TI 2007-092/3
Tinbergen Institute Discussion Paper

Mean and Bold?

Kristof De Witte[1]
Elbert Dijkgraaf[2]

[1] *University of Leuven;*
[2] *Erasmus University Rotterdam, and Tinbergen Institute.*

Mean and bold? On separating merger economies from structural efficiency gains in 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

Elbert Dijkgraaf
SEOR-ECRi
Erasmus University Rotterdam
P.O. Box 1738, H 7-25
3000 DR Rotterdam, The Netherlands
dijkgraaf@few.eur.nl

November 27, 2007

Abstract

The Dutch drinking water sector experienced two drastic changes over the last 10 years. Firstly, in 1997, the sector association started with a voluntary benchmarking aimed to increase the efficiency and effectiveness of the sector. Secondly, merger activity arose. This paper develops a tailored nonparametric model to dissect and distinguish the effects on efficiency of these two evolutions. In particular, we adapt Free Disposal Hull (FDH) to estimate robust and conditional non-oriented efficiency estimates. Parametric COLS (Fourier) tests show the robustness of the model with respect to the specification and its variables. We classify the merger economies into scale economies and increased incentives to fight inefficiencies. Although we detect a significant efficiency enhancing effect of benchmarking, we find insignificant merger economies due to the absence of scale economies and the absence of increased incentives to fight inefficiencies.

Keywords: Mergers and acquisitions, efficiency, scale economies, water sector, non-parametric and parametric estimation

JEL Classification: C13, C14, D20, G34, L95

1 Introduction

In 2003, the Dutch parliament decided to reserve the Dutch drinking water sector as a public domain, implying a moratorium on private investments. This decision was justified

---

*We would like to thank Jan van Helden for discussions on the research questions. In addition, we want to thank Rui Marques and other seminar participants at Technical University of Lisbon.

by the performance of the sector in terms of efficiency and effectiveness. However, not even a few years before, in 1997, the Dutch Ministry of Economic Affairs launched a program to deregulate monopolistic markets, among which the drinking water sector. The sector, represented by the association of Dutch drinking water utilities Vewin, opted for a voluntary benchmarking to circumvent competition in or for the market. Benchmarking is aimed at seeking excellence through a systematic comparison of performance measures with reference standards. Two different applications of benchmarking can be distinguished (for an overview, see De Witte and Marques, 2007a). Firstly, *yardstick competition* which uses the results of benchmarking to set maximum prices or revenues (Schleiffer, 1985). This approach is applied in e.g. the privatized English and Welsh drinking water sector. Secondly, *sunshine regulation* which compares and publicizes the benchmarking results. The latter approach is applied in the Dutch drinking water sector (as is common practice in the Netherlands, we will identify benchmarking with sunshine regulation henceforth).

In the same time period, intensified merger activity arose. The first merger wave in the sector was about eighty years ago when public companies, owned by the municipalities, took over the municipal services in their search for a minimal scale necessary as a result of new environmental requirements.[1] The second merger wave among the public drinking water utilities, started in 1996. For this wave, several diverging reasons were present, (1) it was believed that operating at a larger scale increases the efficiency (scale economies), (2) the larger scale would enable the owners for more specialized and thus improved supervision (corporate governance), (3) the larger scale would be needed to comply with environmental regulation and (4) national and provincial authorities promoted, for the easiness of a coherent policy, the existence of one drinking water utility per province[2]. While there were hundreds of utilities in 1920, 111 in 1975 and 20 in 1992, only 10 drinking water companies remained in 2007. Several of them are discussing new mergers, possibly resulting in still less utilities in the next few years.

The literature is indecisive whether lean firms are mean or whether bold firms are more beautiful. Much depends on the merger economies. We argue that three specific variants of merger economies could be present (e.g. Roller *et al.*, 2000). Firstly, although drinking water utilities operate in a legal regional monopoly, mergers could increase market power by decreasing the effectiveness of the benchmarking instrument as the number of reference partners declines. Secondly, by operating at a larger scale, merged firms could benefit from scale economies and lower the production costs in comparison to pre-merged firms. Finally, the merger could provide the management an enlarged mandate to fight inefficiencies, especially when cost reductions were promised. Although mergers might be effective to realize the latter as well, it is not a necessary condition. Therefore, managers and regulators should

---

[1] This paper studies both mergers and take-overs, however, for the ease of explanation we use the word 'merger' for both terms.

[2] There are 12 provinces in the Netherlands.

carefully evaluate the merits and demerits of all available instruments to execute the latter. As these three merger economies are partly contradicting, it is crucial to understand whether mergers result in scale economies or in improved inefficiency reductions since mergers are only necessary to realize the former. In the absence of scale effects and with other instruments available to fight inefficiencies, the market power argument advocates a reluctant attitude towards mergers.

This paper makes several contributions to the literature. Firstly, the paper develops a tailored nonparametric model which is inspired on *Free Disposal Hull* (Deprins *et al.*, 1984). We adapt the FDH model to a non-orientation which simultaneously measures multiple input reductions and heterogeneous output expansions without assuming any *a priori* specification on the functional form of the production set. The non-orientation is particularly convenient in the current setting as the firm's managers are simultaneously considering cost reductions (i.e. the inputs) and service ameliorations (i.e. the outputs). Secondly, by adapting the robust order-m efficiencies of Cazals *et al.* (2002), the model accounts for atypical observations. Thus, it neutralizes the dependency on one or a few extra-ordinary observations, the main drawback of deterministic nonparametric techniques Thirdly, the nonparametric model enables us to to discriminate the effects on efficiency of different institutional changes which are happening at the same time. By modifying the procedures of Daraio and Simar (2005, 2007), the model allows for a richer analysis and results in improved information on the effectiveness of the evaluated instruments. Fourthly, the paper estimates merger effects in a benchmarking environment, an issue not explored before. Fifthly, it improves the understanding of merger effects in the drinking water sector, since currently, only evidence is available based on UK-data. Sixthly, the results of the nonparametric model are tested and compared by parametric variants. In particular, we assume a parametric *Corrected Ordinarily Least Squares* (COLS) model to test the robustness of the basic model, and apply a *Fourier* function to examine the robustness of an extended model. The paper is the first in applying a Fourier parametric cost function, the most flexible parametric approach available, to the water sector. Finally, the paper explicitly discriminates between efficiency effects due to scale economies and effects as a result of increased incentives to fight inefficiency. This is important as the first mechanism necessitates mergers, while the second does not.

The paper proceeds as follows. Firstly, we describe the benchmarking project in Dutch drinking water sector and review the literature on merger economies (Section 2). Secondly, we develop a tailored nonparametric model to disentangle the benchmarking efficiency gains from the total merger economies (Section 3). Our results indicate in the period before the benchmarking and shortly afterwards (1993-1999) a negative evolution of efficiency, while in the period 1999-2005 a positive effect of the benchmark is detected (Section 4). Total merger economies seem to be positive in the two years before the merger till one year after the merger. However, from then on, we observe merger diseconomies. We test the robustness of the results by converting the nonparametric model into a parametric COLS specification and find that the results are in general robust (Section 5). We further test the robustness of the model with respect to the included variables and to the assumed specification (Section 6). In addition, this model allows to break down the overall merger economies into scale economies and economies resulting from increased incentives to fight inefficiencies. We find that both scale economies and incentives to fight inefficiencies are absent. Finally, we provide conclusions (Section 7).

2 Benchmarking and mergers in the Dutch drinking water sector

The focus of this paper is on the efficiency effects of mergers. In this respect, the Dutch drinking water sector is interesting as large merger activity is present during recent years. In this paragraph, we explore the possible causes of mergers, its potential impact on efficiency and argue the importance to know whether merger economies follow from scale effects or improved incentives to fight inefficiency. At more or less the same time, the introduction of benchmarking probably influenced efficiency also. Thanks to the enforced incentive regulation and to the pressure of all stakeholders the Dutch drinking water sector currently performs better in terms of efficiency and effectiveness than the sectors in e.g. Australia, Belgium, Portugal or even England and Wales (De Witte and Marques, 2007b). Therefore, analyzing merger economies necessitates to include the role of benchmarking. We start this section by motivating the Dutch rationale at that time for choosing for benchmarking, discuss its specific characteristics and argue the effectiveness of the tool. Next, we discuss the pros and cons of mergers.

2.1 Benchmarking

The Dutch drinking water utilities provide drinking water to domestic and industrial costumers. Efficiency concerns arose at the Dutch drinking water sector agenda in 1997. At that time, the Ministry of Economic Affairs managed a general programme to deregulate markets with monopolistic power. Among other sectors, drinking water was under review for deregulation possibilities as utilities were owned by municipal and provincial governments and operated in a legal monopoly. Due to the potential presence of the *quiet life* and *X-inefficiencies* in the monopolistic framework, efficiency might be at stake. This hypothesis was further supported by the fact that companies were managed by technocrats who preferred an increase in drinking water quality and security of water supply, even with unsure benefits and prohibitive costs.

After studying the possibilities to introduce competition on the market (not applied in other countries), for the market (the French model) or yardstick competition with price regulation (the English and Welsh model), the Dutch parliament decided to introduce benchmarking as a light handed type of incentive regulation (Dijkgraaf *et al.*, 1997). It was expected to enhance efficiency and to maintain or even further improve the high levels of quality, security of supply, investments and the (nearly) absence of leakage. In contrast to other incentive mechanisms, benchmarking does not necessarily require a strict regulation of quality, security and investments, and does not change the institutional form of the sector radically. Therefore, a striking advantage of benchmarking is its absence of an expensive regulation authority as it leaves all decisions on outputs and service targets to the utility (including the definition and monitoring of minimum requirements).[3]

Basically, the benchmark is an information generating instrument. Before 1997, management and owners of drinking water utilities lacked instruments to assess their efficiency. The benchmark generates yearly exhaustive information on costs, quality and service levels which are compared among the utilities by the use of performance indicators.[4] Besides the provision of information at the company level, the analysis is also performed at process level (e.g. production, distribution, sales, support and general management processes) and even at subprocess level (e.g. the cost per meter pipe or the cost per installed water meter). The voluntary benchmark is organized by the sector organization Vewin, the association of Dutch water companies, which contracted an external consulting firm to manage the benchmarking process. Every three years, an external report is published, while each year companies receive a detailed internal report.

From a theoretical point of view, it might be expected that benchmarking is effective. As benchmarking introduces the same mechanisms as competition on the market, this light handed type of incentive regulation might even be as effective as the competitive model if the proper rewards and penalties are available. Similarly with competition on the market, where management and owners make decisions which are expected to result in better performances than competing firms, benchmarking triggers the race to the top.[5]

In practice, benchmarking in the sense of the Dutch model can only be effective in an adequate '*naming and shaming*' framework which depends on internal carrots and external sticks. Internal carrots arise from the use of benchmarking as a source of information to stimulate managers to improve performances. Indeed, since 1997, owners explicitly negotiate contracts with the management about efficiency improvements in relation to financial rewards. Owners have incentives to supply such carrots as they are penalised externally if the company does not improve efficiency. External penalties originate from public publications (e.g. in newspapers or sector magazines) and the public debate initiated by these publica-

---

[3] The Ministry of Environmental Affairs sets and monitors bottom-line minimum requirements for quality.

[4] Note that the first official benchmark report was published in 1999. However, between 1997 and 1999 companies could use information from other reports (e.g., Dijkgraaf *et al.*, 1997) and preliminary results from the benchmark analysis.

[5] Kwoka (2006) makes a similar statement discussing competitive possibilities in the U.S. electricity sector. He states that (p. 128) "How close benchmarking approximates the cost discipline of direct competition is an empirical question". His empirical analysis shows that both options arrive at comparable cost savings.

tions. The owners are the municipalities who are represented by the municipal council. As they are elected by the citizens living in the supply area of the utility, they are sensitive to this public debate.

Recent evaluations of the role of benchmarking proof the effectiveness of the incentive in the Dutch drinking water sector. According to Dijkgraaf et al. (2007) and Vewin (2007) efficiency improved by 23%, while also quality, service and investments have been improved. This exceeds all expectations as earlier study (Dijkgraaf *et al.*, 1997) estimated a potential increase in efficiency of 15%. Indeed, as the benchmarking project goes on, the best performing utilities also improve further their performances. Currently, potential efficiency gains are estimated at about 20% (Dijkgraaf *et al.*, 2007).

2.2 Mergers

The relationship between efficiency and mergers has been extensively studied in the literature. Roller *et al.* (2000) point to three consequences of mergers in their literature review. Firstly, mergers might result in increased market power which could be exploited by higher prices and profits. Secondly, mergers could create positive scale economies which enables them to lower production costs in comparison to the pre-merger firms. Thirdly, merged firms might have more incentives to fight inefficiencies. While the two latter mechanisms are positive from a welfare perspective, the former is not. In this paragraph, we briefly discuss the three mechanisms.

Firstly, although even pre-merger drinking water utilities work in a monopolistic environment (i.e. by law only one company has the right to supply drinking water in a particular region), effective market power can expand due to mergers as incentive regulation becomes less practical. For an effective and efficient benchmarking, a minimum number of firms is required as the comparison of different entities only generates relevant information if data are well comparable. A lower number of firms potentially decreases the power of benchmarking as companies are heterogeneous (major differences exist in e.g. network intensity, the type of raw water used and the output mix). The one remaining company is efficient by default after the last merger.[6]

Secondly, scale economies arise when mergers allow to exploit cost advantages by operating at a larger scale. The literature provides ample evidence on the water sector to distrust the scale economies argument. Although scale economies are found for very small companies (with a scale far below the smallest company in the Netherlands), constant returns to scale or even scale diseconomies are generally found for larger companies (OECD, 2004). The

---

[6] A minimum number of firms is not only necessary for benchmarking in the sense of sunshine regulation, also in the sense of yardstick competition the analysis becomes intricate when the number of firms decreases. For price or revenue cap regulation, it is essential for a regulator to know the true cost level, which is only attainable with comparable data. For this reason, the UK Monopoly and Mergers Commission is very reluctant to mergers in the drinking water sector in England and Wales (OECD, 2004).

literature provides evidence for this conclusion using data for drinking water companies in Germany, France, Italy, Portugal, the United States and Japan.[7]

Thirdly, increased and accelerated incentives to fight inefficiencies might arise when the management obtains an enlarged mandate or incentive after the merger to fight inefficiency. Particularly, when managers convinced the owners to agree with the merger on the basis of cost arguments and when positive scale economies are not present, management has to realize its promise by reducing inefficiency. The literature provides only spare evidence on the size of this effect. For the drinking water sector, only Ballance *et al.* (2004) analyze efficiency in relation to mergers, using data for England and Wales. They find no difference in efficiency between post-merger and no-merger firms. However, this seems to be in line with studies from other network sectors. Sung and Gort (2006), for instance, find no positive welfare effects in the American telecom sector due to the absence of scale effects, productivity increases and efficiency gains, while shareholders' value remains stable. Kwoka and Pollitt (2007) find negative effects of acquisitions in the American electricity sector as efficiency of the purchaser remained stable while efficiency of the bought firm decreased. However, Ivaldi and McCullough (2005) arrive at a significant welfare increase in the railroad sector.

Overall, the literature is suspicious regarding the effects of mergers. Gugler *et al.* (2003), for example, studying $114,000$ mergers world wide, conclude that 70% of mergers does not increase welfare. They conclude that 50% of all mergers results in less profit, while another 20% was only able to increase profits by setting higher prices thanks to more market power.

To conclude, an important difference between the arguments of scale economies and increased incentives to fight inefficiencies is that mergers are only required to realize scale economies. Although mergers might be effective to increase incentives to fight inefficiencies, this is not a necessary condition and managers should evaluate the pros and cons of all available instruments to realize this. In the absence of scale effects and with other instruments available to fight inefficiencies, the market power argument advocates a reluctant attitude towards mergers. Given the trade-off between efficiency and market power and the lack of evidence for efficiency effects of mergers, it is worthwhile to evaluate in the next sections the merger economies.

3 A nonparametric model

To analyze merger economies, both parametric and nonparametric techniques can be employed. The former assumes a particular model to the data and tries to fit the data according to this model. The major advantages of this approach lie in the sound statistical properties that can be derived from the model and the easiness to include variables correcting for the exogenous environment. In contrast, the second methodological group does not assume any

---

[7] See, respectively, Sauer (2005), Garcia and Thomas (2001), Fabbri and Fraquelli (2000), Martins and Fortunato (2006), Kim (1987) and Mizutani and Urakami (2001).

particular functional form, and thus '*lets the data speak for themselves*'. Therefore, nonparametric models can easily handle multiple inputs and heterogeneous outputs scenarios. Thanks to the recent work of, among others, Simar and Wilson (1998) and Daraio and Simar (2005, 2007) statistical properties of the estimators can be deduced for nonparametric estimators. In this section, we design a tailored nonparametric non-oriented model (Section 3.1) to estimate the efficiency of the Dutch drinking water companies. The traditional nonparametric models as developed by, among others, Charnes *et al.* (1978) and Deprins *et al.* (1984) assume that any deviation from the best-practice frontier is attributed to inefficient management. This is a rather strong assumption, however, as atypical and outlying observations could influence the frontier (Section 3.2). To dissect and distinguish the merger economies from the benchmarking results, we exploit the ideas of Daraio and Simar (2005, 2007) and adapt them to our model (Section 3.3).

3.1 Non-oriented efficiency estimation

The bulk of the nonparametric literature on the measurement of efficiency deals with estimating input-oriented (i.e. for a given output level minimization of the inputs) or output-oriented efficiency (i.e. for a given input level maximization of the produced outputs). Nevertheless, in many practical observations firm's managers design simultaneous input reducing and output increasing schemes to improve the performance of the entity. Some attempts for non-oriented models have been made by Charnes *et al.* (1985), Färe *et al.* (1985) or Portela *et al.* (2003).[8] However, none of these models formulated a non-oriented model for the traditional non-convex Free Disposal Hull (FDH) model, as developed by Deprins *et al.* (1984). By the use of the directional distance functions of Chambers *et al.* (1998), we develop a non-oriented FDH estimator. The FDH model is convenient as it requires only two minimal assumptions. Firstly, the production set Ψ is assumed to envelop all observed observations χ. Secondly, we require monotonicity, i.e. inputs and outputs are freely disposable. This means that more input (less output) never implies a decrease of the maximally achievable output (minimal required inputs).

Consider a sample χ with n utilities which use p inputs to produce q outputs and label the corresponding input and output vectors, respectively, as $x \in \mathbb{R}_+^p$ and $y \in \mathbb{R}_+^q$. The set of all feasible input - output combinations is defined as:

$$\Psi = \left\{ (x, y) | x \in \mathbb{R}_+^p ; y \in \mathbb{R}_+^q ; (x, y) \text{ is feasible} \right\}. \tag{1}$$

Only imposing free disposability on Ψ, i.e. if $(x, y) \in \Psi$ then $(x', y') \in \Psi$ for $x' \geq x$ and $y' \leq y$, and thus allowing for non-convex technologies, the FDH estimator measures the efficiency of an evaluated entity, denoted by (x_o, y_o), relative to the boundary of the Free Disposal Hull. The FDH estimator of Ψ is characterized by:

---

[8] See De Witte and Marques (2007c) for a survey.

$$\Psi_{FDH} = \left\{ (x,y) \in \mathbb{R}_+^{p+q} | x \geq X_i; y \leq Y_i; (X_i, Y_i) \in \chi \right\}. \tag{2}$$

In Figure 1 where we present a two-dimensional graph with one input variable on the horizontal axis and one output variable on the vertical axis, the frontier appears graphically as a step-wise function as it searches for every evaluated observation a reference entity which uses less inputs to produce more outputs. The horizontal distance to the frontier, the input-oriented efficiency, measures the minimal quantity of inputs required to produce a constant amount of outputs. This can be estimated by the following mixed integer linear program problem:

$$\theta_{input}(x_o, y_o) = \max \left\{ \begin{array}{c} \theta | (1-\theta) x_o \geq \sum\limits_{i=1}^{n} \gamma_i X_i; y_o \leq \sum\limits_{i=1}^{n} \gamma_i Y_i; \sum\limits_{i=1}^{n} \gamma_i = 1; \\ \gamma_i \in \{0,1\}; (X_i, Y_i) \in \chi \end{array} \right\}. \tag{3}$$

Clearly, $0 \leq \theta_{input} \leq 1$ and efficient observations obtain an efficiency score of $\theta_{input} = 0$. Alternatively to the linear program formulation, Tulkens (1993) pointed out that a practical two-step vector comparison algorithm can be used to compute θ_{input}. The first step reveals for each evaluated observation the reference partners, i.e. the set of observations dominating (x_o, y_o):

$$D_o(x_o, y_o) = \{ i | X_i \leq x_o; Y_i \geq y_o; (X_i, Y_i) \in \chi \}. \tag{4}$$

The second step computes the input-efficiency by

$$\theta_{input}(x_o, y_o) = \max_{i \in D_o} \left(1 - \frac{X_i}{x_o} \right). \tag{5}$$

Analogous, the vertically measured output-oriented efficiency estimator maximizes the outputs for a fixed amount of inputs. The mixed integer linear program problem corresponds to:

$$\lambda_{output}(x_o, y_o) = \max \left\{ \begin{array}{c} \lambda | x_o \geq \sum\limits_{i=1}^{n} \gamma_i X_i; (1+\lambda) y_o \leq \sum\limits_{i=1}^{n} \gamma_i Y_i; \sum\limits_{i=1}^{n} \gamma_i = 1; \\ \gamma_i \in \{0,1\}; (X_i, Y_i) \in \chi \end{array} \right\}, \tag{6}$$

such that $0 \leq \lambda_{output} \leq 1$ and efficient observations obtain an efficiency score of $\lambda_{output} = 0$. Within the set of dominating observations D_o the output-efficiency can alternatively be measured by

$$\lambda_{output}(x_o, y_o) = \max_{i \in D_o} \left(1 - \frac{y_o}{Y_i} \right). \tag{7}$$

Extending the traditional FDH models with ideas from directional distance functions, as developed by Chambers *et al.* (1998), allow us to estimate the more realistic non-oriented efficiency estimates. In particular, we assume the direction of the distance function as $(g_x, g_y) = (x, y)$, i.e. simultaneously reducing inputs and expanding outputs. This yields the following mixed integer linear program:

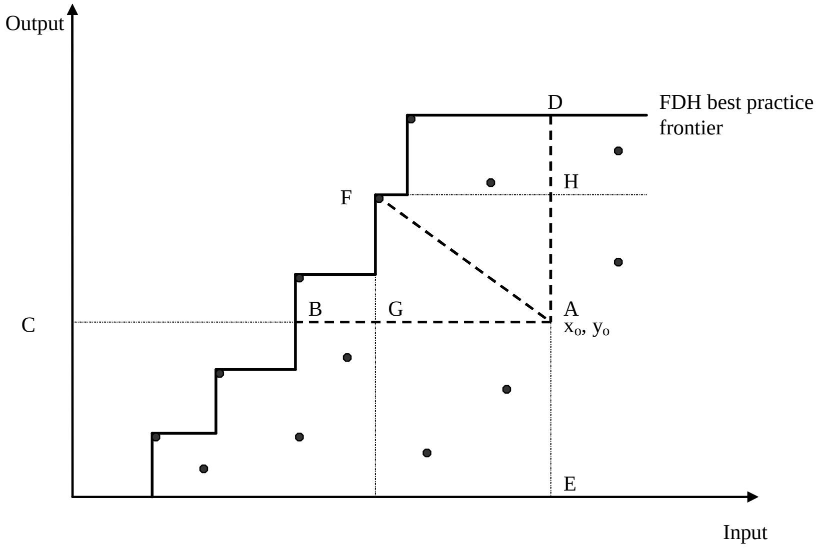

Figure 1: FDH best practice frontier

$$\pi_{non}(x_o, y_o) = \max \left\{ \begin{array}{c} \pi|(1-\pi)x_o \geq \sum_{i=1}^{n} \gamma_i X_i; (1+\pi)y_o \leq \sum_{i=1}^{n} \gamma_i Y_i; \sum_{i=1}^{n} \gamma_i = 1; \\ \gamma_i \in \{0,1\}; i = 1, \ldots, n \end{array} \right\}, \tag{8}$$

An efficient observation which constitutes the frontier obtains a value of $\pi_{non} = 0$, while inefficient observations arrive at positive values. Whereas the input-orientation approaches the frontier horizontally and the output-orientation approaches it vertically, the non-oriented estimator minimizes the distance to any non-dominated observation in the fourth quadrant relatively to observation (x_o, y_o) (see Figure 1). Extending the procedures of Tulkens (1993) and adapting the FDH directional distance function framework of Cherchye *et al.* (2001), we compute first the non-oriented reference observation for (x_o, y_o) for the set of undominated observations as:

$$(x_{i_D}, y_{i_D}) = \min_{i \in D_{undom}} \sqrt{1 - \frac{X_i}{x_o} + 1 - \frac{y_o}{Y_i}}. \tag{9}$$

Secondly, we give the benefit of the doubt to each observation by estimating the non-oriented efficiency as the minimal distance to the frontier in either the input or the output-oriented dimension relatively to this closest reference point (x_{i_D}, y_{i_D}) (with rescaled estimators):

$$\pi_{non} = \max \left(1 - \frac{X_{i_D}}{x_o}, 1 - \frac{y_o}{Y_{i_D}} \right). \tag{10}$$

3.2 Mitigating outlier influence

To mitigate the influence of outlying and atypical observations and to infer statistical properties of the estimators, we construct a partial frontier which does not include all observations. This is important as such observations heavily influence the efficiency socres in deterministic frontier models (as in the full FDH frontier model). In our application, outliers could arise from (1) measurement errors and (2) atypical observations with low or high values for particular variables. The partial frontier approach, introduced by Cazals *et al.* (2002), draws repeatedly ($r = 1, \ldots, R$) and with replacement for every evaluated observation (x_o, y_o) a subsample of size m from the set of dominating observations. For each of the R subsamples the efficiency score $\pi^r_{non}(x_o, y_o)$ is estimated along the mixed integer linear program in (8). Finally, the robust non-oriented 'order-m' estimator is computed as the arithmetic average: $\pi^m_{non}(x_o, y_o) = \frac{1}{R}\sum\limits_{r=1}^{R} \pi^r_{non}(x_o, y_o)$. The observations in the reference group change due to resampling as the evaluated observation (x_o, y_o) will not always be included in the subsample we allow for 'super-efficient' observations. These yield an efficiency score $\pi^m_{non} < 0$ and indicate that the observation is performing more efficiently than the average of m reference observations in the subsample.

We have to specify the parameters m (the number of observations in each drawing) and R (the number of drawings with replacement) to compute the robust efficiency measure. Daraio and Simar (2005, 2007) suggest to select these m and R which keeps the number of super-efficient observations unchanged. As in our application, the number of super-efficient observations is about the same for all values of m and R at less than 1% and as selecting a different value does not significantly changes the results, we arbitrarily set m equal to 60 and R equal to 100.[9]

3.3 Distinguishing incentives from merger economies

The efficiency in the Dutch drinking water sector experiences the influence of two movements. The benchmarking project triggers all drinking water utilities similarly since 1997, while mergers happening in different years influence only some of the companies. Disentangling the effect of mergers and benchmarking involves the use of panel data. Two popular techniques in the nonparametric literature to deal with panel data are Malmquist Indices and Window Analysis (see e.g. Cooper *et al.*, 2004). As the former tool is mainly used as a decomposition tool for efficiency and productivity change, we concentrate on the latter. To detect performance trends of an entity over time, a window analysis operates in a panel data sample of all observations on the principle of moving averages, such that, each observation could possibly be compared with its past or future values which are considered as different observations. We put the procedure into practice by modifying the environment-corrected ef-

---

[9] Detailed results for different values for m $(= 10, 20, \ldots, 80)$ and for R $(= 20, 40, \ldots, 160)$ are available upon request.

ficiency measures of Daraio and Simar (2005, 2007) and by focussing on its reference selection principle.

To introduce environmental variables Z in the efficiency scores of evaluated entities, Daraio and Simar (2005) suggest a probabilistic approach which conditions on the environmental variable of the evaluated entity, z_o. By selecting an appropriate bandwidth h and by applying a Kernel function $K(.)$, the procedure selects all observations (x, y) in the neighbourhood of z_o :

$$D_o(x_o, y_o|z_o) = \{i|X_i \leq x_o; Y_i \geq y_o; |z_o - Z_i| \leq h; (X_i, Y_i) \in \chi\}. \tag{11}$$

The order-m estimator is accommodated by drawing the m observations with a probability equal to $K((z_o - Z_i)/h)/\sum_{j=1}^{n} K((z - z_j)/h)$ such that the *robust non-oriented conditional efficiency measure* $\pi^m_{non}(x_o, y_o|z_o)$ is obtained. In contrast to the bandwidth selection procedure suggested by Daraio and Simar (2005, 2007), i.e. the cross-validation criterion, we select the optimal bandwidth by simple reasoning (and test the robustness with respect to this choice). Two environmental variables are created to distinguish the benchmark incentives from the merger economies.

Firstly, to capture the effects of the benchmarking project on efficiency, we construct a time trend ($TIME$). As the drinking water sector is characterized by the nearly absence of technological development, the effects of benchmarking in $TIME$ will prevail on other influences. We expect to find increased efficiency since 1997, the start of the voluntary benchmarking. Secondly, in addition to this general pattern, we await to find influence on the efficiency of merging companies from two years before the merger till five years after the merger date. To allow for the diverging impact in the years before and after merger, we assign the value $MERGE = 1$ corresponding to the year of the merger; value $MERGE = 0.9$ and 0.8 in, respectively, one and two years before the merger; 1.1 one year after the merger, and continued up to 1.5. Non-merging utilities or utilities without merging influence obtain a time value of $MERGE = 0$. We experienced with other, less flexible specifications as well (e.g. $MERGE = 1$ in all years after the merger and $MERGE = 0$ in all years before the merger) and found similar results.

The determination of the bandwidth for $TIME$, h_{time}, and hence the window size in the window analysis, deserves particular attention; although the literature does not make any suggestion on how to proceed. Indeed, if h_{time} is selected too small, the number of potential reference partners in the analysis is reduced which dramatically diminishes the discrimination in the results. Conversely, if h_{time} is too large, the analysis will not be able to detect the changes over time. In our model we determined h_{time} arbitrarily equal to 3, such that each utility is compared with utilities from two years before and two years after the evaluated time $z_{o,time}$. To test the robustness of the analysis, we experimented with other values of h_{time} as well ($h_{time} = 1, \ldots, 9$) and found similar outcomes. We selected the optimal bandwidth for $MERGE$ as $h_{merge} = 0.3$, which corresponds to the selection of observations from two years before up to two years after the evaluated observation $z_{o,merge}$. Again, other experiments ($h_{merge} = 0.1, \ldots, 0.5$) delivered similar results.

As an exploratory tool to verify the effect of environmental variables on efficiency, Daraio and Simar (2005, 2007) suggest to graphically compare the efficiency estimates $\pi^m_{non}(x_o, y_o)$ and $\pi^m_{non}(x_o, y_o|z_o)$ by nonparametrically regressing the ratio $\pi^m_{non}(x_o, y_o|z_o)$ / $\pi^m_{non}(x_o, y_o)$ against Z_i. An increasing regression line indicates in our non-oriented model a favorable effect of the conditioned environmental variable to efficiency (i.e. increasing efficiency for higher values of Z_i), while a decreasing regression denotes an unfavorable effect to efficiency (i.e. decreasing efficiency for higher values of Z_i). The absence of a graphical first order impact points to the absence of influence of the conditioned variable to efficiency. In the multivariate framework we nonparametrically regress the ratio of the partially conditioned efficiency scores (conditioned on only one environmental variable, say Z_1) to the fully conditioned efficiency scores (conditioned on both environmental variables, say Z_1 and Z_2) against the values of the conditioned variable (i.e. Z_2). As an efficient (unconditioned) observation obtains a score $\pi^m_{non}(x_o, y_o) = 0$, the ratio of conditioned to unconditioned variables would yield undefined values. Therefore, we add the value 1 to the results.

Initially, we estimate the full merger economies (i.e. the sum of scale efficiency and efficiency improvement *ceteris paribus* the scale), although we relax this assumption in the following parametric sections.

4 Nonparametric estimations

We apply our model to a data set deduced from the public annual accounts of the Dutch drinking water companies. The data range from 1992 till 2006. Whereas the Dutch drinking water sector counted 20 companies in 1992, 10 utilities remained in 2006. To artificially increase the number of observations and to allow for the nonparametric measurement of the effects of benchmarking and merger economies, we carefully decompose each of the available variables of the 10 merged companies in 2006 to the initial 20 companies in 1992 (called henceforth sub-utilities) by relatively to its size extrapolating the growth of the variables in the merged company to the sub-utility. This proceeds in two steps. Firstly, as we do have values for the sub-utilities in the year before the merger ($T-1$), we construct for the merged company each of the variables in $T-1$. Secondly, proportionally to the share of the sub-utility in the merged company, we extrapolate the growth of the merged company between $T-1$ and T to the sub-utility. By doing this for every year, we effectively decompose the data set and capture for each sub-utility the claimed advantages of a merger. As such, we obtain a sample of 293 observations.[10] The Appendix provides an overview of the variable

---

[10] Note that we have only incomplete data for one company.

definitions, the descriptive statistics and the mergers in the Dutch drinking water sector.[11] The robustness of the results with respect to the decomposition of the data set is tested in Section 6.

While we experienced with several input variables (various combinations of total costs, the wage base, the capital base, the number of connections, the balance sheet value, length of the mains network as a proxy for capital value and number of employees as a proxy for labour), the results of these models were very similar to each other. In this article, we report only the results of real total costs as input variable. Indeed, real total costs allow for a fair comparison among companies and years as it captures investments in capital and labour, maintenance expenditures for the infrastructure, outsourcing of employees, customer-related services, etc. As presented in Figure 2, average total costs of all companies is increasing from 1992 till 2000, while decreasing between 2003 and 2006. As output variables, we adopted the total volume of drinking water (in million m^3) and the number of connections per m^3. These output variables, which are consensual in the literature, capture respectively the scale of the utility and the number of customers. Both output variables evolve more or less similar to the input variable.

To disentangle the efficiency changes attributed to the benchmark and efficiency effect due to mergers, we apply our tailored robust non-oriented conditional efficiency measure. We present for both conditional and unconditional estimates the averaged results for each year in Figure 3. Notice that the efficiency scores are very high, i.e. on average the scores are close to 0. This is a consequence of the specific model we apply, that is, a (1) robust (2) non-oriented (3) conditional (4) FDH model. Each of these four components give in the measurement of efficiency the 'benefit of the doubt' to the observations. Taken together, they significantly improve the *absolute* efficiency level of the results. As in the determination of efficiency we compare the utilities relatively to each other (cf. relatively to the best practice frontier), the *relative* efficiency scores of the utilities do not change due to our assumptions. Also in the second approach, where we visualize the effects on efficiency of the variables *TIME* and *MERGE*, the (absolute) magnitude of the curvature will be less pronounced in comparison to models which do not give the full benefit of the doubt to the evaluated entities.

Consider Figure 3 where the average conditional and unconditional efficiency scores are presented for each year. We first consider the unconditional efficiency scores in which all utilities are compared with all utilities in the sample. During the first five years of the analysis, the unconditional efficiency scores remain more or less stable at an inefficiency of

---

[11] The costs are made comparable as much as possible. Total costs are defined as gross-income minus environmental taxes (mainly the tax on using groundwater). Large differences in profit policies are corrected by diminishing total profits and adding a calculated 'normal' return of 6% on equity. Large differences in appreciation, as a possible source of outliers (see supra), are corrected by diminishing appreciation on current value (most companies value using historical costs). In addition, corrections are made for incidental appreciations, penalties paid for fast pay back, incidental costs or income, and other non-regular bookkeeping measures.

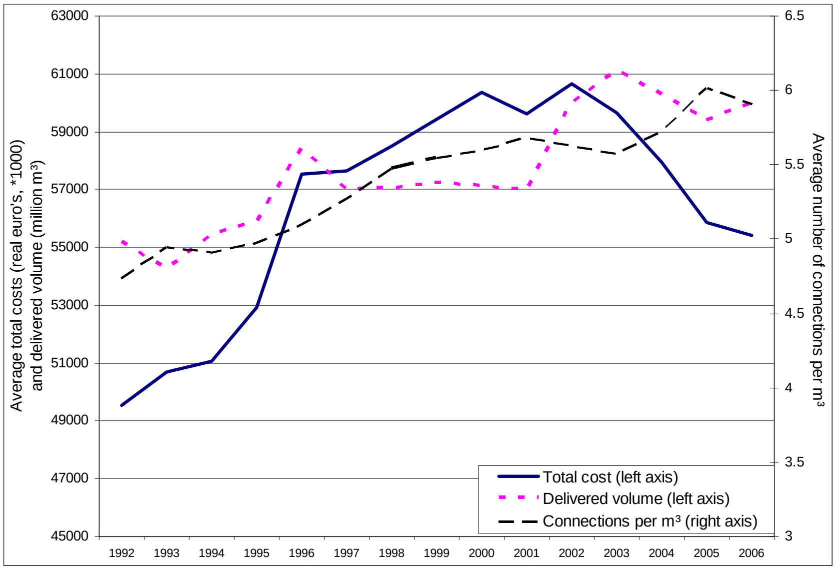

Figure 2: Evolution of descriptive statistics

2.4%. This contrasts to the average efficiency scores from 1998 (i.e. the first year after the sunshine regulation) and 1999 when an initial sharp increase in the efficiency is followed in 2000 by a downswing in efficiency, followed by a steady increase in performances from 2001 on.

A similar pattern can be detected in the efficiency scores conditioned on *MERGE* as only utilities in the same stage of the merger are mutually compared, the claimed benefits of the merger are counterbalanced. We observe a steady efficiency level between 1992-1997, followed by a sharp increase in efficiency around 1998-1999 and a dramatic decrease in efficiency in 2000 after which efficiency gradually improved. In addition, as the average efficiency scores in each year are higher for conditioned values than for unconditioned, an unfavorable effect of the conditioned value *MERGE* could be expected.

Regarding the conditioning on the variable *TIME*, interesting results arise. These efficiency scores capture the efficiency of companies without the effects of the *TIME* variable. The results reveal that without benchmarking efficiency would have been dramatically lower over the period 1992-2006. In addition, it further strengthens the expectation that the increase in efficiency in the Dutch drinking water sector could in particular be attributed to the benchmark while the merger economies do not significantly improve efficiency.

As a complementary explanatory nonparametric approach to detect whether the environmental variables *TIME* and *MERGE* are favorable or unfavorable to efficiency, we follow the procedure described by Daraio and Simar (2005, 2007). The results are presented in Figure 4 and 5. First consider the influence of *TIME*. In the period 1993-1999, we observe a negative

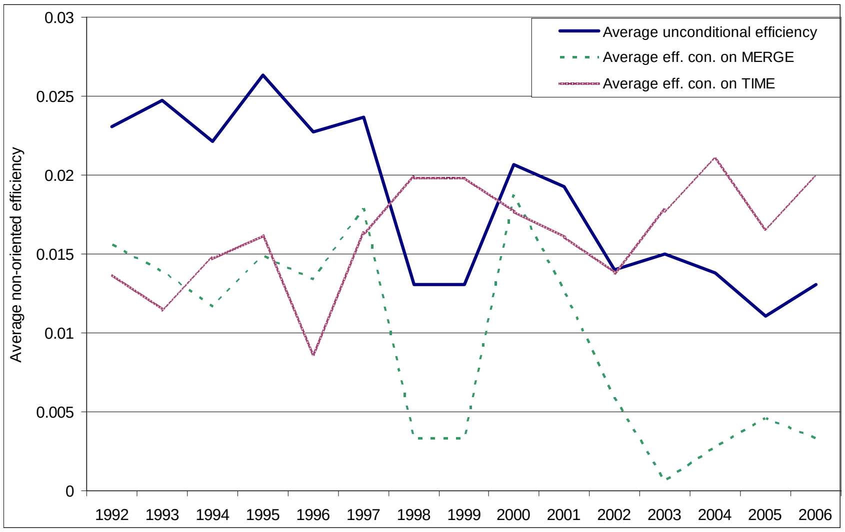

Figure 3: Conditional and unconditional efficiency estimates

first order impact of *TIME* on efficiency. In the period before the benchmarking and shortly afterwards the variable *TIME* has an unfavorable effect on efficiency. In contrast we find a positive first order effect in the period 1999-2005. This allows us to infer the positive role of the benchmark on the efficiency of the Dutch drinking water companies.

Consider next the impact of *MERGE*. In the two years before the merger ($T-2$) until one year after the merger ($T+1$), we detect a positive first order effect. Over this period of time, merging carries a favorable influence over these utilities. From $T+1$ to $T+5$ we find a negative first order impact of *MERGE*. Merged firms experience a unfavorable influence to efficiency in this period. We test the robustness of these results in the next sections.

5 A robustness test

This section tests the robustness of the results with respect to the estimation technique (parametric versus nonparametric) using the same disaggregated data set as before. This contrasts to the next section, where we test parametrically the robustness of the selected variables and the applied disaggregation of the data set.

To relate the costs of drinking water utilities to its inputs and outputs, a cost function is estimated. The unexplained costs (the residuals) are considered as an approximation of the inefficiency. Estimating the cost function in logs, the efficiency is measured by the use of the residual of company i in year t, denoted by $\mu_{i,t}$. In particular, the efficiency equals $e^{\nu_{i,t}}-1$, where $\nu_{i,t}$ equals $\mu_{i,t}$ - $\min(\mu)$. This corresponds to a Corrected Ordinary Least

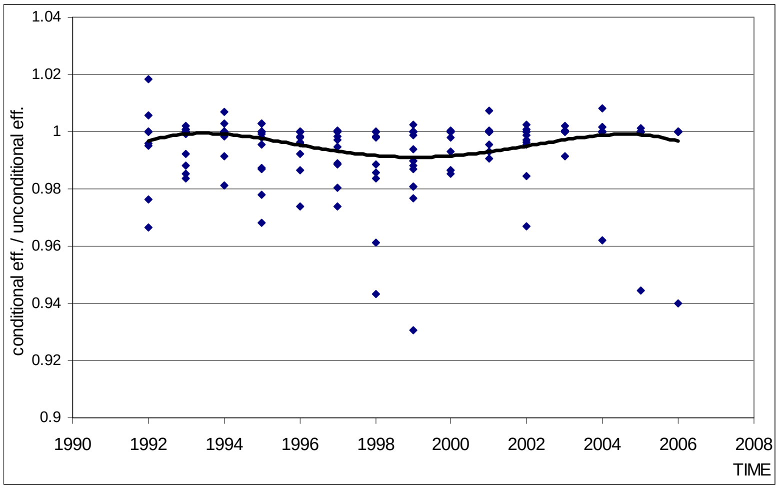

Figure 4: Impact of TIME on efficiency

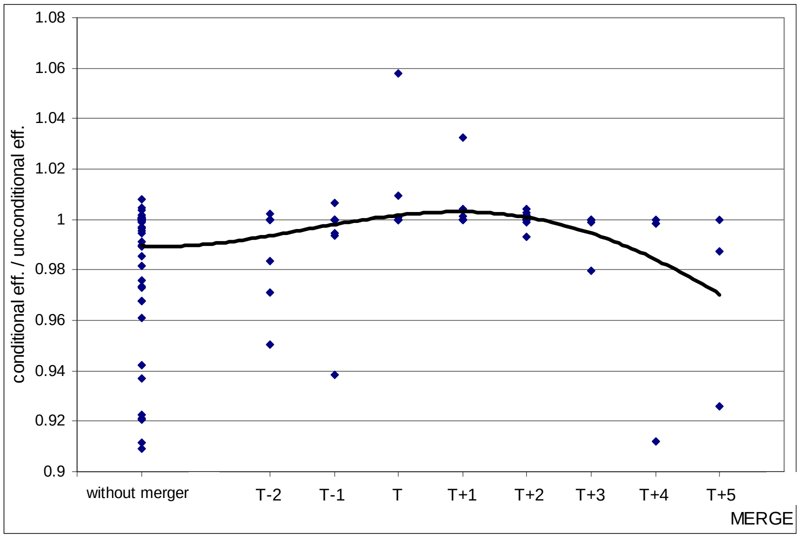

Figure 5: Impact of MERGE on efficiency

Squares (COLS) methodology where the average cost function is pushed inwards towards the most efficient observation.[12] Similar with the nonparametric approach, efficiency is supposed to be a function of benchmarking and merger activity. Again, the merger economies are approximated by the variable *MERGE*, and the benchmarking by *TIME* and its square to allow for non-linear effects. To mimic a parametric variant of the nonparametric model, we opted for a Cobb-Douglas (CD) specification as it reflects a similar relationship between costs and outputs. Compared to the nonparametric specification, the main difference in the CD model is the introduction of the input prices which results in a cost minimizing function with input and output prices on the right-hand side. The model is summarized as:

$$C_{i,t} = \alpha\, Q_{i,t} + \beta\, P_{L_{i,t}} + \gamma\, P_{C_{i,t}} + \zeta\, Con_{i,t} + \theta_1\, TIME + \theta_2\, TIME^2 + \eta\, MERGE_i + \mu_{i,t} \tag{12}$$

where $C_{i,t}$ denotes the total real cost for company i in year t, $Q_{i,t}$ the produced quantity of drinking water, $P_{L_{i,t}}$ the real input price of labour, $P_{C_{i,t}}$ the real input price of capital, $Con_{i,t}$ the number of connections per m^3 and $\mu_{i,t}$ the error term after subtracting the average effect of *TIME* and *MERGE*. All variables are measured in logs as the underlying production function is multiplicative.

The results are presented in Table 1. The coefficient for Q implies decreasing returns to scale as an average increase in production of 1% results in a 1.09% increase in costs. Although its coefficient is close to 1, the constant returns to scale hypothesis (i.e. $Q = 1$) is rejected at the 5%-significance level (but not at 1%). As could be expected, costs are significantly increasing with the input prices. A higher number of connections per m^3 results significantly in a cost increase, which is natural since for a constant level of production more connections have to be served. Similar to the nonparametric estimations, the trend variable *TIME* suggests the effectiveness of the benchmarking tool as a significant bell-shaped relationship between *TIME* and efficiency is detected. While efficiency decreases between 1992 and 1997, it stabilizes in 1998 and 1999 and increases from 2000 on. According to the model, in 2006 the efficiency level increased with 16% compared to 1997. As the benchmark project was able to turn the decreasing efficiency trend, the introduction of the benchmark resulted probably in even higher efficiency gains than if it would not have been established. Finally, we find significant negative merger economies since mergers decrease the efficiency by 7%. In contrast to the nonparametric specification, where only negative efficiency results were detected after the first year of the merger, the parametric specification does not find a bell-shaped effect of mergers (presented in model B of Table 1). Besides the latter remark, the analysis demonstrates the robustness of the nonparametric results with respect to the functional form. In the next section, we proof the robustness of the results

---

[12] See Greene (1980) for technicalities. The drawback of COLS is that the entire residual is supposed to represent inefficiency. Results for Stochastic Frontier Analysis (SFA) (Coelli and Perelman, 1999), which makes it possible to split the residual in an inefficiency term and noise, are highly comparable. Results are available upon request.

Table 1: Parametric robustness test

	Model A	Model B
	Linear effect of MERGE	Quadratic effect of MERGE
Constant	-4.34**	-4.37**
	(0.43)	-0.43
Q	1.09**	1.09**
	(0.02)	(0.02)
P_L	0.44**	0.44**
	(0.10)	(0.10)
P_C	0.39**	0.39**
	(0.02)	(0.02)
Con	0.13**	0.13**
	(0.04)	(0.04)
TIME	0.04**	0.04**
	(0.01)	(0.01)
TIME^2	-0.003**	-0.003**
	(0.001)	(0.001)
MERGE	0.07**	-0.01
	(0.02)	(0.09)
MERGE^2		0.08
		(0.08)
R^2	0.96	0.96

Standard errors between brackets.

** denotes significance at 1%.

with respect to the employed variables and data set.

6 The origins of merger economies

In this section, we take advantage of a particular characteristic of parametric models relative to their nonparametric counterpart. Parametric analysis makes it possible to correct for a larger set of exogenous differences between companies in comparison to the nonparametric models (in their given state of technology). This might be important for two reasons. Firstly, if characteristics differ between merged and non-merged companies, the estimated merger coefficient could be biased if the model does not take the diverging characteristics into account. Assume, for instance, that utilities using groundwater, which only have low purification costs, do not merge, while utilities using river water, which have much higher purification costs, do merge. If not accounted in the model, the merger coefficient will additionally pickup the cost difference between using groundwater and river water suggesting an efficiency decrease after merger. Secondly, it enables us to break down the overall merger effect into scale economies and economies resulting from more incentives to fight inefficiency for a given scale. As the former effect is only attainable by mergers, while the latter effect is also achievable without merger (see supra), it is important to know which merger effect prevails.

Although we use the same data sources of the previous sections (i.e. based on annual reports and sector publications), the disaggregation level differs to allow for separating the merger economies into scale effects and inefficiency reducing abilities. Obviously, the former requires that variables are measured at the true scale as breaking down merged firms into sub-utilities underestimates scale effects in a parametric framework.

Although parametric cost functions generally assume a Cobb-Douglas (CD) or Translog specification[13], we apply the Fourier specification as this functional form has some flexible characteristics. The Fourier cost function adds *sine* (sin) and *cosine* (cos) terms to the Translog model, which adds in turn quadratic terms to the CD specification. The Fourier cost function does not only allow for linear relationships as in CD and non-linear causalities as in Translog, but for almost infinitely flexible relationships.[14] As such, the Fourier specification is a generalization of these models. While CD and Translog specifications only exploit the dominating trend in the data, and provide as such only a local approximation for the unknown function, the Fourier model estimates a global alternative as it exploits the variability over the whole range of data. In fact, the Fourier specification provides a framework to estimate cost parametric function with a flexibility comparable to a nonparametric approach (Kuenzle, 2005). Empirical studies for banking (e.g. Mitchell and Onvural, 1996; Humphrey and Vale, 2004), aviation (Creel and Farell, 2001), farmers (Ivaldi et al., 1996) and electricity (Dashti,

---

[13] See e.g. Kuenzle (2005) and Martins and Fortunato (2006) for a discussion on parametric models for drinking water utilities.

[14] See Galant, 1982 for technical specifications.

2003) show that the Translog function is often too inflexible to reflect the true relationship between output and costs. Although the Translog specification is often applied to the drinking water sector (e.g. Saal and Parker, 2000), the Fourier cost function has to our knowledge never been applied. The Fourier model is specified as:

$$\begin{aligned} C_{i,t} \;=\; & \alpha_1 \, Q_{i,t} \;+\; \alpha_2 \, Q_{i,t}^2 \;+\; \beta_1 \, P_{L_{i,t}} \;+\; \beta_2 \, P_{L_{i,t}}^2 \;+\; \gamma_1 \, P_{C_{i,t}} \;+\; \gamma_2 \, P_{C_{i,t}}^2 \;+ \\ & \delta_1 Q P_{L_{i,t}} + \delta_2 Q P_{C_{i,t}} + \delta_3 \, P_{L_{i,t}} \, P_{C_{i,t}} \;+ \sum_{j=1}^{N} (\lambda_{1j} \; sin(j \; Q_{i,t}) \;+\; \lambda_{2j} \; cos(j \; Q_{i,t})) \;+ \\ & \xi Z_{i,t} + \theta_1 TIME + \theta_2 TIME^2 + \phi \; MERGE_i + \; \mu_{i,t} \end{aligned} \tag{13}$$

where N denotes the number of *sine* and *cosine* terms included (i.e. the grade of the function) and Z the vector of exogenous characteristics (including the intercept). N is determined by testing on the basis of an *F-test* whether the sum of squared residuals (SSR) differs significantly between two values of N. As such, it tries to capture the true relationship between costs and inputs and outputs. For instance, testing whether the SSR for $N = 0$ and $N = 1$ differs significantly, reveals whether a Fourier function of grade 1, which includes $\sin(Q)$ and $\cos(Q)$, is statistically preferred against a function of grade 0, i.e. a Translog model. Comparing $N = 1$ and $N = 2$ makes clear whether a Fourier function of grade 2, which includes $\sin(Q)$, $\sin(2Q)$, $\cos(Q)$ and $\cos(2Q)$, is preferred against a Fourier of grade 1. To define the vector of exogenous characteristics, $Z_{i,t}$, a literature review and extensive discussions with sector experts was undertaken. This resulted in two types of exogenous influences. As a first type we include five sub factors to correct for the provision area characteristics. These are exogenous due to the legal regional monopoly. We account for (1) the number of connections per unit water delivered, (2) network length per unit water, (3) the soil stability[15], (4) the customer mix (i.e. the proportion of water delivered to medium and large sized businesses) and (5) the age of the infrastructure (i.e. ageing infrastructure increases maintenance costs but reduces depreciation and interest). As a second type, we account for diverging purification efforts by including indices measuring (1) the quality difference between raw and drinking water and (2) the type of raw water used. An overview of the variable is presented in Appendix.

We start from the CD specification and test gradually the robustness of the results by different specifications. The estimated effects of benchmarking and merger economies, split into inefficiency reducing abilities and scale economies, are presented in Table 2. As the scale elasticity differs for each particular evaluated Q, the point elasticity is measured for three typical types of firms; (1) the average non-merged firm (Q = 59 million m^3), (2) the average merged firm (Q = 111 million m^3) and (3) the average of the 5 largest observations (Q = 236 million m^3). In comparison to the CD estimations on the disaggregated data set (see Table 1), the benchmarking effect on efficiency increases to 18%. The effect of mergers

---

[15] Some drinking water utilities deliver water in unstable soil regions which reduces the mains' life span.

becomes insignificant using the aggregated (orginal) data.[16] Although this seems puzzling at first sight, the result with aggregated data might be more reliable as the disaggregated analysis in fact duplicates the observations for merged firms (which is only relevant in the parametric analysis). As merged firms are split, degrees of freedom are created that do not exist in reality. In the CD specifications, the scale elasticities are the same for the different types of firms, indicating that a 1% increase in scale reduces efficiency by 0.09% (decreasing returns to scale). The Fourier function of grade 0, the Translog, estimates the effect of benchmarking at 17%, while the effect of MERGE remain insignificantly negative. According to this specification, only non-merged companies experience significant decreasing returns to scale. However, the optimal grade of the Fourier function is higher than 0 as tests point out that in a 'narrow' model with only the number of connections per unit water included as exogenous variable N equals 4, while the grade equals 3 for the 'extended' model which includes all exogenous variables. The CD and Translog specification are thus rejected against the Fourier specification. The narrow Fourier model estimates a significant positive effect of benchmarking and a significant efficiency reducing effect of mergers. It detects increasing returns to scale for the merged utilities and decreasing returns to scale for the 5 largest companies. Contrarily, the 'extended' model finds only a significant effect of the benchmark, and no significant merger economies. Finally, we test if the results change by using only data for companies with a production higher than 25 million m^3, which corresponds to the scale just below the smallest company in 2006. We use this cut-off level as we do not find a rejection of constant returns to scale for companies with a larger scale, while economies of scale are found for very small companies. It is interesting to analyze whether this discontinuity influences the results. By restricting the data set to the larger companies, the number of observations decreases from 242 till 204, while the number of included companies decreases from 20 to 16. The positive effect on efficiency of benchmarking and the insignificant effect of merger economies is robust to this data set specification as well.

Summarizing, benchmarking turns out to have a significant positive effect in all model specifications, while the effect of the merger economies (i.e. scale economies and inefficiency reducing incentives) is more ambiguous in the different models. None of the estimations resulted in a significant and positive effect of mergers on efficiency.

As the sign and size of the scale economies differ for different values of Q, we present these graphically for the Fourier model of grade 0 and grade 3. Figure 6 shows the estimated scale effects for the former, i.e. the Translog model which is frequently employed in the literature. The quantity of water produced is presented on the horizontal axis and the costs per unit water is on the vertical axis. The solid line reflects for each scale level Q the average estimated cost per m^3, while the crosses reflect the 95%-confidence level (these are only depicted for actual observations). Using the Translog, constant returns to scale is not rejected between

---

[16] Additionally, all specifications were tested with the square of $MERGE$ included. This resulted in insignificant coefficients of $MERGE$ and its square.

Table 2: Merger economies analyzed

	Estim. effect Benchmark	Inefficiency reduction of merger	Scale elasticity Non-merged (Q=59)	Merged (Q=111)	Large (Q=236)
Cobb-Douglas	18%**	-3%	-0.09**	-0.09**	-0.09**
	(6%)	(2%)	(0.02)	(0.02)	(0.02)
Fourier ($N = 0$) (i.e. Translog)	17%**	-3%	-0.09**	-0.07	-0.06
	(6%)	(2%)	(0.02)	(0.06)	(0.08)
Fourier ($N = 4$)	11%**	-5%*	0.03	0.37**	-0.59**
- Z = number of connections	(5%)	(2%)	(0.12)	(0.14)	(0.28)
Fourier ($N = 3$)	20%**	0%	-0.28	0.45	-0.09
- Z = all exogenous variables	(4%)	(2%)	(1.35)	(1.41)	(1.43)
Fourier ($N = 3$)	12%**	3%	-0.16	-0.05	-1.61
- Z = all variables - only large comp.	(4%)	(2%)	(1.31)	(1.36)	(1.40)

** and * denotes, respectively, significance at a 1 and 5%-level

35 and 350 million m^3, the confidence intervals are however rather large from 150 million m^3 on.

Figure 7 presents the results for the Fourier specification of grade 3 (based on the sample excluding observations for very small firms).[17] This figure arrives at two scale levels with minimal costs. The first is around 30 million m^3, while the second suggests a comparable low cost level at a much larger scale (i.e. around 200 million m^3). Note, however, that the confidence band is much wider for the latter due to the scarce number of observations around 150 million m^3. Indeed, no significant evidence is found for the hypothesis that mergers would imply positive scale economies. In fact, average costs are lower for non-merged companies (which have an average scale of 59 million m^3) compared with merged companies (which have an average scale of 111 million m^3), although again, this difference is not significant.

Summarizing, the results of the parametric approach show the robustness of the previous analysis. We detect an efficiency enhancing effect of benchmarking, and insignificant merger economies due to, in general, the absence of both scale economies and increased incentives to fight inefficiencies.

---

[17] Note that we exclude companies with a scale smaller than 25 million m^3 as estimates show that including these companies destabilizes the results for larger companies. We find scale economies for these small companies, which is in accordance with the literature.

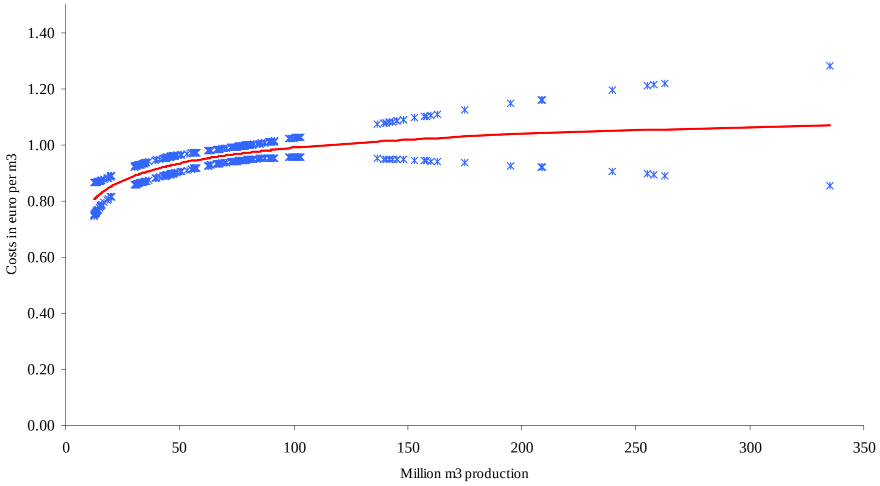

Figure 6: Scale economies with Fourier (N=0; Translog)

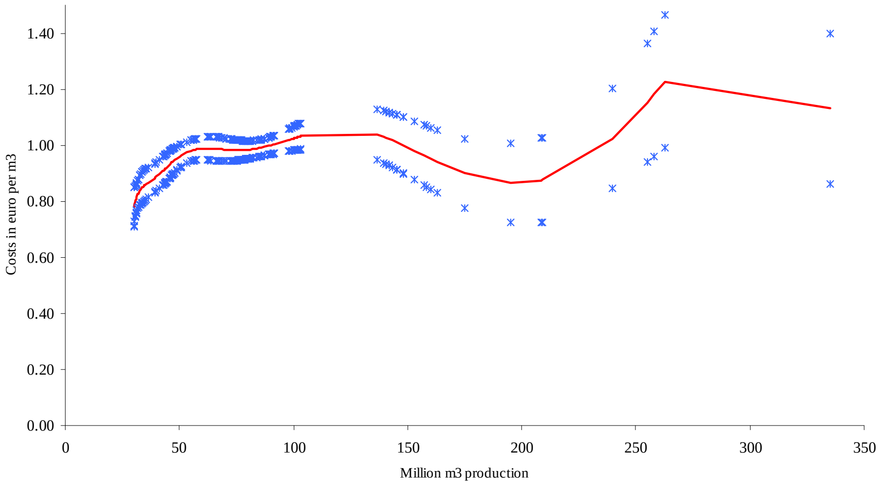

Figure 7: Scale economies with Fourier (N=3)

7 Conclusion

Studying the effects of benchmarking and mergers in the Dutch drinking water sector, the paper does not find an impact of mergers on efficiency. In particular, it ascertains the absence of the two underlying mechanisms to merger economies, i.e. scale economies and increased incentives to fight inefficiencies. Although scale economies are present for very small companies, larger utilities are not producing at lower cost. This observation is especially important as in many countries the drinking water sector is involved in a debate on the optimal scale of its utilities. In addition, the paper shows the effectiveness of the incentive regulation, i.e. a voluntary benchmarking project since 1997. This incentive regulation increased significantly the efficiency of the Dutch drinking water companies.

The paper allows to draw some policy implications with respect to mergers which go beyond the specific Dutch drinking water sectors. Regulators should be cautious with respect to factors which undermine the effectiveness of incentive regulation. Particularly mergers undermine the effectiveness of incentive regulation as the remaining utilities could easier invoke atypical circumstances. If the qualityh of incentive regulation is based on the quality of information about utilities, which is generally the case, mergers probably decrease the effectiveness of regulation. It is suggested that regulators should not give the benefit of doubt to merger projects of utilities in an incentive regulation environment if the number of firms is allready relatively low, which is the case in the Netherlands. On the contrary, these regulators should consider to break down larger companies into smaller utilities in order to obtain more comparable units and, hence, further increase the effectiveness of the incentive regulation.

Although the paper finds a positive effect of benchmarking on efficiency, the Dutch drinking water benchmark is still incomplete. Firstly, the benchmark is a voluntary initiative (although, currently, all companies cooperate). The Minister of Housing, Spatial Planning and Environment, who regulates the water sector, recently made a proposal to oblige participation to the benchmark. Secondly, the current benchmark does not account for exogenous characteristics among the drinking water utilities, which makes the presented information partly incomparable. Thirdly, accounting rules are still not harmonized which results ine.g. differences in depreciation rates. Finally, the few number of Dutch companies decreases the effectiveness of the benchmark. An international benchmark study which increases the scope of the analysis could deliver promising results. It could be expected that the elimination of these hiatuses will further enhance the effectiveness of the benchmark for the Dutch drinking water sector.

Appendix

Table 3: List of variables

Variable	Description
$Costs$	Total real cost per year in 1000 euro
Q	Production of drinking water in million m^3
P_C	Capital real costs per unit assets, based on price deflator investments
P_L	Labour real costs per employee, based on price deflator labour costs
$Connect$	Number of technical connections per m^3
$Length$	Length pipes (transport- and main lines) per m^3
$Soil$	Measures instability soil
$Purif$	Measures intensity of purification
Q_{large}	Deliveries to large size customers ($> 10,000$ m^3 per year) in $\%$ Q
Q_{mid}	Deliveries to middle size customers (300 - $10,000$ m^3 per year) in $\%$ Q
$Dunes_N$	Use of natural dune water
$Dunes_I$	Use of infiltrated dune water
Age	Age infrastructure (book value as $\%$ of purchase value)

Table 4: Descriptive statistics

	All companies				Merged	Non-merged
	Average	Max.	Min.	St. dev.	Average	Average
$Costs$	87	381	10	64	133	76
Q	69	335	12	49	111	59
P_L	50	66	27	6	52	50
P_C	44	103	12	18	35	46
$Connect$	5.2	7.3	1.6	1.3	5.3	5.2
$Length$	0.09	0.16	0.02	0.04	0.10	0.09
$Soil$	1.10	1.35	1.00	0.12	1.06	1.11
$Purif$	8.3	21.8	2.0	6.5	9.1	8.0
Q_{large}	0.18	0.45	0.04	0.09	0.17	0.18
Q_{mid}	0.19	0.37	0.05	0.07	0.22	0.19
$Dunes_N$	0.02	0.19	0.00	0.04	0.00	0.02
$Dunes_I$	0.13	1.03	0.00	0.29	0.23	0.11
Age	0.57	0.74	0.25	0.10	0.53	0.58

Table 5: Mergers in the Dutch drinking water sector 1992 - 2006

Name merger	Merged companies*	Year of merger
WMO	WMO + WOT	1996
DZH	DZH + main part EWR	1996
Wgeld	WMG + WOG	1997
Wgron	Waprog + GWG	1998
WOB	WOB + 2 municipalities	1998
Vitens 1	WMO + Wgeld + NuonWF	2002
BW1	WNWB + WOB	2002
Evides	WBE + Delta	2004
BW2	BW1 + TWM	2006
Vitens 2	Vitens 1 + HydronMN + HydronFL	2006

* Abbreviations from Vewin (2007)

References

[1] Ballance, A., S. Reid and D. Saal (2004), *Investigation into evidence for economies of scale in the water and sewerage industry in England and Wales*, Stone & Webster Consultants, Londen.

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

[3] Chambers, R., Y. Chung and R. Färe (1998). Profit, directional distance functions, and nerlovian efficiency. *Journal of Optimalization Theory and Applications*, 98, 351-364.

[4] Charnes, A., W.W. Cooper and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research,* 2 (6), 429-444.

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

[6] Cherchye, L., T. Kuosmanen and T. Post (2001), FDH directional distance functions with an application to European commercial banks, *Journal of Productivity Analysis,* 15 (3), 201-215.

[7] Coelli, T. and S. Perelman (1999), A comparison of parametric and non-parametric distance functions: With application to European railways, *European Journal of Operational Research*, 117 (2), 326-339.

[8] Cooper, W.W., L.M. Seiford and J. Zhu (2004), *Handbook on data envelopment analysis*, Kluwer Academic.

[9] Creel, M. and M. Farell (2001), Economies of scale in the US airline industry after deregulation: a Fourier series approximation, *Transportation Research Part E*, 37 (5), 321-336.

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

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

[12] Dashti, I. (2003), Inference from concave stochastic frontiers and the covariance of firm efficiency measures across firms, *Energy Economics*, 25 (6), 585-601.

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

[14] De Witte, K. and R. Marques (2007a), Towards a benchmarking paradigm in the drinking water sector, *Public Money and Management*, forthcoming.

[15] De Witte, K. and R. Marques (2007b), Designing incentives in local public utilities. An international comparison to the drinking water sector. *Mimeo University of Leuven.*

[16] De Witte, K. and R. Marques (2007c), Influential observations in frontier models, a robust non-oriented approach to the water sector. *Mimeo University of Leuven.*

[17] Dijkgraaf, E., R. de Jonge.g. v.d. Mortel, A. Nentjes, M. Varkevisser and D. Wiersma (1997), Mogelijkheden tot marktwerking in de Nederlandse watersector (Possibilities to introduce competition in the Dutch watersector). *Ministry of Economic Affairs, The Hague.*

[18] Dijkgraaf, E., S.A. van der Geest and M. Varkevisser (2007), The efficiency gains of benchmarking Dutch water companies. *Working Paper Erasmus University Rotterdam.*

[19] Fabbri, P. and G. Fraquelli (2000). Cost and structure of technology in Italian water industry, *Empirica*, 27 (1), 65-82.

[20] Färe, R., S. Grosskopf and K. Lovell (1985), *The measurement of efficiency of production.* Boston (USA): Kluwer Nijhoff Publishing, p. 216.

[21] Galant, A.R. (1982), Unbiased determination of production technologies, *Journal of Econometrics*, 20, 285-323.

[22] Garcia, S. and A. Thomas (2001), The structure of municipal water supply costs: application to a panel of French local communities, *Journal of Productivity Analysis*, 5 (1), 5-29.

[23] Greene, W. H. (1980), Maximum Likelihood Estimation of Econometric Frontier Functions, *Journal of Econometrics*, 13 (1), 27-56.

[24] Gugler K., D.C. Mueller, B.B. Yurtoglu and C. Zulehner (2003), The effects of mergers: an international comparison, *International Journal of Industrial Organization*, 21 (5), 625-653.

[25] Humphrey, D.B. and B. Vale (2004), Scale economies, bank mergers and electronic payments: A spline function approach, *Journal of Banking & Finance*, 28 (7), 1671-1696.

[26] Ivaldi, M., N. Ladoux, H. Ossard and M. Simioni (1996), Comparing Fourier and translog specification of multiproduct technology: Evidence from an incomplete panel of French farmers, *Journal of Applied Econometrics*, 11 (6), 649-667.

[27] Ivaldi, M. and G. McCullough (2005), Welfare tradeoffs in U.S. rail mergers, *Center for Economic Policy Research*, London.

[28] Kim, H.Y. (1987), Economies of scale in multi-product firms: an empirical analysis, *Economica*, 54 (214), 185-206.

[29] Kuenzle, M. (2005), Cost efficiency in network industries: Application of stochastic frontier analysis, Phd. thesis, *Swiss Federal Institute of Technology*, Zurich.

[30] Kwoka, J. (2006), The role of competition in natural monopoly: Costs, public ownership and regulation, *Review of Industrial Organization*, 29 (1-2), 127-147.

[31] Kwoka, J. and M. Pollitt (2007), Industry restructuring, mergers, and efficiency: Evidence from Electric power, *Cambridge University.*

[32] Martins, R. and A. Fortunato (2006), Cost structure of the Portuguese water industry, *Universiteit van Coimbra.*

[33] Mitchell, K. and N.M. Onvural (1996), Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form, *Journal of Money, Credit and Banking*, 28 (2), 178-199.

[34] Mizutani, F. and T. Urakami (2001), Identifying network density and scale economies for Japanese water supple organizations, *Paper in Regional Science*, 80, 211-230.

[35] OECD (2004), *Competition and regulation in the water sector*, Paris.

[36] Portela, M., P. Borges and E. Thanassoulis (2003), Finding closest targets in non-oriented DEA models: the case of convex and non-convex technologies. *Journal of Productivity Analysis,* 19 (2-3), 251-269.

[37] Röller, L.H., J. Stennek and F. Verboven (2000), Efficiency gains from mergers, *The Research Institute of Industrial Economics*, Stockholm.

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

[39] Sauer, J. (2005), Economies of scale and size optimum in rural water supply, *Water Resource Research*, 41 (1), 1-13.

[40] Shleifer, A. (1985). A theory of yardstick competition. *Rand Journal of Economics*, 6 (3), 319-327.

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

[42] Sung, N. and M. Gort (2006), Mergers, capital gains, and productivity: Evidence from U.S. telecommunications mergers, *Contemporary Economic Policy*, 24, 382-394.

[43] Tulkens, H. (1993), On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts, and urban transit, *Journal of productivity analysis* 4 (1), 183-210.

[44] Vewin (2007), Reflections on performance 2006, *Association of Dutch Water Companies*, Rijswijk.