Simar, L., 2007. Advanced robust and nonparametric methods in efficiency analysis. Methodology and applications, Series: Studies in Productivity and Efficiency, Springer.

[6] De Witte, K., Marques, R., 2007. Designing incentives in local public utilities, an international comparison of the drinking water sector. *CES Discussion Paper Series DPS* 07.32., 1-32.

[7] Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labor efficiency in post offices. In: Marchand, M., Pestieau, P., Tulkens, H. (eds.), The Performance of Public Enterprises: Concepts and Measurements. Amsterdam, North-Holland. pp. 243-267.

[8] Fried, H., Schmidt, S., Yaisawarng, S., 1999. Incorporating the operating environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis* 12 (3), 249-267.

[9] Lovell, C.A.K., 2003. The decomposition of malmquist productivity indexes. *Journal of Productivity Analysis* 20 (3), 437-458.

[10] Morita, H., 2003. Analysis of economies of scope by data envelopment analysis: comparison of efficient frontiers. *International Transactions in Operational Research* 10 (4), 393-402.

[11] O'Donnell, C., Rao, D., Battese, G., 2007. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics* (to appear), 1-35.

[12] Portela, S., Thanassoulis, E., 2001. Decomposing school and school-type efficiency. *European Journal of Operational Research* 132 (2), 357-373.

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

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

[15] Woodbury, K., Dollery, B., 2004. Efficiency measurement in Australian local government: the case of New South Wales municipal water services. *Review of Policy Research* 21 (5), 615-636.

[16] 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), 187-204.