Highlights

A regulatory robust conditional approach to measuring the efficiency of wholesale water supply and wastewater treatment services

Miguel Alves Pereira, Hermilio Vilarinho, Giovanna D'Inverno, Ana Santos Camanho

- Construction of an unprecedented collaborative regulatory framework.
- Benchmarking of water supply and wastewater treatment services.
- Efficiency measurement of wholesale operators in the 2017-2021 period.
- Extension of Data Envelopment Analysis to robust and conditional approach.
- Estimated 2%-3% potential savings for water supply and wastewater treatment services.

A regulatory robust conditional approach to measuring the efficiency of wholesale water supply and wastewater treatment services

Miguel Alves Pereira[a,b,*], Hermilio Vilarinho[a], Giovanna D'Inverno[c], Ana Santos Camanho[a]

[a] *Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal*
[b] *CEGIST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisboa, Portugal*
[c] *Dipartimento di Economia e Management, Università di Pisa, Via Cosimo Ridolfi 10, 56124, Pisa, Italia*

Abstract

A benchmarking evaluation instrument was designed with a European Union country regulatory authority for water supply and wastewater treatment services to determine the efficient operating cost of its service providers that operated in the wholesale market segment in the 2017-2021 period. To this end, the non-parametric Data Envelopment Analysis technique was adapted to a robust and conditional approach. The results point to similar mean efficiency scores between water supply and wastewater services in the five-year period, despite the greater heterogeneity in the latter. Furthermore, the estimated potential cost savings for both services ranged from about 2% to 3%.

Keywords: Data Envelopment Analysis, Water supply, Wastewater treatment.

1. Introduction

Over the last few years, debates about some of the major challenges of the twenty-first century have intensified, from pandemics to wars, population growth to poverty, and climate change to energy crises. However, an issue at the centre of these discussions is often forgotten: the scarcity of water resources. According to the most recent report of the United Nations (UN; 2022), the world's water-related ecosystems are being degraded at an alarming rate, with more than 85% of the Earth's wetlands being lost in the past 300 years. Besides, over 700 million people live in countries with high and critical water stress levels. In the end, the UN predicts that, at the current rate, 1.6 billion people will lack safely managed drinking water by 2030 and a fourfold increase in the pace of progress will be necessary to meet water supply and sanitation targets, despite the progressive convergence of its Member States towards them (Pereira & Marques, 2021, 2022a).

If we narrow our scope to the European context, we encounter a set of policies and strategies established by the European Commission to halt deterioration in European water bodies and improve their status. First, the Water Framework Directive (2000/60/EC) legislates the quality

*Corresponding author
Email address: miguel.alves.pereira@fe.up.pt (Miguel Alves Pereira)

and quantity of groundwater and the quality of surface water. Most Central and Northern European countries have reported their River Basin Management Plans, while most Southern and Eastern European countries are still up to the public consultation stage. Second, the European Green Deal intends to guide the efficient use of resources in the sustainable circular economies of the future, resting on the Circular Economy Action Plan aimed at reducing the pressure on natural resources -, with clean water as one of its primary goals.

However, the scenario for the water sector can be revamped by doing better with less. The Organisation for Economic Co-operation and Development (OECD; 2015) stated that this sector is highly aligned with the features of multi-level governance in that water connects all sectors, places, and people, and its management is both a global and local concern. The monopolistic nature of this sector (and associated market failures) turns policy-making into an intricate task. Indeed, the World Bank (1992) had already aligned a country's quality of governance - how power is utilised in managing a country's resources - with its level of development.

Furthermore, the OECD (2015) has also asserted that "water crises are often primarily governance crises". Despite lacking evidence supporting a one-size-fits-all solution to address global water challenges, a highly context-dependent governance framework, enhanced with bottom-up and inclusive decision-making in designing effective water policies, is key to overcoming bottlenecks and solving current and future water challenges. Essentially, the robustness of future public policies should rely on the dimensions of water governance put forward by the OECD (2015): effectiveness (concerning governance's contribution to set, implement, and meet targets), efficiency (concerning governance's contribution to maximise user sustainable water value for money), and trust and engagement (concerning governance's contribution to building public confidence and democratically involve all stakeholders).

The OECD's focus is more comprehensive than the user satisfaction-based effectiveness dimension (Vilanova et al., 2015). From an economic standpoint, it seeks to minimise resource consumption to produce the outcomes expected by the users. Fundamentally, it rests on four principles: data and information, financing, innovative governance, and regulatory frameworks. In particular, the latter is seen as a critical principle since regulatory authorities play a major role in supervising operators, monitoring all areas of water-related services, and deploying policies to balance the needs and expectations of the stakeholders (Akhmouch & Correia, 2016). Regulation is vital to control the operators' market position, quality of service, and prices - something that can be accomplished via benchmarking (Pereira & Marques, 2022b). Benchmarking actions are recognised as an essential tool to promote efficiency improvements in the water sector (Henriques et al., 2020). Regardless of the type of regulatory model, the literature has shown that benchmarking not only empowers the regulatory authority in guiding decision-making but also introduces artificial competition in naturally monopolistic sectors and, consequently, incentivises improvements (Pinto et al., 2017b).

Nonetheless, Mehta et al. (2013) claim that the full potential of benchmarking tools in regulated sectors, such as urban water supply and sanitation, is yet to be achieved. Therefore, a systematic effort is needed to develop efficiency measurement tools to monitor water supply and wastewater treatment operators conducive to improving their price-quality relationship, quality of service, and system sustainability in the long term.

In this study, we conceive a comparative evaluation instrument to measure the efficiency of water supply and wastewater treatment service providers operating in the wholesale market segment in a European Union (EU) country in collaboration with its water and waste services regulatory authority. The result of this collaboration is to ascertain the efficient operating expenditure (OPEX) of each operator for the 2017-2021 period and support budget drafting for the next regulatory period, with clear impacts on the country's contributions towards the aforementioned European Commission's Water Framework Directive and European Green Deal policies in terms of water quality. In particular, we employ the ubiquitous Data Envelopment Analysis (DEA) non-parametric method, extended to a conditional and robust order-m setting to mitigate the impact of atypical operators and understand the influence of exogenous factors on the operational activity of the service providers. This way, our proposal is aligned with the state of the art of scientific literature on non-parametric efficiency measurement and uses a conditional and robust order-m approach as a benchmarking tool aimed at improving operational practices through peer learning. This work is also innovative since it comprehends a collaborative empirical application in the water supply and wastewater treatment wholesale market segment. The approach proposed in this study is innovative in terms of regulation in the European space, placing this country as a pioneer in terms of formative regulation for wholesale operations. It is tailored to promote continuous enhancement in the sector by providing regulatory authorities with tools that allow them to define improvement objectives based on comparisons with the best practices observed in other entities, taking into account the context in which they operate.

This paper is structured as follows: Section 2 addresses the literature reviewed in the pursuit of the knowledge gap and identifies the aspects that differentiate this study from previous works; Section 3 details the methodology proposed for the efficiency analysis; Section 4 describes the case study built alongside the regulatory authority; Section 5 presents the results and discusses their regulatory and decision-making implications; Section 6 highlights the main achievements, limitations, and research prospects of the study.

2. Knowledge gap

There are numerous scientific publications regarding efficiency in the water sector, ranging from articles (see, e.g., Gidion et al., 2019; Fu & Jacobs, 2022) to book chapters (see, e.g., Davis, 2005) and reviews (see, e.g., Vilanova et al., 2015; Santos et al., 2019) to conference proceedings (see, e.g., Vieira et al., 2015; Dziedzic & Karney, 2014). These studies, and many others in the literature, cover different perspectives in which measuring the efficiency of services provided in this sector is included. When it comes to studies on the efficiency of water supply and wastewater treatment services, the paradigm is quite different, especially if we consider the two segments of this market - wholesale and retail. First, focusing on the wholesale market segment was a request from the country's regulatory authority, which intended to begin an analytical endeavour upstream regarding the water supply and wastewater treatment value chain before delving into any regulatory market changes. Second, addressing potential barriers wholesale market operators create, in light of the sustainability issues raised in Section 1, is crucial for efficient governance, especially at the local government level (Caplan et al., 2022). Third, to the best of the authors' knowledge, there is only one study on the water supply service (Lo Storto, 2013) and two studies on the wastewater treatment service (Carvalho & Marques, 2014; Henriques et al., 2020) that partially meet the wholesale market segment requirement - none of which have provided clear evidence for inefficiencies in the country's wholesale market. It should be noted that the number of studies in this area on the retail market segment is vastly broader, with much more detailed insights into the country, although there is no concrete focus on this niche of the literature in the analysis at the level of companies, municipalities, regions, or even other countries.

Indeed, on the one hand, Lo Storto (2013) measured the efficiency of 53 wholesale water suppliers in Italy (which supply water to 42% of Italian municipalities) and evaluated the influence of contextual factors on their efficiency. Hence, the author employed a two-stage DEA approach. A traditional input-oriented variable returns-to-scale (VRS) DEA model in its multiplier formulation was used in the first stage. This choice was "justified by the great variance of operators, the size of service provided, and the goal of the analysis typical of this kind of study which is generally oriented towards cost reduction, as the demand that operators" face "remains almost stable" (Lo Storto, 2013). The model considered the *Aqueduct network length*, the *Sewerage network length*, and the *Total production cost* as inputs and the *Revenue from service delivered* as a single output. In the second stage, a bootstrapped DEA model and a Tobit regression were used, considering the *Number of municipalities*, the *Number of connections*, the *Number of inhabitants*, the ratio of the *Number of connections* to the *Total network length*, and the ratio of the *Number of connections* to the *Number of municipalities* as contextual continuous factors and *Geography* and *Ownership* as contextual dichotomic factors. The results revealed several inefficiencies of wholesale water suppliers and a statistical significance of the two contextual dichotomic factors.

On the other hand, both Carvalho & Marques (2014) and Henriques et al. (2020) conducted a benchmarking study of the wastewater treatment services operating in the wholesale market segment in Portugal. Nevertheless, while the former also included water supply operators and the retail market segment, the latter focused exclusively on wastewater treatment operators in the wholesale and retail market segments. First, Carvalho & Marques (2014) studied the existence of economies of vertical integration between the two market segments, economies of scope between water supply and wastewater treatment services, and economies of scale in the wholesale market segment using robust conditional order-α DEA. The authors considered a total sample of 74 operators between 2002 and 2008 and evaluated them according to 3 different models (depending on the type of economies under assessment), always considering *Labour costs*, *Capital costs*, and *Other operational costs* as inputs and a mix of volumes as outputs (in particular, regarding wastewater treatment, they have considered the *Volume of collected wastewater* and the *Volume of treated wastewater*). Ultimately, the authors found evidence of economies of scale in wastewater treatment services operating exclusively in the wholesale market segment. Second, Henriques et al. (2020) proposed a benchmarking framework to support performance-based sunshine regulation in wastewater treatment services. Using DEA's 'Benefit-of-the-Doubt' (BoD) approach, formulated with a directional distance function and incorporating weight restrictions, the authors assessed the performance of a total of 212 wastewater treatment retailers and wholesalers in Portugal in 2018, considering the three dimensions (user interface suitability, service management sustainability, and environmental sustainability) and fourteen indicators proposed by the Portuguese regulatory authority for Water and Waste Services (Cardoso et al., 2019) to evaluate the quality of service provided by wastewater treatment retailers and wholesalers. Their framework also included a second-stage contextual analysis. At last, the results of this study pointed to an exemplary level of performance in 6 of the 12 operators of wastewater treatment services that operate in the wholesale market segment in the three considered dimensions but did not find evidence of exogenous variables capable of explaining the dispersion observed in the levels of inefficiency associated with the remaining operators. However, only 3 of the 200 operators of wastewater treatment services operating in the retail market segment achieved notable results along the three dimensions, even though, in this case, there is a positive impact on the quality of service by a larger scale, investment subsidies, and energy production as well as concessions and urbanisation.

The three studies described above address different aspects of the water supply and wastewater treatment sector in the wholesale market segment but share some similarities. Table 1 encapsulates their main features in terms of application context and model structure.

Table 1: Overview of the application context and model structure of studies on efficiency measurement on wholesale water supply and wastewater treatment services.

Reference	Application context			Model structure			
	Country	Sample	Year(s)	Methodology	Indicators		Contextual variables
					Inputs	Output(s)	
Lo Storto (2013)	Italy	53 wholesale water supply services	2009	Two-stage DEA approach: input-oriented VRS DEA + bootstrapped DEA	*Aqueduct network length* + *Sewerage network length* + *Total production cost*	*Revenue from service delivered*	*Number of municipalities* + *Number of connections* + *Population* + Ratio of the *Number of connections* to the *Total network length* + Ratio of the *Number of connections* to the *Number of municipalities* + *Geography* + *Ownership*
Carvalho & Marques (2014)	Portugal	74 wholesale and retail water supply and wastewater treatment services	2002-2008	Robust conditional order-α DEA	*Labour costs* + *Capital costs* + *Other operational costs*	*Volume of delivered water (retail)* + *Volume of delivered water (wholesale)* + *Volume of collected wastewater* + *Volume of treated wastewater*	-
Henriques et al. (2020)	Portugal	12 wholesale wastewater treatment services + 200 retail wastewater treatment services	2018	Two stage DEA approach: directional BoD + hypothesis tests	Dummy variable (unitary input)	*Physical accessibility of the service* + *Economic accessibility of the service* + *Occurrence of floods* + *Response to complaints and suggestions* + *Coverage of expenses* + *Subscription to the service* + *Rehabilitation of collectors* + *Occurrence of structural collapses in collectors* + *Adequacy of human resources* + *Energy efficiency of lifting installations* + *Physical accessibility to treatment* + *Control of emergency discharges* + *Compliance with the discharge license* + *Adequate forwarding of treatment sludge*	*Management model* + *Typology of the intervention area* + *Collected wastewater* + *Own energy production* + *Investment subsidies*

If we extend the scope of our search to the retail market segment, the outcomes are quite different. There is a myriad of studies in several countries, namely: Ananda (2014) in Australia, Tourinho et al. (2021, 2022) and Pereira & Marques (2022c) in Brazil, Maziotis et al. (2020) and Molinos-Senante et al. (2020) in Chile, Romano et al. (2018) in Italy, Satoh (2015) and Satoh (2019) in Japan, Ablanedo-Rosas et al. (2020) and Salazar-Adams (2021) in Mexico, Carvalho & Marques (2011) and Pinto et al. (2017a) in Portugal, Molinos-Senante & Maziotis (2018) and Williams et al. (2020) in England and Wales, Ferreira da Cruz et al. (2012) in Italy and Portugal, De Witte & Marques (2010a) in Australia, Belgium, England, the Netherlands, Portugal, and Wales, and Ferro & Romero (2011) in Latin America.

Bottom line, similarly to the previously reported publications on the wholesale market segment, these studies tend to use: labour, capital, and operational costs as inputs; volumes of delivered water and collected and treated wastewater as outputs; and geography and ownership as contextual variables (Tourinho et al., 2022). The reader interested in water utility benchmarking is directed to the survey of Berg & Marques (2011) and the bibliometric analysis of Goh & See (2021) for further information on the subject.

Note that the vast majority of the studies mentioned above use some form of DEA (mainly the same approach as Lo Storto (2013)), with only Ferro & Romero (2011) (which also used DEA), Molinos-Senante & Maziotis (2018), Molinos-Senante et al. (2020), and Williams et al. (2020) using econometrics resting on the Stochastic Frontier Analysis (SFA) approach. Regarding robust conditional DEA, the order-m approach is more popular than its order-α counterpart, with only applications of the former being found in the works of De Witte & Marques (2010a), Carvalho & Marques (2011), and Pinto et al. (2017a). However, while De Witte & Marques (2010a) conducted an international benchmarking study to design performance incentives for water utilities, both Carvalho & Marques (2011) and Pinto et al. (2017a) attempted to understand the influence of the operational environment on the Portuguese water utilities.

In the end, as far as we know, no studies apply robust conditional order-m DEA to build a regulatory framework to measure the efficiency of water supply and wastewater treatment services operating in the wholesale market segment (in this EU country or abroad). Furthermore, there are no publications that do so as a result of a collaboration between academia and national regulatory authorities. Hence, the contributions of our proposal are reiterated as being twofold, both in terms of the scientific innovation of the used models and their empirical application to actual data underlying the regulation of water supply and wastewater treatment companies in the context of this EU nation.

3. Methodology

With the importance of benchmarking as a vital analysis for regulation activities in the water sector having already been established in Section 1, it is time to address its methodologies. Parametric and non-parametric frontier methods to measure efficiency have been employed in the sector, ranging from SFA to Data Envelopment Analysis (DEA), respectively, and their adaptations and extensions. However, Goh & See (2021) say that DEA has become the most popular.

As a non-parametric frontier method, DEA measures the relative efficiency of a homogeneous set of decision-making units (DMUs) producing multiple outputs from multiple inputs as the radial distance from each DMU to the estimated production frontier. It returns a group of efficient DMUs, i.e., benchmarks, and a group of inefficient DMUs. DEA optimises the weighting system that enables each DMU to yield its best efficiency score. It was designed by Charnes et al. (1978) based on the concepts proposed by Farrell (1957). Its main advantage concerns the nonnecessity of specifying the functional form of its frontier *a priori*, only making some assumptions regarding the production technology (e.g., convexity, returns-to-scale).

Resting on the literature review conducted in Section 2, it is consensual that an input-oriented VRS DEA model should be adopted in efficiency measurements in this sector. Thus, its envelopment formulation, which seeks the proportional input reduction needed for a certain DMU to reach the frontier assuming distinct scale sizes among the DMUs, for the DMU under assessment (DMU_{j_0}) is shown in Model (1):

$$
\begin{aligned}
Z_I = \min \quad & \theta_{j_0} - \varepsilon \left(\sum_{i=1}^{m} s_{ij_0}^{-} + \sum_{r=1}^{s} s_{rj_0}^{+} \right) && (1)\\
\text{subject to} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_{ij_0}^{-} = \theta_{j_0} x_{ij_0}, \qquad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j y_{rj} - s_{rj_0}^{+} = y_{rj_0}, \qquad r = 1, \ldots, s \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \theta_{j_0} \text{ is free} \\
& \lambda_j,\ s_{ij_0}^{-},\ s_{rj_0}^{+} \geqslant 0, \qquad \begin{cases} j = 1, \ldots, n \\ i = 1, \ldots, m \\ r = 1, \ldots, s \end{cases} \\
& \varepsilon > 0,
\end{aligned}
$$

where: x_{ij} denotes the value of input i for DMU j; y_{rj} denotes the value of output r for DMU j; θ_{j_0} is a decision variable that denotes the radial efficiency score of DMU_{j_0}; λ_j is a decision variable that denotes the intensity variables and assumes a positive value in case a specific DMU j is a peer of DMU_{j_0}; $s_{ij_0}^{-}$ is a slack variable that denotes potential non-radial adjustments to the input levels of DMU_{j_0}; $s_{rj_0}^{+}$ is a slack variable that denotes potential non-radial adjustments to the output levels of DMU_{j_0}; and ε is a non-Archimedean infinitesimal.

After computing the optimal solution of Model (1), we obtain *fully efficient*, *weakly efficient*, and *fully inefficient* DMUs. DMU_{j_0} is fully efficient when its efficiency score, θ_{j_0}, is equal to one and all slacks, $s_{ij_0}^{-}$ for $(i = 1, \ldots, m)$ and $s_{rj_0}^{+}$ for $(r = 1, \ldots, s)$, are equal to zero. A DMU is weakly efficient when its efficiency score, θ_{j_0}, is equal to one, but at least one slack, $s_{ij_0}^{-}$ for $(i = 1, \ldots, m)$ or $s_{rj_0}^{+}$ for $(r = 1, \ldots, s)$, is positive. Finally, a DMU is fully inefficient when its efficiency score, θ_{j_0}, is lower than one. Furthermore, it is possible to compute targets for an inefficient DMU_{j_0} based on the optimal values of the decision variables of its peers (given by the symbol '*') according to Expression (2) and Expression (3):

$$
x_{ij_0}^{T} = \sum_{j=1}^{n} \lambda_j^{*} x_{ij} = \theta_{j_0}^{*} x_{ij_0} - s_{ij_0}^{*-}, \quad i = 1, \ldots, m \tag{2}
$$

$$
y_{rj_0}^{T} = \sum_{j=1}^{n} \lambda_j^{*} y_{rj} = y_{rj_0} + s_{rj_0}^{*+}, \quad r = 1, \ldots, s \tag{3}
$$

Nevertheless, DEA's deterministic nature poses a disadvantage in the face of outliers since any atypical observation belonging to the set of DMUs can shape the so-called *full frontier*. Consequently, their presence may shift that frontier and underestimate the scores of the remaining DMUs (Fusco et al., 2020). Therefore, although detecting outliers to be removed from the sample can be useful, understanding the extent to which they are the best- or worst-performing

DMUs may be of interest, especially in the case of small samples where information is crucial for decision-making (De Witte & Marques, 2010b). The proposal of *partial/robust frontiers* has been put forward in the literature to address these issues, essentially in two ways: order-m (Cazals et al., 2002; Daraio & Simar, 2005, 2007b) and order-α (Aragon et al., 2005; Daouia & Simar, 2007) methods. In particular, order-m and order-α partial frontiers differ since the former concerns a "discrete" notion and the latter a "continuous" notion of partial frontiers given the fundamental distinction between m as a function of n and α as the level of an appropriate non-standard conditional quantile frontier (Daouia & Simar, 2007). If we look at Section 2, there is a quantitative preference in the literature for order-m methods instead of order-α ones supporting the use of the former over the latter.

In particular, unlike in the full frontier estimation via Model (1), where the model was solved iteratively per DMU, partial frontier estimation solves Model (1) B times per DMU following a Monte Carlo simulation, where B is a large number. Since we are dealing with an input-oriented analysis, m DMUs are randomly drawn with replacement among those producing an output level greater than or equal to the DMU$_{j_0}$ in each B iteration. This subsampling procedure aims to mitigate the impact of outliers and compare the DMU$_{j_0}$ with less extreme peers[1]. In the end, the mean of the efficiency scores computed per b-th iteration, with $b = 1, \ldots, B$, $\theta_{j_0}^{b,m}$ is equal to the robust order-m efficiency score $\hat{\theta}_{j_0}^{m}$:

$$\hat{\theta}_{j_0}^{m} = \frac{\sum_{b=1}^{B} \theta_{j_0}^{b,m}}{B} \tag{4}$$

On the one hand, m can be seen as the number of DMUs competing with the DMU$_{j_0}$ to produce greater or equal output levels. On the other hand, it can be seen as a threshold value for the robustness analysis. The choice of m is not elementary since the literature mentions that its value should not be too high or too low because of the possibility of not enveloping all DMUs (Cazals et al., 2002; Rogge & De Jaeger, 2013). Typically, m should be lower than the number of sampled DMUs to decrease the probability of super-efficient DMUs. Henriques et al. (2022) suggest conducting a sensitivity analysis for different m values to support the robustness of the analysis.

When identifying benchmarks and targets in a robust order-m DEA context, there are some possibilities to compute them, following Henriques et al. (2022). First, the number of times a DMU is deemed as a benchmark per partial frontier indicates its benchmarking status. Second, the intensity variables and the targets can be computed as the mean of all B iterations. Since a partial frontier produces a tighter envelope around the sampled data, the generated robust efficiency scores will always be higher than those computed from a full frontier estimation. Thus, the robust target values will be lower and more realistically achievable.

Moreover, understanding the influence of the operational context is of utmost importance in these types of analyses. Ascertaining the role of specific factors surrounding the sampled DMUs is usually recognised by the literature as a relevant aspect, typically addressed via a *one-stage approach* (where contextual variables are classified *a priori* as either inputs or outputs and included in the model or used to guide the sampling procedure) or a *two-stage approach* (where efficiency scores are computed in a first stage and parametrically regressed on non-discretionary

[1]Nonetheless, it may lead to the absence of DMU$_{j_0}$ among the subsampled m DMUs. Hence, it may be located above the partial frontier and be identified as a *super-efficient* DMU if its efficiency score is greater than one.

variables in a second stage) (Daraio & Simar, 2007b). However, depending on the type of contextual variable, a mixed approach can be used, for instance, considering continuous/*modelling* contextual variables in the first stage and categorical/*descriptive* contextual variables in the second stage. Still, according to those authors, there are two issues concerning the bias of first-stage efficiency scores and the need for specifying the parametric regression model *a priori*. For this reason, Daraio & Simar (2005) proposed conditional non-parametric frontier models, whose insights were integrated into the robust conditional DEA employed here, also in line with Daraio & Simar (2007b) and De Witte & Kortelainen (2013).

Compared to the robust order-m DEA described above, the robust conditional order-m DEA demands an adjustment to estimate the unconditional frontier. This way, there is a higher probability of DMUs that operate in a similar context being drawn together and a lower probability of DMUs that operate in a different context being drawn together. Accordingly, the mean of the conditional efficiency scores computed per b-th iteration, with $b = 1, \ldots, B$, $\theta_{j_0}^{b,m,z}$ is equal to the robust conditional order-m efficiency score $\hat{\theta}_{j_0}^{m,z}$:

$$\hat{\theta}_{j_0}^{m,z} = \frac{\sum_{b=1}^{B} \theta_{j_0}^{b,m,z}}{B} \tag{5}$$

Note that R version 4.2.1 was the software used to implement the models above and compute the results. In particular, the ROBUST˙DEA and CONDITIONAL˙DEA of the RCDEA package were employed with slight code modifications to enable the benchmark computation.

4. Case study

This section covers the application context addressed in this study (Subsection 4.1) and the modelling structure used to address it (Subsection 4.2).

4.1. Application context

The challenges faced by the water sector and the goals set to tackle them are two key facets of regulation (Henriques et al., 2020). This case considers an unidentified EU country in which the regulatory authority developed an integrated approach comprising two perspectives: structural regulation and behavioural regulation. Although the former concerns organisational aspects and the latter concerns each utility, both standpoints should interact and evolve (Baptista, 2014). Essentially, the sunshine regulatory model adopted by the regulatory authority is portrayed as collaborative rather than restrictive. It uses a set of transparent key performance indicators to evaluate the service quality of operators according to three dimensions - user interface suitability, service management sustainability, and environmental sustainability - and, ultimately, enable regulation by benchmarking to impact the operators' performances and promote accountability, thus assuming a developmental role in the sector instead of a monitoring and control part.

The market structure of the water sector in that nation is divided into wholesale and retail market segments, with the latter being less developed than the former from service quality, resource management, and sustainability standpoints (Costa et al., 2021). As justified above, this study focuses on wholesale water supply and wastewater treatment services, which currently comprehend 17 and 12 operators covering 72% and 96% of the country's population, respectively. Note that three entities operate simultaneously as both wholesalers and retailers (Costa et al., 2021), although their data sets are completely separate, including cost information.

Bottom line, the water supply or wastewater treatment service providers analysed in this paper are studied in the 2017-2021 period as pooled five-year samples but must remain anonymous due to confidentiality issues. In this work, only 10 of the 17 water supply service providers operating in the wholesale market segment were considered since the remaining 7 provide the service in very restricted areas. Thus, 50 observations populated all water supply (WS) models, resulting from 10 observations per year, i.e., five observations per operator. As for the wastewater treatment service, six instances were removed from the sample. This deletion was due to reasons related to significant changes in the production technology of these service providers in the years considered according to the EU country's regulatory authority. Hence, 54 observations (due to the removal of 6 observations from the initial 60 observations, which resulted from 12 observations per year, i.e., five observations per operator) populated all wastewater treatment (WT) models. Additionally, there are six common operators between the ten water supply and 12 wastewater treatment sets, i.e., 16 distinct operators in total. In the end, 7 out of the ten sampled wholesale water supply operators and 10 out of the 12 sampled wholesale wastewater treatment operators are managed under a municipal or multi-municipal concession agreement, while 3 of 10 and 1 of 12 abide by a delegated State or municipal management solution, and municipal or inter-municipal services or associations directly manage 1 of 12.

4.2. Modelling structure

The assessment carried out in this paper required the collaborative construction with the regulatory authority of the production activities of the WS and WT service providers operating in the wholesale market segment. This way, two efficiency measurement models - a robust unconditional (RU) one and a robust conditional (RC) one - were defined per type of service in the wholesale market segment, depending on whether or not the modelling contextual variable was included in the RC model: WS-RC Model and WT-RC Model if the modelling contextual variable was considered in the RC model; WS-RU Model and WT-RU Model if the modelling contextual variable is not included in the RU model (see Table 2).

Essentially, the rationale behind the production process shown in Table 2 corresponds to the context of the physical configuration of the wholesale water supply and wastewater treatment systems according to the regulatory authority. In terms of the total operating costs, all the water that enters the system through an elevating process and serves retailers through pipelines is considered. Regarding the total operating costs, all wastewater collected from retailers through collectors and raised to be subjected to treatment is considered. It should be noted that an elevating process is for the (waste)water to circulate under pressure and enable it to overcome terrain barriers.

From another angle, following the recommendation of Henriques et al. (2022), we have chosen values of m equal to the number of DMUs in the samples ($m = 50$ for WS Models and $m = 54$ for WT Models) due to the small sample size of our study. This choice is supported by Daraio & Simar (2007a) since the authors state that even if m is independent of the sample size, its values can be fixed by taking into consideration the possible number of competitors of a given firm, which, in a market with a small number of utilities - as is the wholesale water supply and wastewater treatment one in this EU country -, it is sensible to consider all of them as potential competitors. $B = 1000$ was also chosen as the appropriate number of iterations for each model since it requires less computational effort. Running a sensitivity analysis on B did not generate changes, especially for larger values.

Table 2: Modelling structure of WS Models and WT Models.

Model	Indicators		Contextual variables	
	Input	Outputs	Modelling	Descriptive
WS-RC	*Total OPEX* (x_1^{WS})	*Volume of water entering the system* (y_1^{WS}) + *Elevated volume of water* (y_2^{WS}) + *Number of households with effective water supply service* (y_3^{WS}) + *Total length of pipelines* (y_4^{WS})	*Raw water quality* (z_1^{WS})	*Management model* (z_2^{WS}) + *Typology of the intervention area* (z_3^{WS})
WS-RU	*Total OPEX* (x_1^{WS})	*Volume of water entering the system* (y_1^{WS}) + *Elevated volume of water* (y_2^{WS}) + *Number of households with effective water supply service* (y_3^{WS}) + *Total length of pipelines* (y_4^{WS})	-	*Management model* (z_2^{WS}) + *Typology of the intervention area* (z_3^{WS})
WT-RC	*Total OPEX* (x_1^{WT})	*Volume of wastewater treated in treatment plants* (y_1^{WT}) + *Elevated volume of wastewater* (y_2^{WT}) + *Number of households with effective wastewater treatment service* (y_3^{WT}) + *Total length of collectors* (y_4^{WT})	*Effluent quality* (z_1^{WT})	*Management model* (z_2^{WT}) + *Typology of the intervention area* (z_3^{WT})
WT-RU	*Total OPEX* (x_1^{WT})	*Volume of wastewater treated in treatment plants* (y_1^{WT}) + *Elevated volume of wastewater* (y_2^{WT}) + *Number of households with effective wastewater treatment service* (y_3^{WT}) + *Total length of collectors* (y_4^{WT})	-	*Management model* (z_2^{WT}) + *Typology of the intervention area* (z_3^{WT})

4.2.1. Input

It is important to note that the operating costs chosen as the input of each model include the *Cost of goods sold and materials consumed*, the *Cost of external supply and services*, the *Cost of labour*, and *Other operating costs*. The regulatory authority considered these four types of cost to be fundamental for the operational cost structure of wholesale service providers and, consequently, for the definition of OPEX. In other words, OPEX resulted from the sum of the *Cost of goods sold and materials consumed*, the *Cost of external supply and services*, the *Cost of labour*, and *Other operating costs*. Bear in mind that it does not include structure costs.

4.2.2. Outputs

After extensive discussions with the regulatory authority about the duality of operation in the wholesale and retail market segments, it was concluded that the system characteristics of the two types of services considered here would be a basis for choosing the outputs (see Subsection 4.1). Therefore, it is essential to consider as outputs the *Volume of water entering the system*, the *Elevated volume of water*, the *Number of households with effective water supply service*, and the *Total length of pipelines* in the case of the water supply service and the *Volume of wastewater treated in treatment plants*, the *Elevated volume of wastewater*, the *Number of households with effective wastewater treatment service*, and the *Total length of collectors* in the case of the wastewater treatment service.

4.2.3. Contextual variables

The selected contextual variables reconcile the evidence found in the literature and the regulatory authority's preferences. On the one hand, the chosen descriptive contextual variables were based on the ones most commonly used in the literature, namely the *Management model* (concession, delegation, or direct management) and the *Typology of the intervention area* (predominantly rural area, moderately urban area, and predominantly urban area). In particular, regarding the former (Vilarinho et al., 2023): in a concession model, the State establishes a long-term public-private partnership with a third party to operate the system; in a delegation model, the State owns and controls the operation of the system, but delegates its management to an operator via a management contract; in a direct management model, the State owns and operates the system. Regarding the latter, its three typologies are derived from the degree of urbanisation of a territory established by the country's National Institute for Statistics. On the other hand, the modelling contextual variable boiled down to the quality of the product, depending on the model: *Raw water quality* in WS Models and *Effluent quality* in WT Models. Note that the latter was developed internally by the regulatory authority specifically for this analysis.

Briefly, the modelling contextual variable enters each robust conditional DEA model (WS-RC Model and WT-RC Model) in the optimisation process associated with the efficiency measurement, whereas the remaining contextual variables are considered only in a phase after obtaining the efficiency scores of the four models due to their descriptive nature. Thus, the descriptive contextual variables are not taken into account for the estimation of the production frontiers.

4.2.4. Overview

Finally, the descriptive statistics of all the variables used in the WS Models (Table A.11) and the WT Models (Table A.12) in the period 2017-2021 are reported in Appendix A.1. Regarding the descriptive contextual variables, given their use *a posteriori* and their invariable nature over time, their descriptive statistics are presented in the same appendix, but in Table A.13.

That being said, bivariate Pearson correlation tests were performed between all potential variables considered in the discussions with the regulatory authority to validate the choice of inputs and outputs. These varied between total OPEX, OPEX without structure costs, and structure costs for inputs and volumes, installed capacities, and socio-demographic indicators for outputs and modelling contextual variables of WS Models and WT Models in 136 correlations between 16 variables. These findings attest to the legitimacy of the modelling choices given the compliance with the isotonicity property of DEA, i.e., the requirement that the relationship between inputs and outputs is not erratic. For reasons of space, only the results concerning the selected variables are presented (see Table 3 and Table 4).

Table 3: Bivariate Pearson correlation among the variables selected for WS Models.

	$x_1^{WS\,\mathrm{a}}$	$y_1^{WS\,\mathrm{b}}$	$y_2^{WS\,\mathrm{c}}$	$y_3^{WS\,\mathrm{d}}$	$z_1^{WS\,\mathrm{e}}$	$y_4^{WS\,\mathrm{f}}$
$x_1^{WS\,\mathrm{a}}$	1	0.886^{**}	0.796^{**}	0.920^{**}	-0.591^{**}	0.652^{**}
$y_1^{WS\,\mathrm{b}}$	-	1	0.924^{**}	0.887^{**}	-0.405^{**}	0.310^{*}
$y_2^{WS\,\mathrm{c}}$	-	-	1	0.887^{**}	-0.284^{*}	0.161
$y_3^{WS\,\mathrm{d}}$	-	-	-	1	-0.406^{**}	0.433^{**}
$z_1^{WS\,\mathrm{e}}$	-	-	-	-	1	-0.549^{**}
$y_4^{WS\,\mathrm{f}}$	-	-	-	-	-	1

[a] Total OPEX
[b] Volume of water entering the system
[c] Elevated volume of water
[d] Number of households with effective water supply service
[e] Raw water quality
[f] Total length of pipelines
** Significance level of 1%
* Significance level of 5%

Table 4: Bivariate Pearson correlation among the variables selected for WT Models.

	$x_1^{WT\,\mathrm{a}}$	$y_1^{WT\,\mathrm{b}}$	$y_2^{WT\,\mathrm{c}}$	$y_3^{WT\,\mathrm{d}}$	$z_1^{WT\,\mathrm{e}}$	$y_4^{WT\,\mathrm{f}}$
$x_1^{WT\,\mathrm{a}}$	1	0.890^{**}	0.646^{**}	0.934^{**}	0.179	0.808^{**}
$y_1^{WT\,\mathrm{b}}$	-	1	0.697^{**}	0.969^{**}	0.169	0.559^{**}
$y_2^{WT\,\mathrm{c}}$	-	-	1	0.741^{**}	0.300^{*}	0.550^{**}
$y_3^{WT\,\mathrm{d}}$	-	-	-	1	0.289^{*}	0.695^{**}
$z_1^{WT\,\mathrm{e}}$	-	-	-	-	1	0.375^{**}
$y_4^{WT\,\mathrm{f}}$	-	-	-	-	-	1

[a] Total OPEX
[b] Volume of wastewater treated in treatment plants
[c] Elevated volume of wastewater
[d] Number of households with effective wastewater treatment service
[e] Effluent quality
[f] Total length of collectors
** Significance level of 1%
* Significance level of 5%

Several positive and statistically significant correlations between the chosen input and the selected outputs validate the regulatory authority's choices. Additionally, due to the positive and statistically significant correlation between almost all pairs of outputs of each model, adding or replacing some of them with other indicators could have been a reality. Infrastructure-related indicators, namely, the *Number of installations*, were a possibility. Nevertheless, the decision was made by the regulatory authority not to include them due to less strong correlations with other indicators.

It should be noted the considerable effort by the regulatory authority to provide clean panel data for all service providers to enable a careful analysis and an in-depth specification of the variables to be included in the models.

5. Results and discussion

This section contains the results and their discussion regarding the measurement of efficiency, computation of peers, and calculation of the ideal targets of the water supply and wastewater treatment service providers operating in the wholesale market segment. Subsection 5.1, Subsection 5.2, and Subsection 5.3 encompass these findings.

5.1. Water supply

On average, between 2017 and 2021, if we consider only the entities for which there is evidence of inefficiency ($\theta_{j_0} < 1$), water supply service providers obtained a mean score of approximately 0.9528 and 0.9541 according to the WS-RU Model and the WS-RC Model, respectively. In particular, between 2017 and 2021, 40% of the service providers were considered inefficient, according to both models. All scores ranged from 0.8731 to 2.5116 and 0.8731 to 1.0000 in the five considered years, respectively. The influence of outliers is evident when *Raw water quality* is not considered a modelling context variable since the service providers are being compared with very similar peers in the WS-RC Model.

Figure 1 details the evolution of the mean efficiency scores per year and methodology, considering all observations each year. The evolution trend of the efficiency scores points to a decrease during the period considered according to both models, with the lowest average value being reached in 2020. Table A.14 in Appendix A.2 provides further details on these results.

Ideally, service providers located below the efficient frontier should guide their improvement process by considering the performance levels observed in one or more peers. It should be noted that the fact that the study sample was relatively small, combined with using a modelling context variable in the WS-RC Model, resulted in an internal benchmarking exercise. External benchmarking occurred in four cases for the WS-RU Model out of 18 inefficient DMUs. Consequently, per year, each inefficient service provider generally has an ideal peer corresponding to itself in another year. 2017 was the year that emerged more frequently in the generated peers.

Thus, it is relevant to study the role of the descriptive contextual variables on the computed efficiency scores. Consequently, non-parametric hypothesis tests appear as the indicated approach; hence, the Kruskal-Wallis H test was applied to the groups of sampled utilities to assess the existence of statistically significant differences between their efficiency scores. In particular, the null hypothesis states that k samples are derived from the same population: if the hypothesis is true, the distribution of the obtained efficiency scores is not statistically significant; otherwise, rejection of the null hypothesis occurs at a significance level of 95% if the p-value is equal to or less than 0.05.

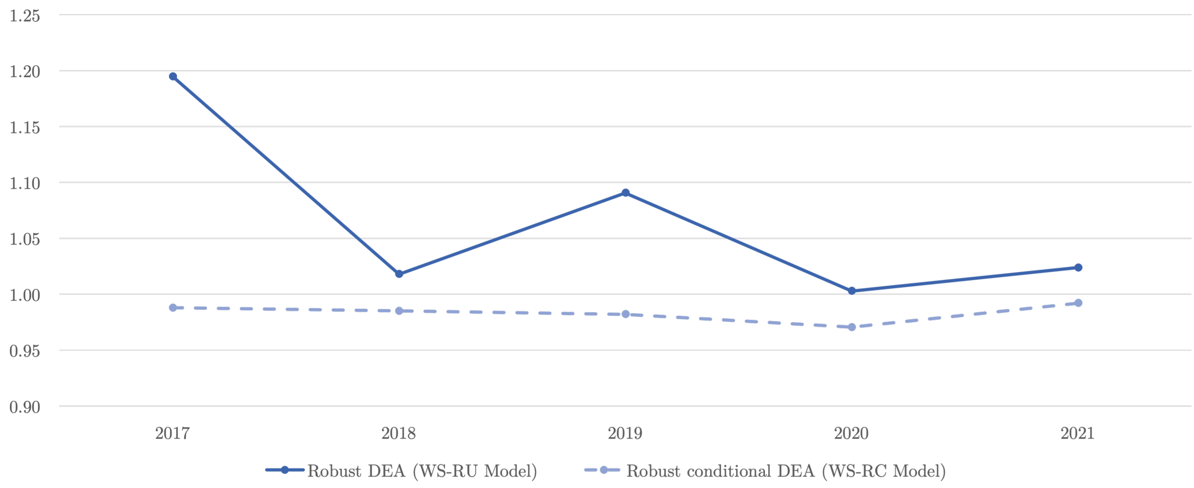

Figure 1: Annual evolution of average efficiency scores for the wholesale water supply service per model.

The results presented in Table 5 point to the retention of the null hypothesis in all cases except for the *Typology of the intervention area* in the WS-RU Model, which indicates that this contextual variable has a statistically significant influence on the efficiency scores of the wholesale water supply service providers when *Raw water quality* is not considered.

Table 5: p-values of the Kruskal-Wallis H tests for the descriptive contextual variables of the water supply service.

Model	***Management model***	***Typology of the intervention area***
WS-RU	0.915	0.001*
WS-RC	0.371	0.415

* Significance level of 5%

Additionally, accounting for the statistically significant differences in the distribution of efficiencies in the WS-RU Model in terms of the *Typology of the intervention area*, it is necessary to perform paired Mann-Whitney U tests to understand the source of statistically significant differences among intervention areas. Table 6 presents the results of these tests already adjusted for the Bonferroni correction. Indeed, the comparisons revealed statistically significant differences for 'Predominantly rural area' vs. 'Moderately urban area' and 'Predominantly rural area' vs. 'Predominantly urban area'. The same did not occur for 'Moderately urban area' vs. 'Predominantly urban area'. In other words, predominantly rural areas are distinct from moderately and predominantly urban areas, evident in their higher mean efficiency scores (1.2379 vs. 1.0012 and 1.2379 vs. 0.9702, respectively). We could not detect statistically significant differences in the WS-RC Model due to the small sample size, and the robust conditional model compares service providers operating in a similar context, which further reduces the comparison potential.

At last, to test the relevance of operating simultaneously in the wholesale and retail market segments on the computed efficiency scores, another Mann-Whitney U test was applied to the groups of sampled utilities. However, the null hypothesis was retained in both the WS-RU

Table 6: p-values of the Mann-Whitney U tests for the *Typology of the intervention area* in the WS-RU Model.

		Typology of the intervention area		
		Predominantly rural area	Moderately urban area	Predominantly urban area
Typology of the intervention area	Predominantly rural area	-	0.012*	0.001*
	Moderately urban area	-	-	0.473
	Predominantly urban area	-	-	-

* Significance level of 5%

Model and the WS-RC Model, implying that operating in the two market segments or only in the wholesale one does not significantly influence the efficiency scores.

5.2. Wastewater treatment

On average, between 2017 and 2021, if we consider only the entities for which there is evidence of inefficiency ($\theta_{j_0} < 1$), wastewater treatment service providers obtained a mean score of 0.9378 and 0.9507 according to the WT-RU Model and the WT-RC Model, respectively - values somewhat similar to those generated by the WS-RU Model and the WS-RC Model. In particular, between 2017 and 2021, 25% of the service providers were considered inefficient, according to both models. All scores ranged from 0.8266 to 3.0230 and 0.8380 to 1.0000 in the five considered years, respectively. In this case, the influence of outliers is evident when *Effluent quality* is not considered a modelling context variable since the service providers are being compared with very similar peers in the WT-RC Model.

Figure 2 details the evolution of the mean efficiency scores per year and methodology, considering all observations each year. The evolution trend of the efficiency scores points to an increase during the period considered according to the WT-RU Model and a slight decrease according to the WT-RC Model. Table A.15 in Appendix A.2 provides further details on these results.

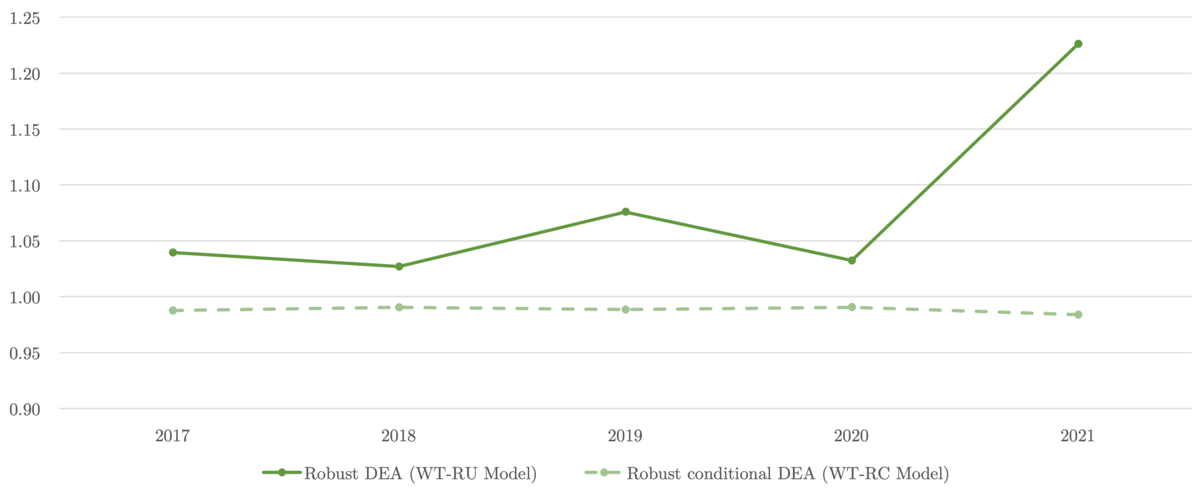

Figure 2: Annual evolution of average efficiency scores for the wholesale wastewater treatment service per model.

As for the water supply service, service providers located below the efficient frontier should guide their improvement process by considering the performance levels observed in one or more peers. Once again, it should be noted that the fact that the study sample was relatively small, combined with using a modelling context variable in the WT-RC Model, resulted in an internal benchmarking exercise. External benchmarking occurred in six cases for the WT-RU Model out of 16 inefficient DMUs. Consequently, per year, each inefficient service provider generally has an ideal peer corresponding to itself in another year. 2018 was the year that emerged more frequently in the generated peers.

Thus, it is relevant to study the role of the descriptive contextual variables on the computed efficiency scores. Similarly to the previous service, non-parametric hypothesis tests in the same conditions appear as the indicated approach. This way, the results presented in Table 7 point to the rejection of the null hypothesis in all cases except for the *Management model* in the WT-RC Model, which indicates that the remaining contextual variables statistically influence the efficiency scores of the wholesale wastewater treatment service providers.

Table 7: p-values of the Kruskal-Wallis H tests for the descriptive contextual variables of the wastewater treatment service.

Model	***Management model***	***Typology of the intervention area***
WT-RU	0.001*	< 0.001*
WT-RC	0.314	0.007*

* Significance level of 5%

Additionally, accounting for the statistically significant differences in the distribution of efficiencies in the WT-RU Model in terms of the *Management model* and the *Typology of the intervention area* and the WT-RC Model in terms of the *Typology of the intervention area*, it is necessary to perform paired Mann-Whitney U tests to understand the source of statistically significant differences among management models and intervention areas. Table 8, Table 9, and Table 10 present the results of these tests already adjusted for the Bonferroni correction. First, the comparisons of the results of the *Management Model* of the WT-RU Model revealed statistically significant differences for 'Concession' vs. 'Delegation' and 'Concession' vs. 'Direct management'. In other words, concession models are distinct from the delegation and direct management models, which is evident in their lower mean efficiency scores (1.0509 vs. 1.0520 and 1.0509 vs. 1.6406, respectively). This finding implies that when the State only participates in an operator's capital instead of owning or operating it, the service seems less efficient. However, note that only one operator is managed under a delegation model and another one under a direct management model, which means that these results should be interpreted cautiously. We could not detect statistically significant differences in the WT-RC Model due to the small sample size and the fact that the robust conditional model compares service providers operating in a similar context, which further reduces the comparison potential. Second, comparing the results of the *Typology of the intervention area* of the WT-RU Model revealed statistically significant differences for 'Predominantly rural area' vs. 'Moderately urban area' and 'Predominantly rural area' vs. 'Predominantly urban area'. In other words, predominantly rural areas are, once again, distinct from moderately and predominantly urban areas, which is evident in their higher mean efficiency scores (1.2464 vs. 1.0603 and 1.2464 vs. 0.9755, respectively). Third, comparing the *Typology of the intervention area* results from the WT-RC Model revealed statistically significant differences for 'Moderately urban area' vs. 'Predominantly urban area'. In other words, moderately urban areas are distinct from predominantly urban areas, given their higher mean efficiency scores (0.9984 vs. 0.9664).

Finally, an additional Mann-Whitney U test was applied to the groups of sampled utilities

Table 8: p-values of the Mann-Whitney U tests for the *Management model* in the WT-RU Model.

		Management model		
		Concession	Delegation	Direct management
Management model	Concession	-	0.032*	0.013*
	Delegation	-	-	1.000
	Direct management	-	-	-

* Significance level of 5%

Table 9: p-values of the Mann-Whitney U tests for the *Typology of the intervention area* in the WT-RU Model.

		Typology of the intervention area		
		Predominantly rural area	Moderately urban area	Predominantly urban area
Typology of the intervention area	Predominantly rural area	-	< 0.001*	0.001*
	Moderately urban area	-	-	1.000
	Predominantly urban area	-	-	-

* Significance level of 5%

Table 10: p-values of the Mann-Whitney U tests for the *Typology of the intervention area* in the WT-RC Model.

		Typology of the intervention area		
		Predominantly rural area	Moderately urban area	Predominantly urban area
Typology of the intervention area	Predominantly rural area	-	1.000	0.072
	Moderately urban area	-	-	0.007*
	Predominantly urban area	-	-	-

* Significance level of 5%

to test the relevance of operating simultaneously in the wholesale and retail market segments on the computed efficiency scores. Once again, the null hypothesis was retained in both the WT-RU Model and the WT-RC Model, implying that operating in the two market segments or only in the wholesale one does not significantly influence the efficiency scores.

5.3. Estimated savings

Finally, the total OPEX target an inefficient service provider needs to achieve based on the comparison with the values of its peers in order to become efficient must be known. Indeed, for regulatory purposes, it is recommended that such targets should be estimated via robust conditional models to ensure a more homogeneous comparison since they correspond to a more complete and conservative approach in terms of what is considered to be the effective potential for improvement in the sector. Nonetheless, we present the results of both models for comparison purposes.

Therefore, using Expression (2), it is estimated that the sum of the ideal total OPEX for the considered period varies between 13,431,583.54 € and 13,699,473.68 € for the water supply service and between 19,188,431.38 € and 26,590,580.70 € for the wastewater treatment service. This would allow average annual savings between 2.13% and 2.18% for the former and 2.36% and 3.22% for the latter. Figure 3 and Figure 4 show the total potential annual savings for each service based on these targets. It should be noted that the ideal total OPEX values are computed based on the outputs produced by each service provider, which justifies the reported annual variations.

It should be noted that entities for which there is evidence of super-efficiency ($\theta_{j_0} > 1$), in line with Mergoni et al. (2022), are DMUs that are doing better than the average m DMUs they are compared with (De Witte & Schiltz, 2018), which means that they do not need to reduce OPEX and their optimal OPEX value is the same as their original one.

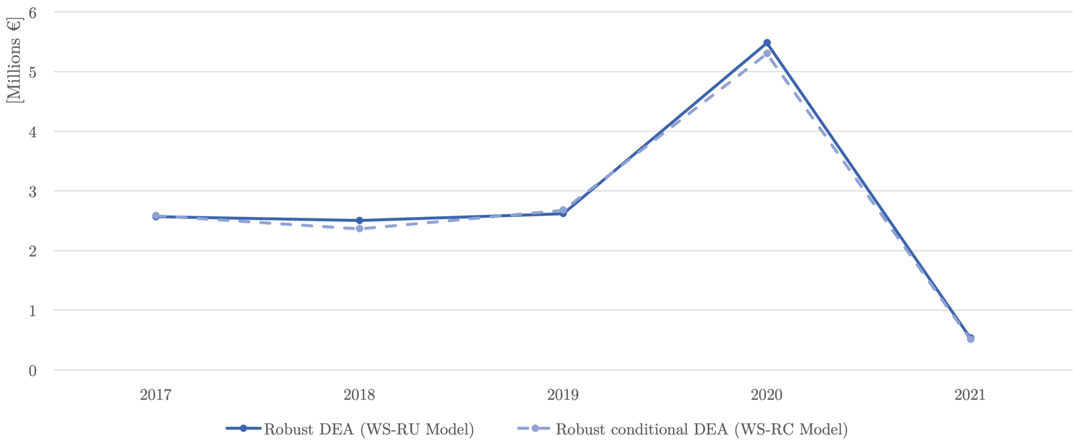

Figure 3: Potential annual savings for the water supply service.

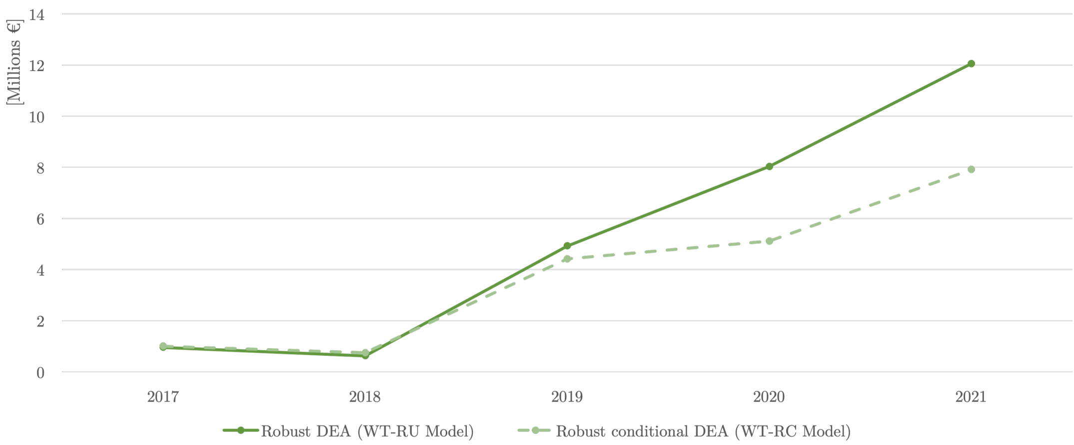

Figure 4: Potential annual savings for the wastewater treatment service.

6. Conclusion

This work addressed the problem of measuring the efficiency of the water supply and wastewater treatment service providers operating in the wholesale market segment in an EU country between 2017 and 2021. For this purpose, two methodologies with consolidated theoretical bases were employed in collaboration with the country's regulatory authority, being able to benchmark these services, avoid the impact of *outliers*, and allow the understanding of contextual factors. Fundamentally, the country's regulatory authority was involved in every process step, from selecting the sample and choosing suitable inputs, outputs, and contextual variables to critically analysing the results.

Nonetheless, robust conditional models should be considered the more accurate and conservative instruments for regulatory purposes. The regulatory authority saw such models as invaluable regulation tools with plenty of potential for future regulatory frameworks, such as monitoring service quality and setting efficient water tariffs. Moreover, by adopting a sunshine regulation strategy, the regulator can leverage the findings of this study to simulate market competition and encourage the operators to meet benchmark efficiency levels. Hence, they can ascertain the best practices of their peers and attain more favourable OPEX levels.

The results point to similar efficiency scores between the wastewater treatment service and the water supply service in the five years in question, although being slightly higher in the former (1.08 and 0.99 vs. 0.99 and 0.98 for the robust approach and the conditional approach to each service). In addition, it is estimated that, for the level of production of the service providers in each service, it is possible to save 2.13% (robust approach) and 2.18% (conditional approach) and 2.36% (robust approach) and 3.22% (conditional approach) of their respective average yearly total OPEX, which corresponds to 2,739,894.74 € (robust approach) and 2,686,316.71 € (conditional approach) for the former and 5,318,116.14 € (robust approach) and 3,837,686.28 € (conditional approach) for the latter. It should be noted that the *Management model* and the *Typology of the intervention area* showed statistically significant differences in terms of their role in influencing the efficiency scores obtained.

As main limitations, we point out three aspects. First, the availability and quality of some data motivated the use of *Raw water quality* and *Effluent quality* in their present form (developed internally by the regulatory authority specifically for this analysis) as *proxies*. Second, the positive and statistically significant correlation between almost all pairs of outputs of each model could motivate the addition or replacement of some of them by other indicators. Third, the conditional nature of one of the methodologies transforms an external benchmarking exercise (given the comparison of a service provider with others) into an internal benchmarking exercise (since each service provider, when inefficient, becomes its own peer); for this reason, and allied to the reduced sample size, it was not possible to study the impact of the modelling contextual variables on the results by comparing the use of conditional and non-conditional models. The continued improvement of these shortcomings will lead to results even more suited to the fine-tuning of this collaborative regulatory framework, which will result in more transparent regulation and more efficient governance of the water sector. Alternative DEA methodologies should also be considered, e.g., window DEA to deal with the multi-period nature of the samples, meta-frontier DEA to account for the categorical contextual variables, and output-side weight restrictions based on the regulatory authority's preferences. Multi-criteria decision analysis-based approaches should also be considered to incorporate multiple stakeholders' value judgements further and ease the consensus-reaching process.

Acknowledgements

Miguel Alves Pereira would like to thank the Portuguese Foundation for Science and Technology (FCT, from the Portuguese abbreviation of *Fundação para a Ciência e a Tecnologia*) for supporting this research via project UIDB/00097/2020 at CEGIST. Hermilio Vilarinho would also like to thank FCT for supporting this research via the PhD research studentship 2021.05244.BD. The authors gratefully acknowledge the support provided by the regulatory authority in the project context for developing a methodology for estimating the efficiency of utilities that provide water supply and wastewater sanitation services. The role played by Anna Mergoni in the methodology phase should also be praised. We would also like to leave a word of appreciation to the Editor-in-chief and the two anonymous referees for their invaluable insights that improved earlier versions of this paper. Our views (and any errors) are our responsibility.

References

Ablanedo-Rosas, J. H., Guerrero Campanur, A., Olivares-Benitez, E., Sánchez-García, J. Y., & Nuñez-Ríos, J. E. (2020). Operational Efficiency of Mexican Water Utilities: Results of a Double-Bootstrap Data Envelopment Analysis. *Water*, *12*, 553. URL: `https://www.mdpi.com/2073-4441/12/2/553`. doi:10.3390/w12020553.

Akhmouch, A., & Correia, F. N. (2016). The 12 OECD principles on water governance – When science meets policy. *Utilities Policy*, *43*, 14–20. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0957178715300886`. doi:10.1016/j.jup.2016.06.004.

Ananda, J. (2014). Evaluating the Performance of Urban Water Utilities: Robust Non-parametric Approach. *Journal of Water Resources Planning and Management*, *140*, 04014021. URL: `http://ascelibrary.org/doi/10.1061/{%}28ASCE{%}29WR.1943-5452.0000387`. doi:10.1061/(ASCE)WR.1943-5452.0000387.

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). NONPARAMETRIC FRONTIER ESTIMATION: A CONDITIONAL QUANTILE-BASED APPROACH. *Econometric Theory*, *21*. URL: `http://www.journals.cambridge.org/abstract{_}S0266466605050206`. doi:10.1017/S0266466605050206.

Baptista, J. M. (2014). *Uma abordagem regulatória integrada (ARIT-ERSAR) para os serviços de águas e resíduos*. Technical Report Entidade Reguladora dos Serviços de Águas e Resíduos Lisbon, Portugal.

Berg, S. V., & Marques, R. C. (2011). Quantitative studies of water and sanitation utilities: A benchmarking literature survey. *Water Policy*, *13*, 591–606. URL: `https://iwaponline.com/wp/article/13/5/591/19988/Quantitative-studies-of-water-and-sanitation`. doi:10.2166/wp.2011.041.

Caplan, E., Greene, Z., Womble, J., MacLaughlin, K., & Bird, L. (2022). The impacts of wholesale market rules and policies on clean energy goals: A primer for local governments.

Cardoso, J., Rosa, J., Silva, N., Rodrigues, R., Faroleiro, P., Ruivo, F., Lobo, F., Videira, C., Rosa, D., Ferreira, D., Santos, C., Nunes, M., Carreira, P., Aleixo, C., Varela, C., Abreu, M. J., & Sá, S. (2019). *Relatório Anual dos Serviços de Águas e Resíduos em Portugal*

(2019): Volume 1 - Caracterização do setor de águas e resíduos. Technical Report Entidade Reguladora dos Serviços de Águas e Resíduos Lisbon, Portugal.

Carvalho, P., & Marques, R. C. (2011). The influence of the operational environment on the efficiency of water utilities. *Journal of Environmental Management*, *92*, 2698–2707. URL: https://linkinghub.elsevier.com/retrieve/pii/S0301479711002052. doi:10.1016/j.jenvman.2011.06.008.

Carvalho, P., & Marques, R. C. (2014). Computing economies of vertical integration, economies of scope and economies of scale using partial frontier nonparametric methods. *European Journal of Operational Research*, *234*, 292–307. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377221713007777. doi:10.1016/j.ejor.2013.09.022.

Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, *106*, 1–25. URL: https://linkinghub.elsevier.com/retrieve/pii/S030440760100080X. doi:10.1016/S0304-4076(01)00080-X.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444. URL: http://linkinghub.elsevier.com/retrieve/pii/0377221778901388. doi:10.1016/0377-2217(78)90138-8.

Costa, A., Cardoso, J., Rodrigues, R., Rodrigues, S., Ruivo, F., Gomes, P., Videira, C., Cunha, A., Ribeiro, C., Silva, J., Mesquita, A., Ramos, A. R., Gonçalves, P., Silva, R., Sá, S., Varela, C., Andrade, I., & Abreu, M. J. (2021). *Relatório Anual dos Serviços de Águas e Resíduos em Portugal (2021): Volume 1 - Caracterização do setor de águas e resíduos*. Technical Report Entidade Reguladora dos Serviços de Águas e Resíduos Lisbon, Portugal.

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, *140*, 375–400. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407606001400. doi:10.1016/j.jeconom.2006.07.002.

Daraio, C., & Simar, L. (2005). Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. *Journal of Productivity Analysis*, *24*, 93–121. URL: http://link.springer.com/10.1007/s11123-005-3042-8. doi:10.1007/s11123-005-3042-8.

Daraio, C., & Simar, L. (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis* volume 4 of *Studies in Productivity and Efficiency*. Boston, MA: Springer US. URL: http://link.springer.com/10.1007/978-0-387-35231-2. doi:10.1007/978-0-387-35231-2.

Daraio, C., & Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, *28*, 13–32. URL: http://link.springer.com/10.1007/s11123-007-0049-3. doi:10.1007/s11123-007-0049-3.

Davis, J. (2005). Private-sector participation in the water and sanitation sector. *Annual Review of Environment and Resources*, *30*, 145–183. URL: https://www.annualreviews.org/doi/10.1146/annurev.energy.30.050504.144635. doi:10.1146/annurev.energy.30.050504.144635.

De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, *45*, 2401–2412. URL: http://www.tandfonline.com/doi/abs/10.1080/00036846.2012.665602. doi:10.1080/00036846.2012.665602.

De Witte, K., & Marques, R. C. (2010a). Designing performance incentives, an international benchmark study in the water sector. *Central European Journal of Operations Research*, *18*, 189–220. URL: http://link.springer.com/10.1007/s10100-009-0108-0. doi:10.1007/s10100-009-0108-0.

De Witte, K., & Marques, R. C. (2010b). Influential observations in frontier models, a robust non-oriented approach to the water sector. *Annals of Operations Research*, *181*, 377–392. URL: http://link.springer.com/10.1007/s10479-010-0754-6. doi:10.1007/s10479-010-0754-6.

De Witte, K., & Schiltz, F. (2018). Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional Benefit-of-the-Doubt approach. *European Journal of Operational Research*, *267*, 1172–1181. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377221717311682. doi:10.1016/j.ejor.2017.12.034.

Dziedzic, R., & Karney, B. (2014). Water Distribution System Performance Metrics. *Procedia Engineering*, *89*, 363–369. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877705814023157. doi:10.1016/j.proeng.2014.11.200.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*, 253. URL: https://www.jstor.org/stable/2343100?origin=crossref. doi:10.2307/2343100.

Ferreira da Cruz, N., Marques, R. C., Romano, G., & Guerrini, A. (2012). Measuring the efficiency of water utilities: A cross-national comparison between Portugal and Italy. *Water Policy*, *14*, 841–853. URL: https://iwaponline.com/wp/article/14/5/841/20033/Measuring-the-efficiency-of-water-utilities-a. doi:10.2166/wp.2012.103.

Ferro, G., & Romero, C. A. (2011). Setting performance standards for regulation of water services: Benchmarking Latin American utilities. *Water Policy*, *13*, 607–623. URL: https://iwaponline.com/wp/article/13/5/607/19991/Setting-performance-standards-for-regulation-of. doi:10.2166/wp.2011.042.

Fu, W., & Jacobs, B. W. (2022). Does increased water efficiency improve financial performance? The important role of operational efficiency. *International Journal of Operations & Production Management*, *42*, 304–330. URL: https://www.emerald.com/insight/content/doi/10.1108/IJOPM-10-2021-0628/full/html. doi:10.1108/IJOPM-10-2021-0628.

Fusco, E., Vidoli, F., & Rogge, N. (2020). Spatial directional robust Benefit of the Doubt approach in presence of undesirable output: An application to Italian waste sector. *Omega*, *94*, 102053. URL: https://linkinghub.elsevier.com/retrieve/pii/S0305048318308867. doi:10.1016/j.omega.2019.03.011.

Gidion, D. K., Hong, J., Hou, B., Shillamkwese, S. S., Adams, M. Z. A., & Khoveyni, M. (2019). Consistency in efficiency benchmarking: urban water utility regulation with performance improvement. *Water Supply*, *19*, 1955–1964. URL: https://iwaponline.com/ws/article/19/7/1955/67546/Consistency-in-efficiency-benchmarking-urban-water. doi:10.2166/ws.2019.072.

Goh, K. H., & See, K. F. (2021). Twenty years of water utility benchmarking: A bibliometric analysis of emerging interest in water research and collaboration. *Journal of Cleaner Production*, *284*, 124711. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959652620347557. doi:10.1016/j.jclepro.2020.124711.

Henriques, A. A., Camanho, A. S., Amorim, P., & Silva, J. G. (2020). Performance benchmarking using composite indicators to support regulation of the Portuguese wastewater sector. *Utilities Policy*, *66*, 101082. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178720300771. doi:10.1016/j.jup.2020.101082.

Henriques, A. A., Fontes, M., Camanho, A. S., D'Inverno, G., Amorim, P., & Silva, J. G. (2022). Performance evaluation of problematic samples: a robust nonparametric approach for wastewater treatment plants. *Annals of Operations Research*, . URL: https://link.springer.com/10.1007/s10479-022-04629-z. doi:10.1007/s10479-022-04629-z.

Lo Storto, C. (2013). Are Public-Private Partnerships a Source of Greater Efficiency in Water Supply? Results of a Non-Parametric Performance Analysis Relating to the Italian Industry. *Water*, *5*, 2058–2079. URL: http://www.mdpi.com/2073-4441/5/4/2058. doi:10.3390/w5042058.

Maziotis, A., Villegas, A., Molinos-Senante, M., & Sala-Garrido, R. (2020). Impact of external costs of unplanned supply interruptions on water company efficiency: Evidence from Chile. *Utilities Policy*, *66*, 101087. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178720300825. doi:10.1016/j.jup.2020.101087.

Mehta, M., Mehta, D., & Immanuel, A. (2013). *A Review of Performance Benchmarking: Urban Water Supply and Sanitation*. Technical Report CEPT University Ahmedabad, India.

Mergoni, A., D'Inverno, G., & Carosi, L. (2022). A composite indicator for measuring the environmental performance of water, wastewater, and solid waste utilities. *Utilities Policy*, *74*, 101285. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178721001193. doi:10.1016/j.jup.2021.101285.

Molinos-Senante, M., & Maziotis, A. (2018). Assessing the influence of exogenous and quality of service variables on water companies´ performance using a true-fixed stochastic frontier approach. *Urban Water Journal*, *15*, 682–691. URL: https://www.tandfonline.com/doi/full/10.1080/1573062X.2018.1539502. doi:10.1080/1573062X.2018.1539502.

Molinos-Senante, M., Maziotis, A., & Sala-Garrido, R. (2020). Evaluating trends in the performance of Chilean water companies: impact of quality of service and environmental variables. *Environmental Science and Pollution Research*, *27*, 13155–13165. URL: http://link.springer.com/10.1007/s11356-020-07918-x. doi:10.1007/s11356-020-07918-x.

Organisation for Economic Co-operation and Development (2015). *OECD Principles on Water Governance*. Technical Report Organisation for Economic Co-operation and Development Paris, France.

Pereira, M. A., & Marques, R. C. (2021). Sustainable water and sanitation for all: Are we there yet? *Water Research*, *207*, 117765. URL: https://linkinghub.elsevier.com/retrieve/pii/S0043135421009593. doi:10.1016/j.watres.2021.117765.

Pereira, M. A., & Marques, R. C. (2022a). From a millennium to a sustainable water and sanitation development: were we there already? *Journal of Water Supply: Research and Technology-Aqua*, *71*, 293–300. URL: https://iwaponline.com/aqua/article/doi/10.2166/aqua.2022.124/86258/From-a-millennium-to-a-sustainable-water-andhttps://iwaponline.com/aqua/article/71/2/293/86258/From-a-millennium-to-a-sustainable-water-and. doi:10.2166/aqua.2022.124.

Pereira, M. A., & Marques, R. C. (2022b). Is sunshine regulation the new prescription to brighten up public hospitals in Portugal? *Socio-Economic Planning Sciences*, (p. 101219). URL: https://linkinghub.elsevier.com/retrieve/pii/S0038012121002111. doi:10.1016/j.seps.2021.101219.

Pereira, M. A., & Marques, R. C. (2022c). Technical and Scale Efficiency of the Brazilian Municipalities' Water and Sanitation Services: A Two-Stage Data Envelopment Analysis. *Sustainability*, *14*, 199. URL: https://www.mdpi.com/2071-1050/14/1/199. doi:10.3390/su14010199.

Pinto, F. S., Simões, P., & Marques, R. C. (2017a). Water services performance: do operational environment and quality factors count? *Urban Water Journal*, *14*, 773–781. URL: https://www.tandfonline.com/doi/full/10.1080/1573062X.2016.1254254. doi:10.1080/1573062X.2016.1254254.

Pinto, F. S., Simões, P., & Marques, R. C. (2017b). Raising the bar: The role of governance in performance assessments. *Utilities Policy*, *49*, 38–47. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178716303629. doi:10.1016/j.jup.2017.09.001.

Rogge, N., & De Jaeger, S. (2013). Measuring and explaining the cost efficiency of municipal solid waste collection and processing services. *Omega*, *41*, 653–664. URL: https://linkinghub.elsevier.com/retrieve/pii/S0305048312001867. doi:10.1016/j.omega.2012.09.006.

Romano, G., Salvati, N., & Guerrini, A. (2018). Governance, strategy and efficiency of water utilities: the Italian case. *Water Policy*, *20*, 109–126. URL: https://iwaponline.com/wp/article/20/1/109/38141/Governance-strategy-and-efficiency-of-water. doi:10.2166/wp.2017.172.

Salazar-Adams, A. (2021). The efficiency of post-reform water utilities in Mexico. *Utilities Policy*, *68*, 101153. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178720301466. doi:10.1016/j.jup.2020.101153.

Santos, L. F., Galvão, A. F., & Cardoso, M. A. (2019). Performance indicators for urban storm water systems: a review. *Water Policy*, *21*, 221–244. URL: https://iwaponline.

com/wp/article/21/1/221/64780/Performance-indicators-for-urban-storm-water. doi:10.2166/wp.2018.042.

Satoh, E. (2015). Nontransferable water rights and technical inefficiency in the Japanese water supply industry. *Water Resources and Economics*, *11*, 13–21. URL: https://linkinghub.elsevier.com/retrieve/pii/S2212428415300037. doi:10.1016/j.wre.2015.08.001.

Satoh, E. (2019). Water Demand Fluctuations, Non-Transferable Water Rights, and Technical Inefficiency in Japan's Water Sector. *Water Economics and Policy*, *05*, 1850028. URL: https://www.worldscientific.com/doi/abs/10.1142/S2382624X18500285. doi:10.1142/S2382624X18500285.

Tourinho, M., Barbosa, F., Santos, P. R., Pinto, F. T., & Camanho, A. S. (2022). Productivity change in Brazilian water services: A benchmarking study of national and regional trends. *Socio-Economic Planning Sciences*, (p. 101491). URL: https://linkinghub.elsevier.com/retrieve/pii/S0038012122002981. doi:10.1016/j.seps.2022.101491.

Tourinho, M., Santos, P. R., Pinto, F. T., & Camanho, A. S. (2021). Performance assessment of water services in Brazilian municipalities: An integrated view of efficiency and access. *Socio-Economic Planning Sciences*, (p. 101139). URL: https://linkinghub.elsevier.com/retrieve/pii/S0038012121001312. doi:10.1016/j.seps.2021.101139.

United Nations (2022). *The Sustainable Development Goals Report 2022*. Technical Report United Nations New York, NY.

Vieira, P., Jorge, C., & Covas, D. (2015). Novel Performance Assessment Indices for Domestic Water Use. *Procedia Engineering*, *119*, 813–819. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877705815026132. doi:10.1016/j.proeng.2015.08.943.

Vilanova, M. R. N., Magalhães Filho, P., & Balestieri, J. A. P. (2015). Performance measurement and indicators for water supply management: Review and international cases. *Renewable and Sustainable Energy Reviews*, *43*, 1–12. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364032114009745. doi:10.1016/j.rser.2014.11.043.

Vilarinho, H., D'Inverno, G., Nóvoa, H., & Camanho, A. S. (2023). The measurement of asset management performance of water companies. *Socio-Economic Planning Sciences*, (p. 101545). URL: https://linkinghub.elsevier.com/retrieve/pii/S0038012123000459. doi:10.1016/j.seps.2023.101545.

Williams, S., Pickard, C., Glass, K., & Glass, A. (2020). Benchmarking Water Retail Cost Efficiency in England and Wales. *International Journal of the Economics of Business*, *27*, 431–467. URL: https://www.tandfonline.com/doi/full/10.1080/13571516.2020.1790979. doi:10.1080/13571516.2020.1790979.

World Bank (1992). *Governance and Development*. Washington, DC: World Bank. URL: http://elibrary.worldbank.org/doi/book/10.1596/0-8213-2094-7. doi:10.1596/0-8213-2094-7.

Appendix A. Descriptive statistics

Appendix A.1. Variables

Appendix A.2. Efficiency scores

Table A.11: Descriptive statistics of the variables used in WS Models.

Year	**Variable**	**Mean**	**Standard deviation**	**Minimum**	**Maximum**
2017	x_1^{WS}	12,395,102.16	10,033,834.25	395,495.00	26,460,650.66
	y_1^{WS}	63,980,827.14	68,994,925.54	2,151,337.00	229,002,657.80
	y_2^{WS}	95,400,689.13	124,276,267.73	133,220.00	411,152,726.57
	y_3^{WS}	363,978.80	315,610.54	13,100.00	831,458.00
	y_4^{WS}	1,038.84	1,196.51	26.80	3,592.50
	z_1^{WS}	2.58	1.04	1.01	4.00
2018	x_1^{WS}	12,279,225.73	9,902,925.74	452,405.58	27,994,392.92
	y_1^{WS}	61,546,785.64	65,922,840.17	1,671,349.70	218,116,734.10
	y_2^{WS}	75,926,301.82	74,837,040.63	97,320.00	216,991,501.40
	y_3^{WS}	363,978.80	315,610.54	13,100.00	831,458.00
	y_4^{WS}	1,038.84	1,196.51	26.80	3,592.50
	z_1^{WS}	2.62	1.09	1.01	4.00
2019	x_1^{WS}	12,534,473.63	10,031,587.68	434,584.29	28,867,824.39
	y_1^{WS}	63,238,860.53	66,926,146.21	2,295,527.00	221,836,249.50
	y_2^{WS}	79,291,274.48	78,851,433.83	92,899.28	227,531,526.53
	y_3^{WS}	363,978.80	315,610.54	13,100.00	831,458.00
	y_4^{WS}	1,038.84	1,196.51	26.80	3,592.50
	z_1^{WS}	2.38	1.16	1.00	4.00
2020	x_1^{WS}	12,885,563.17	10,454,720.57	493,640.65	30,304,818.28
	y_1^{WS}	63,486,330.01	66,564,159.78	1,853,111.00	221,124,927.50
	y_2^{WS}	87,498,453.77	91,680,391.39	90,282.00	291,069,399.21
	y_3^{WS}	363,978.80	315,610.54	13,100.00	831,458.00
	y_4^{WS}	1,038.84	1,196.51	26.80	3,592.50
	z_1^{WS}	2.35	1.16	1.00	4.00
2021	x_1^{WS}	12,220,446.01	9,690,404.96	462,896.00	26,894,013.72
	y_1^{WS}	63,376,911.69	66,530,277.81	1,974,214.00	221,716,594.60
	y_2^{WS}	82,765,968.47	79,201,694.49	80,448.00	238,510,635.86
	y_3^{WS}	363,978.80	315,610.54	13,100.00	831,458.00
	y_4^{WS}	1,038.84	1,196.51	26.80	3,592.50
	z_1^{WS}	2.38	1.17	1.00	4.00

Table A.12: Descriptive statistics of the variables used in WT Models.

Year	**Variable**	**Mean**	**Standard deviation**	**Minimum**	**Maximum**
2017	x_1^{WT}	13,436,619.49	13,472,594.14	412,369.17	40,721,908.47
	y_1^{WT}	42,102,229.49	52,517,519.75	4,271,902.30	183,032,070.20
	y_2^{WT}	39,564,638.97	49,220,985.12	198,887.00	149,159,693.48
	y_3^{WT}	321,232.10	332,894.80	17,489.00	1,127,557.00
	y_4^{WT}	523.84	530.60	31.20	1,645.00
	z_1^{WT}	3.59	1.51	1.00	5.00
2018	x_1^{WT}	13,965,105.88	13,493,353.61	381,752.16	40,073,630.51
	y_1^{WT}	47,561,707.67	55,896,640.65	4,542,814.00	194,233,441.50
	y_2^{WT}	29,316,932.93	48,280,960.23	278,201.00	152,651,482.29
	y_3^{WT}	321,232.10	332,894.80	17,489.00	1,127,557.00
	y_4^{WT}	523.84	530.60	31.20	1,645.00
	z_1^{WT}	3.59	1.51	1.00	5.00
2019	x_1^{WT}	14,622,204.20	14,300,805.38	1,235,023.21	43,327,904.80
	y_1^{WT}	46,021,939.12	53,178,253.80	4,241,844.20	185,062,474.90
	y_2^{WT}	46,484,766.47	65,437,777.76	286,984.00	184,199,337.79
	y_3^{WT}	321,903.70	332,220.02	24,205.00	1,127,557.00
	y_4^{WT}	520.90	533.55	28.00	1,645.00
	z_1^{WT}	3.99	1.26	1.00	5.00
2020	x_1^{WT}	14,013,753.63	13,757,954.24	472,208.75	45,183,937.02
	y_1^{WT}	44,040,373.08	51,521,976.35	4,908,142.00	193,585,060.80
	y_2^{WT}	52,290,736.79	80,634,323.06	311,960.00	275,252,931.11
	y_3^{WT}	296,744.08	313,109.01	17,489.00	1,127,557.00
	y_4^{WT}	479.03	500.67	28.00	1,645.00
	z_1^{WT}	3.65	1.44	1.00	5.00
2021	x_1^{WT}	14,787,598.11	14,392,195.05	507,821.70	47,817,730.33
	y_1^{WT}	46,818,113.98	51,478,321.80	4,843,681.00	184,164,595.40
	y_2^{WT}	45,322,787.43	59,286,303.73	327,424.00	185,498,401.74
	y_3^{WT}	296,744.08	313,109.01	174,89.00	1,127,557.00
	y_4^{WT}	479.03	500.67	28.00	1,645.00
	z_1^{WT}	3.65	1.44	1.00	5.00

Table A.13: Management model and typology of the intervention area per model.

Descriptive contextual variable	Description	Relative frequency
z_2^{WS}	Concession	70%
	Delegation	30%
z_3^{WS}	Predominantly rural area	30%
	Moderately urban area	50%
	Predominantly urban area	20%
z_2^{WT}	Concession	83%
	Delegation	8%
	Direct management	8%
z_3^{WT}	Predominantly rural area	25%
	Moderately urban area	50%
	Predominantly urban area	25%

Table A.14: Descriptive statistics of the efficiency scores generated by WS Models.

Model	Sample perspective	Type	Mean	Standard deviation	Minimum	Maximum
WS-RU	Full	-	1.0660	0.2650	0.8731	2.5116
	Management model	Concession	1.0457	0.1826	0.8901	1.9882
		Delegation	1.1132	0.4017	0.8731	2.5116
		Direct management	-	-	-	-
	Typology of the intervention area	Predominantly rural area	1.2379	0.4383	0.9056	2.5116
		Moderately urban area	1.0012	0.0620	0.8901	1.2176
		Predominantly urban area	0.9702	0.0444	0.8731	1.0002
WS-RC	Full	-	0.9835	0.0299	0.8731	1.0000
	Management model	Concession	0.9868	0.0247	0.9043	1.0000
		Delegation	0.9757	0.0394	0.8731	1.0000
		Direct management	-	-	-	-
	Typology of the intervention area	Predominantly rural area	0.9916	0.0188	0.9388	1.0000
		Moderately urban area	0.9832	0.0274	0.9043	1.0000
		Predominantly urban area	0.9719	0.0452	0.8731	1.0000

Table A.15: Descriptive statistics of the efficiency scores generated by WT Models.

Model	Sample perspective	Type	Mean	Standard deviation	Minimum	Maximum
WT-RU	Full	-	1.0837	0.3332	0.8266	3.0230
	Management model	Concession	1.0509	0.3146	0.8266	3.0230
		Delegation	1.0520	0.0327	1.0285	1.1084
		Direct management	1.6406	0.4701	1.1009	1.9607
	Typology of the intervention area	Predominantly rural area	1.2464	0.5277	0.9787	3.0230
		Moderately urban area	1.0576	0.2627	0.8266	1.9607
		Predominantly urban area	0.9755	0.0578	0.8380	1.0474
WT-RC	Full	-	0.9881	0.0321	0.8380	1.0000
	Management model	Concession	0.9861	0.0344	0.8380	1.0000
		Delegation	1.0000	0.0000	1.0000	1.0000
		Direct management	1.0000	0.0000	1.0000	1.0000
	Typology of the intervention area	Predominantly rural area	0.9930	0.0204	0.9258	1.0000
		Moderately urban area	0.9984	0.0061	0.9703	1.0000
		Predominantly urban area	0.9664	0.0523	0.8380	1.0000