



UvA-DARE (Digital Academic Repository)

Influential observations in frontier models, a robust non-oriented approach to the water sector

de Witte, K.; Marques, R.C.

DOI

[10.1007/s10479-010-0754-6](https://doi.org/10.1007/s10479-010-0754-6)

Publication date

2010

Document Version

Final published version

Published in

Annals of Operations Research

[Link to publication](#)

Citation for published version (APA):

de Witte, K., & Marques, R. C. (2010). Influential observations in frontier models, a robust non-oriented approach to the water sector. *Annals of Operations Research*, 181(1), 377-392. <https://doi.org/10.1007/s10479-010-0754-6>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Influential observations in frontier models, a robust non-oriented approach to the water sector

Kristof De Witte · Rui C. Marques

Published online: 29 May 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract This paper suggests an outlier detection procedure which applies a nonparametric model accounting for undesired outputs and exogenous influences in the sample. Although efficiency is estimated in a deterministic frontier approach, each potential outlier initially benefits of the doubt of not being an outlier. We survey several outlier detection procedures and select five complementary methodologies which, taken together, are able to detect all influential observations. To exploit the singularity of the leverage and the peer count, the super-efficiency and the order- m method and the peer index, it is proposed to select these observations as outliers which are simultaneously revealed as atypical by at least two of the procedures. A simulated example demonstrates the usefulness of this approach. The model is applied to the Portuguese drinking water sector, for which we have an unusually rich data set.

Keywords Nonparametric estimation · Frontier · Non-oriented · Outliers · Water sector

1 Introduction

Non-parametric frontier models, such as Free Disposal Hull (FDH, Deprins et al. 1984) and Data Envelopment Analysis (DEA, Charnes et al. 1978), are convenient tools to estimate the relative efficiency of observations. However, the traditional models faced two major drawbacks. Firstly, they did not allow to incorporate exogenous variables (i.e. variables which

K. De Witte (✉)

Centre for Economic Studies, University of Leuven (KU Leuven), Naamsestraat 69, 3000 Leuven, Belgium
e-mail: kristof.dewitte@econ.kuleuven.be

K. De Witte

Top Institute for Evidence Based Education Research, Maastricht University, Kapoenstraat 2, 6200 MD Maastricht, The Netherlands

R.C. Marques

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

are beyond the influence of the firm's managers) without imposing a separability condition between the inputs and outputs, and the exogenous environment. Therefore, the traditional methods to incorporate these exogenous variables (see Fried et al. 2008 for an overview) assumed that the input-outputs and the exogenous variables are located in a different sphere. Only recently, by the conditional efficiency approach of Cazals et al. (2002) and Daraio and Simar (2005) has this major drawback been resolved. As a second disadvantage, the traditional FDH and DEA models are deterministic in nature in that they assume that there is no noise, nor atypical observations in the sample. This issue has been tackled by the robust efficiency model of Cazals et al. (2002).

Although the robust efficiency model mitigates the influence of outlying observations (i.e. gives them the benefit of the doubt of not being an outlier),¹ for the applied researcher it can be extremely interesting to detect these atypical observations. As atypical observations can present the best or worse practices, in this paper we develop a model that is able to detect a large range of these entities. Once the applied researcher knows the atypical observations, an in-depth case study should be carried out.

This article contributes to the literature from three angles. Firstly, it develops a specially tailored model to the application (i.e. the drinking water sector for which we have an unusually rich data set of Portuguese utilities). As the production technology is unknown, our model is entirely non-parametric, such that it does not assume any *a priori* specification on the functional form of the production function. The developed composite efficiency score is inspired on the directional distance functions of Chambers et al. (1998) and is able to deal simultaneously with potential input reductions and output expansions. In addition, the model makes a distinction between desired and undesired outputs and takes exogenous environmental effects into account (cf. Chung et al. 1997). Following Portela et al. (2003), we label models (and thus also the proposed model which is rooted in the directional distance function approach of Chambers et al. 1998) that simultaneously estimate potential input reductions and outputs expansions as 'non-oriented' models (although one should not necessarily agree with this terminology). 'Non-oriented models' have largely been undervalued in the productivity measurement literature, although they are considerably more realistic in numerous real-world cases in comparison to 'oriented' models. Indeed, in most scenarios a firm's manager intending to improve its entities' efficiency, will consider simultaneous input reductions, output expansions and undesired output diminutions, instead of working only in one of these dimensions separately.

As a second contribution, the paper combines several outlier detection procedures in one model. By combining complementary outlier detection models, we aim at detecting a broad range of atypical observations. Indeed, atypical observations could deserve an increased interest in that they could present the best or worst practices. This contrasts to the outlier detection procedures presented in the literature which have their own characteristics and singularity and, therefore, are unable to detect all atypical observations, but only the observations with similar characteristics (for an overview on the several outlier detection models in the literature, see Appendix). We exploit this conclusion by selecting five complementary methodologies which, taken together, are able to detect a broad range of influential observations. The effectiveness of the model is tested on a simulated data set.

¹The term 'benefit of the doubt' is prominent in the DEA literature as DEA uses an endogenous weighting (e.g. Melyn and Moesen 1991; Despotis 2005; Cherchye et al. 2007). Indeed, DEA retrieves the information on the weights from the data themselves. Therefore, a higher weight is given to variables on which the evaluated observation is doing relatively better.

Thirdly, the model includes some features which initially mitigate the influential impact of atypical observations. Indeed, by exploiting the robust order- m technique of Cazals et al. (2002), the model includes all observations in the sample (thus also the atypical observations) but gives less weight to these observations (see Sect. 2).² Therefore, every potential outlier initially obtains the benefit of the doubt of not being an outlier by mitigating the effect of the outlying observations. In other words, an observation is only an outlier if it is detected by the combined model after the influence of the outlier has been reduced (benefit of the doubt). In addition, the non-parametric model allows for diverging exogenous variables by the conditional efficiency scores of Daraio and Simar (2005, 2007). In our and in similar applications, outliers could arise from (1) the exceptionally high or low variables compared to the remaining utilities in the sample, (2) the measurement errors (especially due to different methodologies to measure, e.g., capital outlays), (3) the different structure in ownership and (4) the utilities which bring to light the best or worst practices.

The paper is organized as follows. Section 2 discusses the non-oriented model and provides the methodological background to estimate robust non-oriented conditional efficiency measures. In addition, it shows how to detect influential observations. Section 3 presents the results of the model on a simulated data set while Sect. 4 applies the model to the efficiency evaluation in the Portuguese drinking water sector. Finally, Sect. 5 draws the conclusions.

2 Methodology

Observations which disproportionately distort the relative efficiency scores of the dominated observations are labeled as influential. These observations are both extremely disturbing as they reduce the informative value of the efficiency scores, and extremely informative as they could indicate the best practices and efficient observations. This ambiguous role of outliers indicates the importance of detecting the proper atypical observations, so that researchers can take a close look at these units. Our outlier detection procedure differs from the traditional methods in the literature in two aspects. Firstly, thanks to the specially tailored ‘robust order- m model’ (see Sect. 2.2), it has the advantage to give the benefit of the doubt to atypical observations as the partial frontier mitigates the influence of outliers. Secondly, by exploiting five complementary procedures, each with their own singularity and capabilities to detect outliers with similar characteristics, the full model is able to detect all ‘proper’ outliers (see Sect. 2.3). We start this section by developing a non-oriented model which is inspired on directional distance functions (see Sect. 2.1).

2.1 Non-oriented efficiency estimation with correction for unwanted outputs

We first present the non-oriented model in the two-dimensional Fig. 1 with one input and one desired output variable and contrast it to the traditional input and output orientation. The input-orientation measures the minimal amount of inputs needed to produce a given amount of output (i.e., the distance CB/CA); the output-orientation searches for the maximal quantity of outputs which can be produced with a given level of inputs (i.e., the distance ED/EA); the non-oriented approach measures simultaneously the potential input decrease and potential output increase. Non-oriented models are introduced in the literature in 1985 with meaningful differences between them by Charnes et al. (1985) and Färe et al. (1985). The former, called the additive model, aims at changing inputs and outputs simultaneously

²The paper can be straightforwardly adapted to the robust order- α context of Daouia and Simar (2007). As indicated by Aragon et al. (2005) and Daouia and Ruiz-Gazen (2006), this could be worthwhile as order- α frontiers are more robust than order- m frontiers.

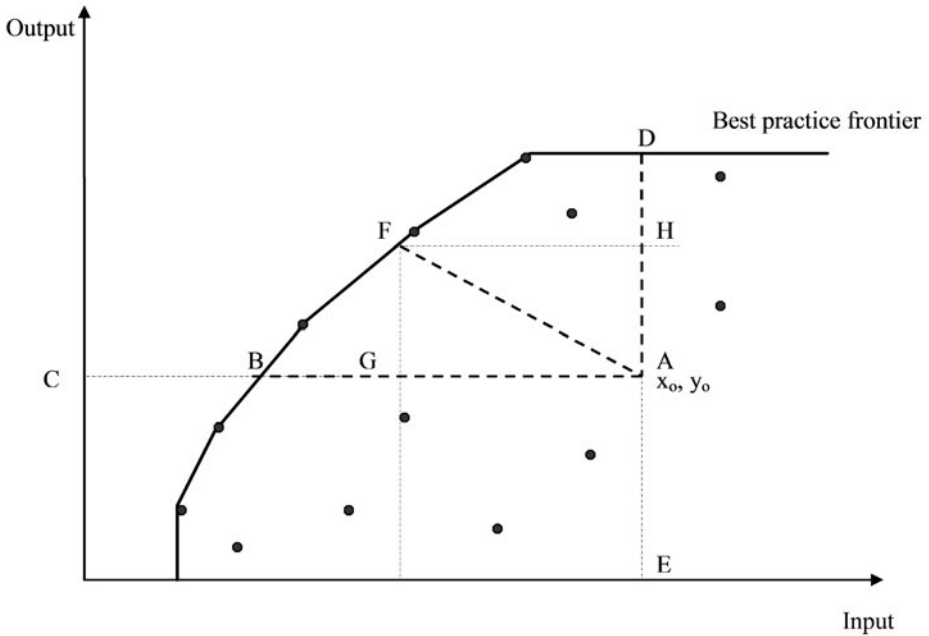


Fig. 1 Oriented and non-oriented models

in the sense of vector optimization, i.e., this model maximizes the L_1 distance (also known as ‘city-block distance’) of the observation under analysis to the observed efficient frontier assuming variable returns to scale (VRS). The latter, known as the hyperbolic model, maximizes the equiproportionate expansion of the output vector and the reduction of the input vector that places a given observation on the boundary of the technology (in particular, the hyperbolic model is non-linear and estimates a hyperbolic path $(\beta x, \beta^{-1}y)$ from the evaluated observation to the frontier).

Besides these two non-oriented models, variants such as those based on the Range Adjusted Measure (RAM) of Cooper et al. (1999) and on the directional distance function of Chambers et al. (1998), were developed. As Portela et al. (2003) point out, all these non-oriented models share the common feature of maximizing slacks but rarely satisfy the desired features concerning the inclusion of all the sources of inefficiency and the meaning of radial inefficiency measures at the same time. Exceptions are the measures proposed by Brockett et al. (1997) and Portela et al. (2003). Notice that the non-oriented model has several advantages and therefore various potential empirical applications, however, in the water sector, only two studies were found in the literature (Aida et al. 1998 and Lambert et al. 1993, see De Witte and Marques 2007 for an overview of the literature). Our specially tailored nonparametric model is also well-suited for other environments where the production technology is unknown (e.g., inefficiency measurement in other utilities, education, health, social sector, etc.).

Next, consider a general setting of n utilities which transform p inputs $x \in \mathbb{R}_+^p$ into q desired outputs $y \in \mathbb{R}_+^q$ and r undesired outputs $b \in \mathbb{R}_+^r$. The inefficiency of the utilities is measured relatively to a frontier constituted from the best practices, so that any deviation from the frontier is assumed to be attributed to inefficient management. The production

technology set Ψ defines the set of all feasible input-output combinations:

$$\Psi = \{(x, y, b) \text{ such that } x \text{ can produce } y \text{ and } b\}. \tag{1}$$

The production set Ψ is typically unknown in empirical studies and is nonparametrically approximated from the set of observations drawn from the technology. Following Ray and Mukherjee (2007), we assume free disposability of the inputs and desired outputs and reverse disposability of the undesired outputs, i.e., let $\tilde{x} \in \mathbb{R}_+^p$, $\tilde{y} \in \mathbb{R}_+^q$ and $\tilde{b} \in \mathbb{R}_+^r$, with $\tilde{x} \geq x$, $\tilde{y} \leq y$ and $\tilde{b} \geq b$ if $(x, y, b) \in \Psi$ then $(\tilde{x}, y, b) \in \Psi$, $(x, \tilde{y}, b) \in \Psi$ and $(x, y, \tilde{b}) \in \Psi$.³ This free disposability assumption delivers the FDH model of Deprins et al. (1984). The DEA model of Charnes et al. (1978) is obtained by assuming in addition to free disposability also convexity: if $(x_1, y_1, b_1), (x_2, y_2, b_2) \in \Psi$, then $\forall \alpha \in [0, 1]: (x, y, b) = \alpha(x_1, y_1, b_1) + (1 - \alpha)(x_2, y_2, b_2) \in \Psi$. We believe these two minimalistic assumptions to be appropriate in the context under study.

For every observation, we can estimate the inefficiency relatively to this efficient boundary of Ψ by estimating the scalar β :

$$\delta\Psi = \{(x, y, b) \in \Psi | ((1 - \beta)x, (1 + \beta)y, (1 - \beta)b) \notin \Psi, \forall 0 < \beta < 1\}. \tag{2}$$

The efficiency is estimated relatively to this efficient boundary. As such, β estimates the distance to the frontier (i.e. the distance AF in two-dimensional Fig. 1).

To operationalize the non-oriented model, we start from a directional distance function as developed by Chambers et al. (1998) which measures the radial reduction in inputs and expansion in outputs to the frontier in a preassigned direction (g_x, g_y, g_b) . For the non-oriented model, we assume the direction of the distance $(g_x, g_y, g_b) = (x_o, y_o, b_o)$ such that:

$$\begin{aligned} \vec{D}(x_o, y_o, b_o, g_x, g_y, g_b) &= \sup\{\beta | x_o - \beta g_x, y_o + \beta g_y, b_o - \beta g_b \in \Psi\} \\ &= \sup\{\beta | (1 - \beta)x_o, (1 + \beta)y_o, (1 - \beta)b_o \in \Psi\}, \end{aligned} \tag{3}$$

where x_o, y_o and b_o denote the evaluated observations, $g_x (= x_o)$, $g_y (= y_o)$ and $g_b (= b_o)$ the direction vectors in which \vec{D} is defined and β is a scalar. Note that this approach is close to the hyperbolic efficiency idea of Färe et al. (1985). We apply the directional distance function ideas to the Data Envelopment Analysis (DEA) estimator of the production set. Assuming Variable Returns to Scale (VRS), which is in the context under study an appropriate assumption (De Witte and Marques 2007) and which is consistent whatever the hypothesis on returns to scale (Daraio and Simar 2007), the estimator of the non-oriented inefficiency score can be obtained by solving the linear program:

$$\begin{aligned} \hat{\beta}(x_o, y_o, b_o) &= \max \left\{ \beta | (1 - \beta)x_o \geq \sum_{i=1}^n \lambda_i x_i; (1 + \beta)y_o \leq \sum_{i=1}^n \lambda_i y_i; (1 - \beta)b_o \right. \\ &\quad \left. \geq \sum_{i=1}^n \lambda_i b_i; \sum_{i=1}^n \lambda_i = 1; 0 \leq \beta \leq 1; i = 1, \dots, n \right\}. \end{aligned} \tag{4}$$

³In the nonparametric literature, undesired outputs are alternatively assumed as weakly disposable. However, we prefer to follow the approach of Ray and Mukherjee (2007) as a reduction of the bad output is a natural assumption in many applications. An extension of the model to weakly disposable bad outputs is straightforward.

An efficient observation constituting the best practice frontier obtains a value of $\hat{\beta} = 0$, while inefficient observations reach positive values.

2.2 Reducing the influence of outliers and environmental factors

The deterministic nature of nonparametric frontier models can be moderated by limiting the number of potential reference partners which constitute the frontier. This so-called robust order- m approach, as suggested by Cazals et al. (2002), draws a partial frontier which is more robust to atypical observations and has some attractive statistical properties. We adapt the method for the specific inefficiency measure $\hat{\beta}(x_o, y_o, b_o)$ defined above. For an evaluated observation (x_o, y_o, b_o) a subsample of size m is repeatedly drawn ($r = 1, \dots, R$ times) with replacement among the original sample (x, y, b) . For each subsample, the inefficiency score $\hat{\beta}_m^r(x_o, y_o, b_o)$ is estimated along the linear program in (4). Finally, the robust non-oriented estimator is defined as the arithmetic average:

$$\hat{\beta}_m(x_o, y_o, b_o) = \frac{1}{R} \sum_{r=1}^R \hat{\beta}_m^r(x_o, y_o, b_o). \quad (5)$$

The more references in the subsample (and thus the higher m), the higher the probability an influential observation will constitute the frontier and the closer the estimate $\hat{\beta}_m$ will approximate $\hat{\beta}$. Remark that some observations could obtain negative inefficiency scores as due to the resampling the evaluated entity (x_o, y_o, b_o) is not included in the subsample. Following Andersen and Petersen (1993), we label these units as ‘super-efficient’.

Inspired on the robust order- m approach of Cazals et al. (2002), Daraio and Simar (2005, 2007) suggest a conditional approach to take into account environmental influences z without imposing separability (i.e. without assuming that the inputs and outputs are not immediately influenced by the exogenous variables). Favorable environmental influences behave as substitutive inputs, while unfavorable environmental variables behave as substitutive outputs needing additional inputs to compensate. To immediately include the exogenous variables in the non-parametric model, Daraio and Simar (2005) suggest to adapt the robust order- m efficiency model in a way that the reference set of size m is drawn with a particular probability. In particular, they suggested to compare like with like by non-parametrically smoothing (by a Kernel estimator $K(\cdot)$ with a bandwidth h as estimated by the likelihood cross-validation criterion) the exogenous variable z . The Monte-Carlo algorithm of the robust estimator is accommodated by drawing the m observations with a probability equal to:

$$K((z_o - z)/h) / \sum_{j=1}^n K((z_o - z_j)). \quad (6)$$

As shown before, we adapt the conditional efficiency model to our specific inefficiency measure, such that the obtained robust conditional inefficiency estimate $\hat{\beta}_m(x_o, y_o, b_o | z_o)$ does not suffer from a separability condition as the value of z influences the position of the frontier (see also De Witte and Korteleinen 2008).

2.3 Detecting influential observations

A typical phenomenon in the operations research literature is to carefully analyse atypical observations in an attempt to find best practices (Yin 2003). This section describes a procedure which combines five complementary outlier detection models. By combining the models, we are able to detect a broad range of (potentially interesting) atypical observations.

As a first outlier detection procedure, we select the ‘leverage’ concept which explicitly exploits the disproportional influence of atypical observations (Sousa and Stocic 2005). For each observation, the leverage (l_o) is measured as the standard deviation of the inefficiency estimates relative to the full sample without the evaluated observation ($\hat{\beta}_{m,i,o}$), and the inefficiency estimates relative to the full sample ($\hat{\beta}_{m,i}$). Formally, we measure:

$$l_o = \sqrt{\frac{\sum_{i=1; i \neq o}^n (\hat{\beta}_{m,i,o} - \hat{\beta}_{m,i})^2}{n-1}}. \quad (7)$$

A higher leverage value, l_o , provides evidence for an influential observation. In our robust order- m model, the leverage equals the average leverage over the R drawings in the sub-samples. Remark that, thanks to the robust order- m model we apply, our methodology gives every potential outlier initially the ‘benefit of the doubt of not being an outlier’ (see before). In addition, due to the resampling in order- m , also inefficient observations will obtain positive leverage values.

The second outlier detection procedure, the *peer count*, suggested by Charnes et al. (1985), consists of the computation of how many observations are influenced by a certain efficient observation. Both the observations with a higher or lower peer count are candidates to be outliers. Again, the peer count is reinterpreted as the average frequency that an observation influences the other observations (i.e., frequency over number of drawings R).

Third, the *super-efficiency* concept, developed by Andersen and Petersen (1993), determines to what extent the efficient observations can simultaneously increase their inputs and undesired outputs and reduce the intended outputs by keeping themselves technically efficient. In numerical terms, the procedure consists in removing the evaluated observations from the sample at the moment of the DEA-evaluation, so that the observations are compared with the linear combination of all the others, except for themselves. The lower the super-efficiency (i.e., the more negative), the higher the probability that an entity is an outlier. However, this methodology, in spite of sorting the efficient units with regard to the efficiency surpluses, does not say anything about their sorting according to the importance of the efficient DMUs as reference or benchmarking element for the inefficient observations of the sample.

Fourth, the *order- m* method for detecting outliers, as developed by Simar (2003) based on the work of Cazals et al. (2002), determines observations far above the frontier (i.e., a value considerably below 0) as outliers. Therefore, the model simply estimates and ranks the above mentioned order- m inefficiency scores.

Finally, the *peer index* computes for each of the input and output variables of a unit the potential input reduction and output expansion (Torgersen et al. 1996). As such, the peer index is computed for every variable k of each efficient observation j . Summarizing from Torgersen et al. (1996), the peer index ρ_j^k is constructed as the ratio of the total weighted saving potential of the inefficient units in the reference set of the peer and the total saving potential in the complete data set. The measure ρ_j^k expresses the percentage of the potential reduction of an input that is represented by the inefficient unit which depend on the efficient unit j . The higher the ρ_j^k , the larger the possibilities of employing that observation for benchmarking or in other perspective the larger the possibility that it is an outlier. Both observations with a higher or lower peer index indicate potential outliers. Similar as to the other outlier detection methods, the peer index is reinterpreted as the average potential input reduction and output expansion (i.e., the average over number of drawings R).

Having their own singularity and set-up, as shown in the next section by a simulation, these five procedures complement each other. Taken together, they are able to detect a broad

range of influential observations (see simulation exercise in Sect. 3). Indeed, the leverage and peer count draw special attention to the influence of the atypical observation on the other observations in the sample; the super-efficiency and the order- m method emphasize in particular the atypical inefficiency scores; while finally the peer index attaches great importance to the individual variables of the evaluated entity. Individually, these procedures only consider a specific variant of atypical observations, while taking the particular characteristics together, these procedures are supposed to reveal all ‘proper’ outlying observations. To fully exploit the complementarity of the selected outlier detection procedures, we consider observations which prevail in at least two different procedures as outliers (labelled as $u = 2$). The appropriate value of u results from the complementarity of the model since there are, except for the atypical variables, always two different procedures for each source of outliers (i.e., there are two different procedures for atypical observations; two different procedures for atypical inefficiency scores). The robustness of the analysis improves by increasing u , although not all atypical observations will be detected anymore. Indeed, using $u = 2$, both influential outliers (by leverage and peer count), atypical inefficiency scores (by super-efficiency and order- m) and atypical variables (by peer index) will be detected. This is tested on a simulated data set in the next section. Note that, the robustness of the analysis improves by increasing u , although less observations will be labelled as outliers.

Remark that our approach is analogous to the ideas of Wilson (1995) in that he constructs a similar composite outlier detection procedure relying on the principles of super-efficiency. However, our approach is more balanced than the Wilson procedure as it, following the super-efficiency, mainly emphasizes the atypical inefficiency scores. Other outlier detection procedures are surveyed in Table 4 in the Appendix.

3 Simulation

To test whether the conditional non-oriented model, which gives potential outliers initially the benefit of the doubt, suitably detects the ‘proper’ outliers both in the sense of atypical efficient units (i.e., atypical little use of inputs and atypical large production of outputs) and atypical inefficient observations (i.e., abundant use of inputs), we simulate a data set of 65 observations each with two inputs, two outputs, one undesired output, one favorable environmental variable and one unfavorable environmental variable. The variables are generated as follows. The input variables are randomly drawn from a uniform distribution between 0 and 1 (x_1 and $x_2 \sim U(0, 1)$). The first output variable is the aggregate of the two input variables ($y_1 = x_1 + x_2$), while the second output variable is constructed by $y_2 = \text{undesired output} + ((1 + x_1) * \text{environmental}_1) + (x_2 * \text{environmental}_2)$. The undesired output variable is generated as the product of an exponential distributed variable with mean 3 and the first input variable ($\text{undesired output} = \exp(3) * x_1$). The favorable environmental variable is drawn from an exponential distribution with mean 2 ($\text{environmental}_1 \sim \exp(2)$), while the unfavorable exogenous factor is drawn from an exponential distribution with mean 1 ($\text{environmental}_2 \sim \exp(1)$). In this simulated exercise, we determined $R = 50$ and the size of the partial frontier $m = 40$ as from these values on, the percentage of super-efficient observations is more or less stable (cfr. Daraio and Simar 2005, 2007).

The simulated data are enriched by 5 atypical observations a . Observation a_1 uses significantly more inputs than the observation with the highest input value, a_2 consumes significantly less inputs to produce the outputs, a_3 produces significantly more outputs than the predicted output according to the data generation, a_4 produces an atypical large amount of the undesired outputs while a_5 possesses an extremely favorable environmental factor.

A priori, the model is expected to compensate for environmental variables which are beyond the scope of the firm’s managers and, thus, we should not detect a_5 as an influential observation. In addition, as managers are able to deal with undesired outputs, a_4 should not be detected as atypical, conversely, it should obtain a low inefficiency score. The model should not mitigate super-inefficient (i.e. a_1) and super-efficient observations (a_2 and a_3) which we would like to trace as atypical.

Our findings are in line with the anticipated results. On the one hand, the individual procedures only detect their own singular outliers, while on the other hand, the five methodologies prove to be complementary to discover all atypical observations. Average inefficiency of the simulated sample amounts to 0.069 with a standard deviation of 0.163. Firstly, consider the outcomes of the individual procedures. The results are presented in Table 1 where we denoted the detected outliers by O , the efficient observations according to the procedure by E and the inefficient observations as I .⁴ The observed values are presented in brackets. First consider the leverage procedure. The leverage equals on average 0.0675. Observations a_1 , a_2 and a_3 according to the model considered as outlying observations. Note that these influential effects are already mitigated by the robust inefficiency estimates which give each observation the benefit of the doubt of not being an outlier. The second procedure of peer count does not detect unit a_2 as influential, but delivers for the other observations similar results to the leverage. Thirdly, although super-efficiency ascertains unit a_2 as atypical, it does not trace a_1 as so.⁵ In addition, the order- m inspired procedure of Simar (2003) struggles

Table 1 Outlier procedures in a simulated sample

		Leverage	Peer count	Super-eff	Simar
a_1	abundant inputs	O (0.1191)	O (0)	I (0.4753)	I (0.4793)
a_2	little inputs	O (0.0783)	E (1)	O (-0.8734)	E (0.000)
a_3	huge output	O (0.1092)	O (29)	O (-0.1748)	O (-0.1745)
a_4	undesired output	I (0.0554)	I (3)	I (0.6345)	I (0.0597)
a_5	environmental	E (0.058)	E (1)	E (0.000)	E (0.000)
average		0.0675	3	0.0150	0.0788
st. dev		0.0222	4.1	0.5105	0.1717
max		0.1191	29	0.6794	0.4793
min		0.0325	0	-2.1162	-0.1745
		ρ_j^{input1}	ρ_j^{input2}	$\rho_j^{output1}$	$\rho_j^{output2}$
		$\rho_j^{bad_output}$			
a_1	I (0.000)	I (0.000)	I (0.000)	I (0.000)	I (0.000)
a_2	E (0.000)	E (0.000)	E (0.000)	E (0.000)	E (0.000)
a_3	O (0.0260)	O (0.0254)	O (0.0408)	O (0.0530)	O (0.0653)
a_4	I (0.000)	I (0.000)	I (0.000)	I (0.000)	I (0.000)
a_5	E (0.000)	E (0.000)	E (0.000)	E (0.000)	E (0.000)

Where O denotes outlier, E = efficient, I = inefficient; estimated values in brackets

⁴As discussed before, only for the peer index, the categorization in efficient and inefficient is not clear from the figures (i.e., $\rho_j^k = 0$). One should look at the underlying data.

⁵The estimation of the Wilson (1995) procedure to detect outliers yielded similar results to the super-efficiency approach on which it heavily draws.

with a_2 and a_1 . The peer index only detects observation a_3 as atypical (and not observation a_1 and a_2). These findings confirm that each outlier detection procedure has its own peculiarities and caution should be taken if selecting one or another procedure.

Secondly, consider the singularity of the different outlier detection procedures, and combine these procedures into one outlier detection model. Here, the picture changes dramatically. Observations which prevail in at least two different procedures ($u = 2$) are considered as influential observations (i.e., observations a_1, a_2 and a_3). This simulated example indicates the appropriate working of the model.⁶

4 Application: efficiency in Portuguese drinking water sector

4.1 Model

In 2004, the Portuguese regulator, Institute for the Regulation of Water and Solid Waste (IRAR), started a performance indicator based benchmarking project for the private Portuguese water utilities. It is expected that this benchmarking exercise (by the use of a sunshine regulation, i.e., the computation, comparison and publicizing of performance) will enhance efficiency and productivity in the drinking water sector. In this section, we apply the specially tailored benchmarking tool to nonparametrically assess the inefficiency of the Portuguese water utilities while simultaneously allowing for input reductions, output expansions, penalization for undesired outputs and correction for environmental variables. The sample consists of 73 water utilities, covering 64% of the Portuguese population (6.6 millions of inhabitants) and corresponding to both the largest water services and to those that have separate accounts for the water activities. It includes private companies, municipal companies under private law, municipal companies under public law and municipal services. The data refer to the year 2005 and are obtained by the annual account reports of the water services.

The variables in the model are constructed as follows: as inputs we adopted the operational expenses (OPEX) and the capital expenses (CAPEX). As all water utilities belong to a rather homogeneous group (i.e., to the same country), the monetary units are well comparable. Caution, however, was taken regarding the CAPEX values owing to the different account rules depending on the ownership (see *supra*). The use of monetary variables creates some advantages since these incorporate all the inputs including their quality. The quantity units often employed, mainly in international benchmarking studies, hardly capture all the input issues. For example, the use of mains length as a proxy of CAPEX does not account for the quality of the network (e.g., ageing) and it disregards the other assets in the analysis (e.g., treatment plants and pumping stations). As desired outputs, we adopted the delivered (revenue) water volume and the number of customers, which are consensual in the literature. Moreover, in Portugal, as in the water tariff system there is a fixed component corresponding to the meter (which is a link to the number of customers) and another variable component related to the sales of water (expressed in m^3), these outputs make sense together. Our model considers one undesired output, i.e., the water losses volume (m^3). In Portugal, a substantial area of the country suffers from the lack of water resources, eventually leading to the rationing of consumption in some water utilities in summer times. Finally, as explanatory

⁶We estimated the simulated data with and without weight restrictions, and with and without exogenous and bad variables. Using $u = 2$, each of these models specifications detected the appropriate atypical observations.

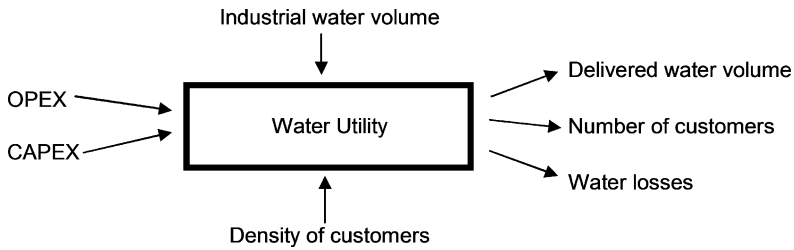


Fig. 2 Specification of the model

Table 2 Descriptive statistics of the Portuguese water utilities sample

	Inputs		Outputs			Explanatory factors	
	OPEX (milj. euro)	CAPEX (milj. euro)	Delivered vol. (m ³)	Customers (no.)	undesired losses (m ³)	Customers density	Industrial volume
average	6,184	2,024	5,837,947	42,472	2,898,830	65.9	972,782
std. dev.	9,645	4,614	8,460,023	53,256	5,030,354	42.8	587,555
median	3,479	923	3,196,112	26,500	1,501,235	56.3	639,542
min.	381	80	334,774	4,364	171,828	12.2	3,348
max.	71,489	38,390	62,297,940	341,764	36,655,050	239.5	14,483,586

variables, we introduce the density of customers (number/km) and the industrial water volume (m³). First consider the former: the density of customers results in a typical U-shaped cost. Indeed, on the one hand, water utilities with a high density of customers will face higher OPEX and CAPEX (and thus more total expenses, TOPEX) since the networks turn out to be more complex and require more inputs to produce the same proportion of outputs. On the other hand, a low density of customers presents high costs since they entail valuable maintenance expenses, water transport over long distance and long networks. Concerning the second explanatory factor, the industrial water volume (m³), if not corrected for, the higher its value, the more efficient the water utility will be. Indeed, water utilities which sell more water for industrial customers present less costs per m³ provided.⁷

Figure 2 displays the model adopted while Table 2 presents the descriptive statistics of the sample. Following suggestions of water organisations, we use a non-oriented model as water utilities focus on both input reductions and output expansions. Following Daraio and Simar (2005), we select m and R as the values for which the percentage of super-efficient observations remain stable; in our application, this amounts to $m = 40$ and $R = 50$. Detailed results for other values of m and R are available upon request.

⁷As there is not a consensus in the literature about considering CAPEX and delivered water volume as fixed variables, we decided not consider these variables as fixed. Indeed, we consider CAPEX in monetary units and, as such, an increase (or reduction) is possible in this variable. Concerning the delivered water volume, we considered that the water volume has a substantial component that has some elasticity (in particular for the non-essential use and industrial/commercial use). For this reason, for example, in some countries (like the Netherlands, UK, Portugal and Belgium) the larger customers are liberalized. Therefore, in our opinion, it makes sense to consider the possibility of expansion (reduction) of this variable.

The causes of potential outlier's presence in the sample are four-fold. First, a water utility can be considered as an outlier because particular variables have exceptionally high or low values when compared with the remaining. The wider values could arise from specific features intrinsic to the water utility, for example, a high peak factor or a high cost of water treatment (due to, e.g., desalination). A second cause is linked to the measurement errors, where improper data collection, data reporting and record keeping are included. A particular source for measurement errors in the Portuguese drinking water sector are the CAPEX, since, depending on the ownership, water utilities often employ different methodologies to measure capital outlays. Third, due to the different structure in ownership (e.g., the data set contains both private and public owned utilities) some entities could be atypical and influence the inefficiency scores of the other utilities. Finally, a fourth source of outliers is related to those water utilities which bring to light the best or worst practices. The latter, despite being the best practices and theoretically useful to benchmark, make part of the efficient frontier influencing the results of the other companies and so require special attention.

4.2 Detecting influential observations

Applying the suggested model to the data set (i.e., estimating the robust directional distance functions for each of the complementary outlier detection methods) results in a range of potential outlying observations. Table 3 summarizes the ten water utilities with higher likelihood of being outliers according to the five complementary methodologies.

The five procedures detected 25 different potential atypical observations, from which 14 are found by at least two different procedures (i.e., our criterion to select an observation as outlier). We briefly discuss these 14 outliers next and look into the atypical observations more closely. Pombal has a very penalizing environment (due to a very low customer density) while Loulé operates in an extremely favorable environment (thanks to many industrial customers). Lousada and VilaF depict an unusually high capital input with a low bad output (i.e., thanks to a relatively flat relief, the pressure in the mains can be lower such that there are lower losses). Lisboa, Sintra, O/A and Gaia are by far the largest water utilities of the country (and, as such, atypical). Paços Ferreira, subject to a recent privatization, presents an excess of inputs for the outputs produced in a favourable environment. C Branco provides also the subsidized solid waste service so that the quality of data can not be assured. Feira presents one of the most critical circumstances in Portugal concerning the water service coverage which has led to its privatization a few years ago. Finally, a set of more surprising outliers consists of Cascais, Viseu, Coimbra and Vila Franca, which comprise a combination of various variables with very high values, although no single variable presents an extreme value. In the following, to obtain more accurate inefficiency estimates, we remove the atypical observations from the sample, such that we obtain a data set of 59 water utilities.

4.3 Efficiency estimates

The average robust non-oriented conditional inefficiency estimate of the Portuguese drinking water utilities in a sample without the influential observations amounts to 0.0414 (with a standard deviation of 0.0047). This indicates that, on average, the utilities could simultaneously decrease their inputs and undesired outputs and increase their desired outputs by 4.1% if they would produce as efficiently as their references (which have, thanks to the conditioning, similar environmental characteristics). Although conditional non-oriented models have to our best knowledge never been applied to the Portuguese drinking water sector, these results are in line with previous research (for an overview, e.g., Marques and Silva 2006).

Table 3 Influential observations

Leverage			Super-efficiency		Simar			Peer count	Peers
			value	std. dev					
Pombal	0.0357	Pombal	-0.7201	0.0029	Pombal	-0.1410	0.0524	Lousada	13
Loulé	0.0342	Loulé	-0.634	0.0036	Loulé	-0.0953	0.0056	Loulé	13
O/A	0.0216	VilaF	-0.4035	0.0000	P. Ferreira	-0.0423	0.0117	Viseu	9
Sintra	0.0204	Batalha	-0.2699	0.0460	Coimbra	-0.0415	0.0203	Valongo	9
Lousada	0.0203	Lousada	-0.2618	0.0426	Lousada	-0.0345	0.0122	Feira	8
Cascais	0.0202	Cantanhede	-0.2284	0.0023	Feira	-0.0270	0.0072	Pombal	8
Lisboa	0.0094	Sintra	-0.1662	0.0000	Sintra	-0.0265	0.0023	CBranco	7
VilaF	0.0093	P. Ferreira	-0.1485	0.0230	Cascais	-0.0229	0.0088	Cascais	6
P. Ferreira	0.0090	Coimbra	-0.1347	0.0231	Gaia	-0.0209	0.0150	Matosinhos	6
Lagos	0.0090	Cascais	-0.1336	0.0104	Viseu	-0.0182	0.0068	Braga	6

Peer index									
input1		input2		output1		output2		undesired	
Carrazeda	0.0438	CBranco	0.0377	Carrazeda	0.0400	O/A	0.0409	O/A	0.0625
O/A	0.0404	O/A	0.0366	O/A	0.0377	Carrazeda	0.0330	Carrazeda	0.0305
Sintra	0.0221	Carrazeda	0.0358	Setubal	0.0266	Setubal	0.0267	CBranco	0.0261
PontaD	0.0216	Loulé	0.0203	Sintra	0.0216	Sintra	0.0240	Pombal	0.0256
VianaC	0.0211	Gaia	0.0199	Loulé	0.0215	CBranco	0.0234	Loulé	0.0213
Viseu	0.0208	Sintra	0.0186	CBranco	0.0210	Viseu	0.0218	Espinho	0.0196
Setubal	0.0207	Viseu	0.0175	Espinho	0.0201	Espinho	0.0216	Sintra	0.0191
Loulé	0.0089	Lousada	0.0172	Viseu	0.0198	Loulé	0.0202	Viseu	0.0187
VilaF	0.0000	VilaF	0.0000	VilaF	0.0000	VilaF	0.0000	VilaF	0.0000
Lisboa	0.0000	Lisboa	0.0000	Lisboa	0.0000	Lisboa	0.0000	Lisboa	0.0000

Nevertheless, in past research, the inefficiency scores were lower since (1) the outliers were not eliminated, (2) only input or output efficiency was considered, and (3) the models had more degrees of freedom. In addition, there is consistency with the extreme values (worst and best practices). Due to the robust order- m approach applied, some observations are labeled as super-efficient (e.g., the best performing observation obtains an inefficiency score of -0.0795 with a Monte-Carlo standard deviation of 0.0247) which indicates that, e.g., for this super-efficient observation, it performs 7.95% more efficient in both input and output dimensions than the average of m reference observations randomly drawn from the population of utilities with the same characteristics. Detailed results for individual utilities are available from the authors upon request.

5 Conclusion

This paper proposes a nonparametric model to measure the efficiency of utilities for which the production technology is unknown. In particular, it develops for the Portuguese drinking water sector, for which we have an unusually rich data set, a non-oriented model which accounts for undesired outputs and exogenous influences in the sample. These features make the model attractive for real-world applications as firms' managers simultaneously take into

account input reductions and output expansions. The simulated example pointed out that, although our model gives the benefit of the doubt to atypical observations of not being influential, each outlier detection procedure has its own peculiarity and, hence, caution should be taken if one or another procedure is selected. To exploit the singularity of the leverage and the peer count (i.e., attention to the influence on other observations), the super-efficiency and the order- m method (i.e., stress the atypical inefficiency scores) and the peer index (i.e., importance to the individual variables of the evaluated entity), it is proposed to select these observations as outliers which simultaneously are revealed as atypical by at least two of the procedures. The simulated example demonstrates the usefulness of this approach.

Applying the model to the Portuguese drinking water utilities, we detected and described the atypical observations. After deleting these influential units from the sample, the robust non-oriented conditional inefficiency estimate of the Portuguese drinking water utilities amounts, on average, to 4.1% which indicates that the observations could simultaneously increase the inputs and decrease the outputs by 4.1% if producing as efficient as the benchmarks.

Acknowledgements We would like to thank Laurens Cherchye, Elbert Dijkgraaf, Wim Moesen, Erwin Ooghe, Sergio Perelman, Emmanuel Thanassoulis and two anonymous referees for their very useful comments. In addition, we want to thank Gary Simpson, Tony Flegg and other seminar participants at the OR49 conference in Edinburgh for insightful discussions. Kristof De Witte is grateful to his doctoral committee for additional suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Table 4 Survey of outlier detection procedures

Study	Method	Comments
Wilson (1993)	Statistical test to check if the change in efficiency due to the deletion of i observations from the sample is relevant.	Computational burden and difficulties in considering the masking problem.
Wilson (1995)	Descriptive statistic method to prioritize observations in the efficient subset for further scrutiny considering different factors.	Necessary to know the production process to decide when to stop the scrutiny.
Pastor et al. (1999)	Statistical test to check if the change in efficiency due to the deletion of i observations from the sample is relevant.	Propose a sign-test approach to examine influential observations.
Fox et al. (2004)	Determination of a measure of absolute dissimilarity based on the additive aggregation of a separate mix and scale measure of dissimilarity.	Treat observations (efficient and non-efficient) symmetrically.
Jahanshahloo et al. (2004)	The influence measure of an observation on the efficiency of the inefficient observations is specified by using a half-line and a simple formulation.	Only developed for radial models.

Table 4 (Continued)

Study	Method	Comments
Chen and Johnson (2006)	Unified model for detecting outliers by examining their effect on the boundaries of the convex hull constructed from a data set.	Only focuses on the ranking of outliers but not on any other decisions, such as whether or not to remove them from the data set.
Banker and Chang (2006)	Employ the Banker and Gifford super-efficiency method to identify outliers.	Specially applied when data is contaminated.
Johnson and McGinnis (2006)	Use a methodology considering a iterative procedure using a hyperbolic distance function.	Outliers are determined based on an efficient and an inefficient frontier.
Mahlberg and Raveh (2007)	Employ multi-dimensional scaling approach (co-plot) methodology.	Co-plot is a graphical help to explain the DEA results to users identifying the atypical observations.

References

- Aida, K., Cooper, W., Pastor-Ciurana, J., & Sueyoshi, T. (1998). Evaluating water supply services in Japan with RAM: a range-adjusted measure of inefficiency. *Omega International Journal of Management Science*, 26(2), 207–232.
- Andersen, P., & Petersen, N. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), 1261–1264.
- Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21(2), 358–389.
- Banker, R., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research*, 175(2), 1311–1320.
- Brockett, P., Rousseau, J., Wang, Y., & Zhou, L. (1997). *Implementation of DEA models using GAMS* (Research Report 765). University of Texas, Austin.
- Cazals, C., Florens, J., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106, 1–25.
- Chambers, R., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, 98, 351–364.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 428–449.
- Charnes, A., Cooper, W., Golany, B., Seiford, L., & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1), 91–107.
- Chen, W., & Johnson, A. (2006). Detecting efficient and inefficient outliers in data envelopment analysis. *Working Papers Series*. Available at SSRN.
- Cherchye, L., Lovell, C. A. K., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, 51, 749–779.
- Cooper, W., Park, K., & Pastor, J. (1999). RAM: a range adjusted measure of efficiency for use with additive models, and relations to other models and measures in DEA. *Journal of Productivity Analysis*, 11, 5–42.
- Chung, Y., Fare, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management*, 51(3), 229–240.
- Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24, 93–121.
- Daraio, C., & Simar, L. (2007). Advanced robust and nonparametric methods in efficiency analysis. In *Series: studies in productivity and efficiency*. Berlin: Springer.
- Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: a multivariate conditional quantile approach. *Journal of Econometrics*, 140(2), 375–400.
- Daouia, A., & Ruiz-Gazen, A. (2006). Robust nonparametric frontier estimators: influence function and qualitative robustness. *Statistica Sinica*, 16(4), 1233–1254.
- Depriens, D., Simar, L., & Tulkens, H. (1984). Measuring labor efficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: concepts and measurements* (pp. 243–267). Amsterdam: North-Holland.

- Despotis, D. (2005). A reassessment of the human development index via data envelopment analysis. *Journal of the Operations Research Society*, 56, 969–980.
- De Witte, K., & Marques, R. (2007). Designing incentives to local public utilities, an international comparison to the drinking water sector. *CES Discussion Paper Series DPS* 07.32.
- De Witte, K., & Korteleinen, M. (2008). Blaming the exogenous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *CES Discussion Paper Series DPS* 08.33.
- Färe, R., Grosskopf, S., & Lovell, K. A. (1985). *The measurement of efficiency of production*. Boston: Kluwer.
- Fox, K., Hill, R., & Diewert, E. (2004). Identifying outliers in multi-output models. *Journal of Productivity Analysis*, 22(12), 73–94.
- Fried, H., Lovell, C. A. K., & Schmidt, S. (2008). *The measurement of productive efficiency and productivity growth*. London: Oxford University Press, 638 p.
- Jahanshahloo, G., Hosseinzadeh, F., Shoja, G., Tohidi, G., & Razavyan, S. (2004). A method for detecting influential observation in radial DEA models. *Applied Mathematics and Computation*, 147(2), 415–421.
- Johnson, A., & McGinnis, L. (2006). *An outlier detection methodology with consideration for an inefficient frontier*. Paper presented at the NAPW IV, June 2006, New York.
- Lambert, D., Dichev, D., & Raffiee, K. (1993). Ownership and sources of inefficiency in the provision of water services. *Water Resources Research*, 29(6), 1573–1578.
- Mahlberg, B., & Raveh, A. (2007). *Co-plot: a useful tool to detect outliers in DEA*. Paper presented at the EWEPX X, June 2007, Lille.
- Marques, R., & Silva, D. (2006). Statistical inference of efficiency estimators obtained with the DEA nonparametric frontier technique. A bootstrap methodology. *Portuguese Operational Research Journal*, 26(2), 89–110.
- Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public Economics Research Paper* 17. KU Leuven.
- Pastor, J., Ruiz, J., & Sirvent, I. (1999). A statistical test for detecting influential observations in DEA. *European Journal of Operational Research*, 115(3), 542–554.
- Portela, M., Borges, P., & Thanassoulis, E. (2003). Finding closest targets in non-oriented DEA models: the case of convex and non-convex technologies. *Journal of Productivity Analysis*, 19(23), 251–269.
- Ray, S., & Mukherjee, K. (2007). Efficiency in managing the environment and the opportunity cost of pollution abatement. *Working papers* 2007-09, University of Connecticut, Department of Economics.
- Simar, L. (2003). Detecting outliers in frontier models: a simple approach. *Journal of Productivity Analysis*, 20, 391–424.
- Sousa, M., & Stocic, B. (2005). Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurement of outliers. *Journal of Productivity Analysis*, 24, 157–181.
- Torgersen, A., Førsund, F., & Kittelsen, S. (1996). Slack-adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis*, 7(4), 379–398.
- Wilson, P. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics*, 11(3), 319–323.
- Wilson, P. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, 6(1), 27–45.
- Yin, R.K. (2003). *Applications of case study research*. Thousand Oaks: Sage