Measuring the Environmental Pressure of Portuguese Water and Waste Utilities: A Composite Indicator Approach

Anna Mergoni[a,*], Giovanna D'Inverno[a], Laura Carosi[b]

[a]*KU Leuven, Faculty of Business and Economics (FEB), Leuven, Belgium*
[b]*University of Pisa, Department of Economics and Management, Pisa, Italy*

Abstract

The aim of this paper is to measure and benchmark the environmental performance of Portuguese utilities jointly active in the three sectors of water supply, water collection and waste management. To do so, we suggest the use of a traditional (optimistic) directional distance Benefit of the Doubt index. We complement the analysis by considering also the pessimistic version of the proposed BoD and by implementing a robust and conditional approach. The obtained results show that there is space for improvement in the pressure balance of these utilities, especially for small and very large units, mostly operating in urban areas.

Keywords: Environmental sustainability, Environmental pressure indicator, Benefit of the Doubt, Composite Indicator, Robust and conditional analysis.

---

1. Introduction

Environmental sustainability is defined as the set of rules for the "maintenance of the natural capital" (Goodland, 1995, p.10) or as "the ability to maintain the qualities that are valued in the physical environment" (Sutton, 2004, p.1). Within this framework the environmental pressure indicators, i.e. the indicators which focus on the exchanges between the human activities and the environment, play a fundamental role. The idea of controlling for the release of substances and for the use of resources is intrinsic in the definition itself of environmental sustainability and has been widely used in the literature (see for example: Moldan et al., 2012; Dahl, 2012; D'Amato et al., 2017; Purvis et al., 2019). However, the term 'environmental pressure indicator', introduced by Smeets and Weterings (1999), has not received equal attention and there are only a few scholars who adopted explicitly this terminology: Nikolaou (2001); Munksgaard et al. (2005); Giannouli et al. (2006); Geelen et al. (2009); González-Benito and González-Benito (2010); Liang et al. (2014).

The idea of a pressure indicator directly relates with the necessity of measuring the pressures on the environment exerted by the activities involved in the economic and the social development. Nowadays, as the population grows, climate change threats and economic activity spreads

*Corresponding author.
Email addresses: anna.mergoni@kuleuven.be (Anna Mergoni), giovanna.dinverno@kuleuven.be (Giovanna D'Inverno), laura.carosi@unipi.it (Laura Carosi)

irresponsibly, this necessity is even more compelling. To this purpose, the monitoring of water and waste sector is fundamental (Lombardi et al., 2019; Degli Antoni and Marzetti, 2019; Das et al., 2019). The water and waste sector responds to important social necessities (such as finding sufficient water sources and sustainable solutions to waste disposal) by involving high level of energy usage, pollution emission, physical infrastructure and financial inputs. For this reason the literature about water and waste services sustainability is wide (see Walter et al., 2009; Juwana et al., 2012; Simões and Marques, 2012; Worthington, 2014; Allesch and Brunner, 2014; Vilanova et al., 2015; Margallo et al., 2015; da Silva et al., 2020; Zeller et al., 2020, and the references therein).

Nevertheless, it seems that there is a lack of consensus on a widely accepted method to assess the environmental sustainability of water supply, collection and waste management services (Simões and Marques, 2012; Marques et al., 2015; Sala et al., 2015; Pérez et al., 2018). The major challenges can be attributed to the multidimensionality of the phenomenon and regard, on the one hand, the choice of the relevant indicators, on the other, the choice of a fair aggregating method.

As for the former challenge, several authors suggest different possible sets of indexes for measuring the environmental sustainability of the urban waste and water services. Despite the relevant differences among the suggested approaches, all the proposed criteria can be interpreted as sub-indicators of 'good' or 'bad' pressures (Marques et al., 2015; Molinos-Senante et al., 2016; Pinto et al., 2017; Pérez et al., 2018).

As for the second challenge, i.e. the choice of a fair aggregating method, a wide variety of methodologies is available. Data Envelopment Analysis (DEA) and DEA-like approaches are among the most used ones (see Romano and Guerrini, 2011; Molinos-Senante et al., 2017; Marques et al., 2018; Caldas et al., 2019, among others).

The Composite Indicator (hereafter CI) implemented in this paper belongs to this family and is based on the model by Zanella et al. (2015), which Rogge et al. (2017) defined as a directional distance version of the Benefit of the Doubt (BoD) model. Such approach allows us to select the benchmarking units in a completely data driven way, to evaluate the utilities along desirable and undesirable dimensions and to ensure the best possible rank to each unit. To obtain information about the weakest environmental areas and their potentially harmful impact, we complement this traditional (optimistic) approach with a pessimistic version of the BoD model. Besides, in its robust and conditional form, the directional distance BoD model allows to account for the possible presence of outliers and to ensure a context-unbiased evaluation.

We add to the previous literature with a number of contributions. First, we develop a pressure indicator to evaluate the utilities jointly and simultaneously active in the areas of water supply, water collection and waste management. In particular, our indicator evaluates the utilities according to their ability of reducing the resource usage, the release of noxious substances in

the environment, in line with the indication provided by OECD Environment Directorate (2008); OECD (2020) and Dong and Hauschild (2017). Second, from a theoretical perspective, we complement the CI proposed by Zanella et al. (2015) in two ways. On the one hand, we introduce the formulations of its pessimistic version, following insights from Zhou et al. (2007) and Rogge (2012). On the other hand, we use the robust and conditional analysis introduced by Cazals et al. (2002), following the path of Rogge et al. (2017), Lavigne et al. (2019), Fusco et al. (2020) and D'Inverno et al. (2020). To do so, a revision of the definition of the CI has been necessary. Third, we implement the suggested approach to the Portuguese case. By evaluating the entities that are active both in the water and waste sectors, we account for the possible interactions and synergies that may occur in the joint management of these sectors.
To the best of authors knowledge, there is no previous study which accounts for these two sectors together, in the framework of environmental performance. There are only a few studies that treat the water and waste sector jointly, specifically Bel and Warner (2008) and Caldas et al. (2019). However these papers focus on the economic aspect, respectively the presence of privatization and of scale economies, and the environmental perspective is not considered. Finally, from a policy perspective, by benchmarking in comparative terms the utilities, we promote information exchange and encourage the imitation of the best performing practices.

The rest of the paper is organized as follows. In section 2 we briefly justify the choice of Portugal and we present the data. In section 3 we present the methodology and the path that brought us to the choice of the directional distance BoD CI, both in its optimistic and pessimistic formulation, and to its implementation in a robust and conditional framework. In section 4 we report and comment the obtained results. Section 5 concludes the paper with some final remarks on the policy relevance of the proposed tool.

2. Empirical Framework and Data

2.1. Water and Waste in Portugal

The idea of measuring the environmental pressure of waste and water utilities is implemented by looking at the Portuguese case. In this country the system for water supply, water collection and waste management shows a number of relevant characteristics that have drawn the attention of many scholars, generating a flourishing scientific debate (see among others the works of R.C. Marques, A.P. Antunes, M.C. Cunha and the recent papers by Martins et al. 2020; Henriques et al. 2020; Marques and Simões 2020; Silva and Rosa 2020). From a juridic perspective, private, state and municipal owned utilities coexist. These utilities operate in the water supply, water collection and waste management and present a strong interdependence among the three main areas. From an environmental perspective, the peculiarities of these sectors make Portugal an interesting laboratory for testing the suggested composite indicator.

First, these sectors are vulnerable. Portugal is prone to seasonality, with abundance of water in winter and scarcity in summer, especially in the south and it is suffering climate change, which is impacting the quality and the availability of surface and underground drinking water sources, with serious consequences for the water provision (Serra et al., 2021; EurEau Association, 2021). Besides, the economic growth has increased in absolute and in relative terms the waste production (Kaza et al., 2018).

Second, these sectors are dynamic and constantly evolving. During the last decades Portugal has committed considerable resources in the mentioned sectors yielding an increasing attention of the public debate and a positive thrust to the quality and the coverage of the offered services. While in 1994 the coverage for the services of water supply, water collection and waste management was, respectively, 81.5%, 60.7% and 98%, nowadays it increased up to the 96%, the 85% and the 100% (for the Portuguese mainland), corresponding to 9.6, 8.6 and 10 million of inhabitants.

Third, in Portugal the water and waste sectors are deeply integrated. Though they involve three distinct macro-areas - water supply, water collection and waste management - they are regulated and supervised by the same authority, and often they are managed by the same entities. Specifically, 48% of the utilities are jointly active in the three macro-areas.

Fourth, the Entidade Reguladora dos Serviços de Água e Resíduos (ERSAR - Regulatory Entity for Water and Waste), created in 1997 under the name of IRAR, is the fundamental body for the strategic decision-making planning and the management of water supply, collection and waste management. ERSAR acquired its regulatory power in 2009 and become an independent administrative entity in 2014, however since 2004 it is responsible for the performance assessment and benchmarking of the utilities active in the sector. This responsibility has two direct consequences: first, by evaluating the quality of the utilities, ERSAR implicitly decides which are the important criteria to be assessed and the target to be reached (Gonçalves et al., 2014); second, ERSAR collects the necessary data to analyze the performance of the utilities.

Fifth, these sectors are increasingly involved in the environmental cause, by addressing circular economy strategies and including waste recycling in agriculture (Serra et al., 2019). At the beginning of the new millennium, Portugal faced the challenge of increasing the coverage and improving the performance of these services (Correia and Marques, 2010; Marques et al., 2018). Today, the challenge is to protect their sustainability by providing and implementing solutions to minimize the negative impact on the environment and to ensure the continuous supply of high quality water, the collection and treatment of wastewater and to reduce the amount of waste, for present and future generations (UN General Assembly, 2015).

2.2. The Data

The database at our disposal contains information about the whole population of water supply, collection and waste management utilities in Portugal mainland in 2018 (ERSAR, 2018).

We restrict our focus on the retail utilities simultaneously active in the three macro-areas, i.e., on the utilities providing jointly the three services of water supply, water collection and waste management for the households. This allows us to construct a comprehensive indicator which fulfills the homogeneity assumption (see Dyson et al., 2001, p. 247), since all the units in our sample have similar productive processes. In Portugal there are 180 utilities active in the three sectors, however it was possible to include in the analysis only the 149 units which provided sufficient information along the dimensions of interest. The units employed for our analysis provide more than 223 billion m^3 water per year, collect almost 279 billion m^3 wastewater per year and collect more than 2 million tons of urban waste per year, providing the services of water supply and waste management, respectively, to 2,207,000 and 2,240,000 of households.

Among the sub-indicators collected by ERSAR, four have been selected to measure the pressure on the environment by water and waste utilities: 1) water losses 2) structural collapses 3) gas emission and 4) recycled waste. The choice of these indicators has been driven by the idea of accounting for the pressure (in terms of release of substances) that the water and waste utilities exert on the environment (Marques et al., 2015; Molinos-Senante et al., 2016; Pinto et al., 2017; Pérez et al., 2018). This leads us to the choice of our four sub-indicators. These indicators comprehensively represent the multidimensional environmental pressure framework. Moreover, we remark that the inclusion of less informative sub-indicators would be paid by the exclusion of several units due to missing values, without changing drastically the main findings (see also Henriques et al., 2020).

The first and the second criteria, water losses and structural collapses, are indicators of bad pressure. Uncontrolled water release is bad for the environment on different levels. First, it promotes soil erosion, which is one of the greatest environmental threats to sustainability (Zhu et al., 2019). Second, it is associated with leaching and nutrient loss, leading to groundwater contaminations with nitrate and other soluble compounds (Serra et al., 2019). Then, water quality also has an effect on soil quality, modifying soil conditions and altering mineral nutrition (García et al., 2008). The third indicator, gas emission, is also an indicator of the bad pressure exerted by the utilities on the environment in the form of release of greenhouse gas (ERSAR, 2020). It refers to the total amount of CO2 emissions from undifferentiated collection vehicles per ton of waste collected in the management area. The last criteria, i.e. recycled waste - criterion 4, instead, is a measure of positive pressure exerted by the utilities on the environment, if properly managed. By recycling the waste collected from the households, the utilities control and prevent an otherwise inevitable release of polluting substances, as long as duly managed.

As Serra et al. (2021) report, Portugal mainland is characterized by considerable heterogeneity in terms of climate, orography and land use. To account for the possible impact of these external factors on the behaviour of the utilities, we implemented a conditional analysis. Specifically, three control variables have been selected as possibly influential external variables: 1) geographical position 2) intervention area and 3) volume of water supplied. These variables do not directly enter in the construction of the composite indicator, but they might still affect the assessed environmental pressure of the utilities. Specifically, their location is directly related to their service provision since water utilities operate as natural monopoly. Volume of water supplied is used as a proxy for the size (note that volume of water supplied is highly correlated with the volume of water and waste collected). Similarly to what happens for the economic assessment, the size might influence also the environmental pressure. The urban areas reveal different needs and challenges with respect to the rural or the semi-urban ones, especially from an environmental perspective.

For more details on the definitions of the sub-indicators and the control variables see tables 1 and 2. As it can be noticed, the variables are measured in different scale, but this is not an issue as the implemented methodology does account for this.

Table 1: Definition of the environmental pressure sub-indicators chosen to construct the composite indicator.

Sub-indicator	Pressure	Definition
Real water losses	Bad	The volume of actual losses per unit length of conduit in a day, measured in volume of losses / connections in a day. ERSAR database code: AA12b.
Structural collapse	Bad	The number of structural collapses in 100 km of collectors in a year. ERSAR database code: AR08b.
Gas emissions	Bad	Total amount of CO2 emissions from undifferentiated collection vehicles per ton of waste collected in the management area of the management body. ERSAR database code: RU17b.
Recycled waste	Good	Ratio among the ton of waste recycled and the target ton of waste recycled in the year. ERSAR database code: RU07b.

Table 3 shows that there is heterogeneity among the units located in different intervention area and with different volumes of activity, especially for the indicators gas emission and recycled waste. Instead, the differences along the geographical position are not significant. Specifically, it emerges that, considering the variable intervention area, the units located in urban and semi-urban emit, on average, less gas, while the units located in rural area produce, on average, more gas. This pattern can be explained by the fact that in rural areas the households are located further one to the other, so that the companies are more prone to cover longer distances to deliver the services, and therefore, to emit more gas. According to the volume of activity, small

[1] See `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016DC0666&from=EN`

Table 2: Control variables used in the conditional analysis to account for possible heterogeneity of the context where the utilities operate.

Control variable	Definition
Geographical position	Portugal is divided into five macro regions: the region of the North, the region of the Centre, the region of Lisbon, Alentejo and Algarve. In line with ERSAR reports, we consider just three macro domains: the North, equivalent to the region of the North, the Centre, equivalent to the region of the Centre plus the Lisbon district, and the South, composed by Alentejo and Algarve.
Intervention area	We consider the typology of areas according to the definition of the Deliberations n. 488/98 and n. 2717/2009, followed also by the Portuguese national institute of Statistics. Three intervention areas are identified: *predominantly rural areas*, *medium urban areas* and *predominantly urban areas*.
Volume of activity	Volume of water (in m3) supplied in a year. The 'Drinking Water Directive' (Council Directive 98/83/EC) distinguishes between large and small water utilities: '*large water supplies provide either more than 1,000* m^3 *drinking water per day as an average or serve more than 5,000 persons*' [1]. For the present application we refer to the volume of water.

units do significantly worse than the average in the emission of gas and in the recycling waste, while the large units do significantly better than the average according with these indicators.

Figure 1 complements Table 3 by showing the geographical variability of the four environmental pressure sub-indicators over the Portuguese territory. Inspired by the ERSAR reports, we display in red the utilities that are exerting a high (negative) and so an unacceptable level of pressure, in yellow a medium level and in green a low and so a less urgent level. From a policy-making perspective, choosing to address one issue, e.g. the water losses, might lead to overlook utilities unsatisfactorily performing in other domains.

Table 3: Mean distribution of the sub-indicators in 2018.

	N	Coverage	Water loss l/day		Structural collapse (n/km.year)		Gas emission kg (CO2/t)		Recycled waste (%)	
Overall	149	83%	146.1		3.82		20.15		86.58	
	min		1.8		0		6		28	
	max		502		173		52		281	
Geographical location										
North	39	81%	158.8		4.308		20.87		81.49	
Centre	60	90%	141.51		5.480		20.90		81.63	
South	50	78%	141.84		1.438		18.7		96.5	
Intervention area										
Rural	112	80%	135.42		2.97		21.98	*	82.6	
Semi-urban	29	97%	181.6	*	5.76		15.28	***	103.3	*
Urban	8	89%	167.88		8.575		12.25	***	81.88	
Volume of activity										
Small	53	75%	125.8		5.113		25.08	***	77.25	**
Medium	48	87%	162.4		0.51	**	19.9		83.77	
Large	48	97%	152.4		5.69		14.98	***	99.71	**

Note: The significance of the difference between the overall distribution and the distribution per groups has been computed through the *t*-test. *p<0.1; **p<0.05; ***p<0.01.

Source: Authors' own elaboration based on data from ERSAR.

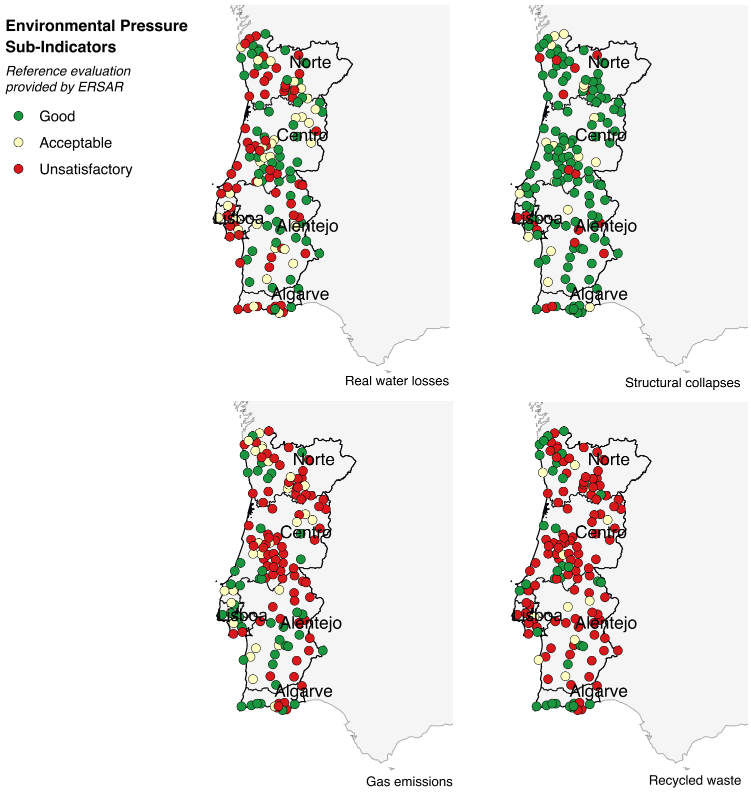

Figure 1: Geographical distribution of the environmental pressure sub-indicators.

Source: Authors' own elaboration based on data from ERSAR relative to 2018.

3. Methodology

A crucial issue in the construction of Composite Indicators (CIs) is the aggregating method as, in most cases, there is only disparate expert opinion available about the appropriate weights to be used in the aggregator function. The Benefit of the Doubt (BoD) approach, presented by Melyn and Moesen (1991) and then popularized by Cherchye et al. (2007), allows to overcome this problem. It endogenously assigns weights so that the overall score depicts each analyzed decision making unit (DMU) in the best possible light relatively to the other observations. So every DMU is granted with the *'Benefit of the Doubt'* and the approach is strongly data oriented. These two qualities explain a major part of the appeal of the BoD-based CIs in real settings.

3.1. The traditional BoD model: An optimistic approach

The BoD approach has its root in the Data Envelopment Analysis (DEA) model of Charnes et al. (1978); it actually can be seen as an input-oriented DEA model with unitary input and the sub-indicators as outputs. Therefore, we can translate also the interpretation of the score: a good relative performance of a DMU, in one particular sub-indicator, indicates that the evaluated unit considers that specific dimension as relatively important.

The value of the performance is obtained by aggregating all the sub-indicators values, weighting them in the most convenient way for the unit under analysis, subject to two constraints: 1) the weights have to be positive and, 2) the value of the CI, for no unit in the sample can exceed a given threshold (usually fixed at 1).

The BoD model has been designed to deal with 'desirable' sub-indicators (meaning the higher the better). Nevertheless, it may occur that some relevant dimensions of the analyzed units are described by means of 'undesirable' sub-indicators (meaning the lower the better). Whenever both 'desirable' and 'undesirable' sub-indicators are considered, the standard BoD model cannot be applied. To overcome this drawback, Zanella et al. (2015) propose an alternative formulation on the basis of the directional distance function approach of Chung et al. (1997). Like the BoD model, a dummy input is fixed at a unitary level and like the directional distance function models, a suitable directional vector g is considered to allow the simultaneous contraction of the undesirable indicators and expansion of the desirable ones. According to Zanella et al. (2015)(p.523), in this paper, CIs are computed by solving the following maximization problem:

$$\begin{cases} \max & \beta \\ s.t. & \sum_{j=1}^{n} b_{kj}\lambda_j \leq b_{kj_0} - \beta g_b, \qquad k = 1, ..., l \\ & \sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rj_0} + \beta g_y, \qquad r = 1, ..., s \\ & \sum_{j=1}^{n} \lambda_j \ = \ 1 \\ & \lambda_j \ \geq \ 0, \qquad j = 1, ..., n \end{cases} \tag{1}$$

where l, s, n respectively represent the number of undesirable sub-indicators, the number of desirable ones and the number of DMUs, respectively; j_0 is the evaluated DMU, b_{kj} is the value of the undesirable sub-indicator k of the unit j, while y_{rj} is the value of the desirable sub-indicator r of the unit j. The vector $g = (-g_b, g_y)$ represents the direction along which the simultaneous contraction of the undesirable indicators and expansion of the desirable ones is possible. The current literature proposes alternative directions values: for example, directions values equal to one, $g = (1, 1)$; equal to the current indicator values of the unit under evaluation, $g = (-b_{kj_0}, y_{rj_0})$; or equal to the average values across all the units under analysis, $g = (-\bar{b}_k, \bar{y}_r)$ (for a further discussion, we refer to Rogge et al., 2017). Different directions give rise to different interpretations. For our model we choose $g = (-g_b, g_y) = (-b_{kj_0}, y_{rj_0})$, so that each utility follows its own improvement path and a great level of flexibility and proportional interpretation of the results are granted. β is the value of the directional distance function for the evaluated DMU and it measures the room of possible improvements along the direction g; the optimal value of the problem, β^*, belongs to $(0, +\infty)$ and, accordingly, the associated CI is defined as $\frac{1}{1+\beta^*}$. This formulation allows to 'control' the value of the β^*, so the value of the CI belongs to (0, 1]. The higher the value of the CI, the closer the DMU is to the best-practice frontier. DMUs on the frontier assume a CI = 1 (see also Zanella et al., 2015; Rogge et al., 2017; Lavigne et al., 2019).

3.2. A complement to the traditional BoD: A pessimistic approach

By construction, the weights assigned by the traditional Benefit of the Doubt allow to evaluate each utility under the best possible light. This is obtained by overemphasizing the dimensions where the units perform the best and mostly neglecting where they perform the worst. This endogenous weighting mechanism grants a fair evaluation and mostly avoids complaints among the evaluated units. In spite of the fairness granted to the utilities under evaluation, the BoD analysis might overlook very poor performances along some dimensions, thus cannot be completely informative from an environmental footprint perspective and might suggest inappropriate policy measures. To avoid this issue, we complement the traditional (optimistic) evaluation with the so-called "pessimistic" version of the BoD model (Dardha and Rogge, 2020). From an intuitive point of view, the pessimistic approach evaluates how close is each utility to the worst performing utilities in the sample under the least favorable evaluation conditions, that is, assigning high weights on areas where the utility exerts a relatively high environmental pressure level and low weights where it exerts relatively low environmental pressure level (Rogge, 2012).

To the best of the authors' knowledge, this is the first application of the pessimistic scenario adapted to the main model proposed by Zanella et al. (2015) following insights from Zhou et al. (2007) and Rogge (2012). The problem (1) adjusted for its pessimistic counterpart then becomes the following (in Appendix A we provide also the multiplier formulation for both the optimistic and the pessimistic directional distance BoD model):

$$\begin{cases} \min & \beta_P \\ s.t. & \sum_{j=1}^{n} b_{kj}\lambda_j \geq b_{kj_0} - \beta_P g_b, \qquad k = 1, ..., l \\ & \sum_{j=1}^{n} y_{rj}\lambda_j \leq y_{rj_0} + \beta_P g_y, \qquad r = 1, ..., s \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j \geq 0, \qquad j = 1, ..., n \end{cases} \tag{2}$$

where l, s, n respectively represent the number of undesirable sub-indicators, the number of desirable ones and the number of DMUs, respectively; j_0 is the evaluated DMU, b_{kj} is the value of the undesirable sub-indicator k of the unit j, while y_{rj} is the value of the desirable sub-indicator r of the unit j.

Coherently with the optimistic version, we set $g = (-g_b, g_y) = (-b_{kj_0}, y_{rj_0})$. Intuitively, this means moving along the direction opposite to the optimistic one or, in another way, the direction along which the simultaneous expansion of the undesirable indicators and contraction of the desirable ones is possible, so to reach the worst-case scenario. β_P is the value of the directional distance function for the evaluated DMU. In the pessimistic case, the optimal value of the problem is non-positive, as β_P^* belongs to $(-1, 0]$. Hence, the worst performing units assume $\beta_P^* = 0$, while β_P^* tends to -1 for the least worst performing units. Similarly to the optimistic case, the associated CI_P is defined as $\dfrac{1}{1+\beta_P^*}$. Accordingly, the value of the CI_P belongs to $[1, +\infty)$. The lower the value of the CI_P, the closer the DMU is to the worst-case scenario. The worst-performing DMUs assume $CI_P = 1$ (see also Zhou et al., 2007; Rogge, 2012).

3.3. Beyond the deterministic nature of BoD: A robust and conditional approach

Previous literature (see e.g. Nardo et al., 2005, or Daraio and Simar, 2007) highlighted some typical limitations related to the use of a non-parametric approach. In particular, the deterministic nature of the CI leads to three issues: 1. statistical inference is difficult, 2. the scores are sensitive to outliers and 3. to the sample size. To face these problems, we complement the model with a robust and a conditional analysis, by applying the methodology proposed by Cazals et al. (2002) and by Daraio and Simar (2005, 2007) (see also Rogge et al., 2017; Fusco et al., 2020; Lavigne et al., 2019; D'Inverno et al., 2020).

The robust evaluation of Cazals et al. (2002), also called 'order-m', consists of a Monte Carlo simulation. Each DMU is evaluated B times with respect to m units randomly drawn with replacement from the original sample Γ (with $n > m$). This allows to control for extremes and outliers.

B sub-samples $\Gamma^{b,m}$ are generated for each DMU_{j_0} under analysis and B scores are calculated. $\beta^{b,m}$ is the directional distance BoD score computed for the DMU_{j_0}, using the b^{th} sub-sample of dimension m. Therefore, in the robust version of Model 1, a given DMU j appears in the constraints only if $j \in \Gamma^{b,m}$. Once obtained the B $\beta^{b,m}$ coefficients, we define $\beta^m = \sum_{b=1}^{B} \dfrac{1}{B}\beta^{b,m}$.

It is important to notice that the DMU_{j_0} under analysis may not be drawn in the sub-sample used as reference set. For this reason each $\beta^{b,m}$ belongs to $\mathbb{R}$ (and so does β^m). The more negative the β^m the further the DMU$_{j_0}$ over the frontier, in the sense that it is performing better. If $\beta^m < 0$ the DMU is referred to as *super-performing*[2]. Since $\beta^m \in \mathbb{R}$, the previous formulation of the CI loses its explanatory power. This is due to two main reasons: first, the function $CI(\beta^m) = \dfrac{1}{1+\beta^m}$ is defined over $(-\infty, -1) \cup (-1, +\infty)$ and not over $\mathbb{R}$ (note that in the deterministic model this was not a problem as β^m belonged to $[0, +\infty)$. Second, interpretation problems arise for those DMUs having a value of β^m lower than -1; although they are super-performing, their corresponding CI is negative and, accordingly, they are judged worse than the bad performing ones, i.e. those with a high and positive value of β^m. To avoid these problems, we propose the following construction of the robust Composite Indicator:

$$CI^m(\beta^m) = \begin{cases} \dfrac{1}{1+\beta^m} & \text{if} \quad \beta^m \geq 0 \\ \log(1-\beta^m)+1 & \text{if} \quad \beta^m < 0 \end{cases} \tag{3}$$

The performance score $CI(\beta^m)$ is now defined over $\mathbb{R}$ and it is continuous and differentiable. As in the deterministic case, it is decreasing with respect to β and preserves the interpretation proposed by Daraio and Simar (2007): a value of $CI(\beta^m)$ greater than one indicates that the unit j_0 is better performing than the average of m peers randomly drawn from the population (p.71).

To properly account for the influence of the exogenous characteristics and therefore to ensure a fairer evaluation, we allow the benchmarking frontier to 'adapt' according to the exogenous characteristics of the unit under analysis, i.e., we adopt the conditional analysis (see developed in Cazals et al., 2002; Daraio and Simar, 2005, 2007; De Witte and Rogge, 2011). The basic idea is to condition the choice of the reference set for the DMU j_0 under evaluation according to its own exogenous characteristics. While in the robust scenario the units of the reference group $\Gamma_{j_0}^{b,m}$ are drawn with replacement from a uniform distribution, in the conditional case they are included in the reference group $\Gamma_{j_0}^{b,m,z}$ according to the probability of being similar to the observation j_0 (with $\Gamma_{j_0}^{b,m,z} = \Gamma_{j_0}^{b,m}|Z$). Similarity is measured by means of the probability distribution for the joint Z variables, estimated by a kernel function (see also De Witte et al., 2013; Li and Racine, 2003). Using a Monte Carlo simulation procedure, B sub-samples $\Gamma^{b,m,z}$ are generated for each DMU$_{j_0}$ and B $\beta^{b,m,z}$ are obtained by running model (1) considering only the DMUs belonging to $\Gamma_{j_0}^{b,m,z}$. Then, the mean of the obtained B values of $\beta^{b,m,z}$ is calculated, and the corresponding Composite Indicators $CI^{m,z}$ is computed by using function (3).

The interpretation of the conditional Composite Indicator CI^{m,z} has to go arm in arm with the comparison between this indicator and the robust one, namely CI^m. To investigate the source

[2]The terminology used in the literature is *super-efficiency*; since we refer to a composite indicator, we prefer to talk of 'performance' instead of 'efficiency, in line with Rogge et al. (2017) and Lavigne et al. (2019).

of the difference between them, the ratio $CI^m/CI^{m,z}$ is considered, as suggested by Daraio and Simar (2007)[3]. If the ratio is increasing along the environmental variable, it means that this variable has a positive influence on the performance of the utilities we are measuring. Vice versa a decreasing ratio shows an unfavorable environment. We regress the ratio of the robust over the conditional on the environmental variables using a non-parametric regression (as suggested by Daraio and Simar, 2007 page 113):

$$\frac{CI^m}{CI^{m,z}} = g(Z_i) + \epsilon_i, \quad i = 1, ..., n.$$

4. Results and Discussion

The environmental pressure index was computed for 149 utilities that provide both waste and water services in Portugal. For the estimation we followed the methodology described in the previous section, so to get an aggregate indicator that measures how well the operators are coping with the environmental pressure they exert on the environment. First, we explored the obtained findings for the deterministic case. Second, we explored the results considering the optimistic and the pessimistic environmental scenario. Third, we gave insights on the robust and the conditional analyses. Finally, we investigated the influence of the operating context through statistical inference.

4.1. Results from the traditional BoD model

Table 4 shows the descriptive statistics of the environmental pressure Composite Indicator (CI) scores for the deterministic case. The mean value of 0.7398 suggests that there is room for improvements in environmental pressure reduction if all the entities would perform on the four sub-indicators as well as the best performing entities. The minimum value of 0.5819 together with the first quartile of 0.6614 denotes the widespread presence of poorly performing operators, i.e., operators which are outperformed despite being evaluated in the most favorable way along different measures of environmental pressure. Previous literature had already detected the need for a performance enhancement of the Portuguese water and waste sectors (see among others Ferreira da Cruz et al., 2012; Marques et al., 2015; Molinos-Senante et al., 2016; Pérez et al., 2019). Our findings complemented this evidence by giving specific emphasis on the environmental sustainability issue and particularly from an environmental pressure perspective.

We identified 11 best performing operators (CI = 1) out of the 149 in the sample. This means that a relatively small percentage of our sample (7.38%) can be considered as best practice for the others that report CI scores lower than one. We also explored the characteristics of these units by looking at the distribution of the CI along the operating context variables introduced in section 2, namely the geographical location, the area of intervention and size. At first sight, the

[3]Daraio and Simar (2007) use the inverse of this ratio. The reason of our choice is that it simplifies the interpretation of the estimated relationships (Rogge et al., 2017).

utilities that report the highest mean and median values are located, more likely, in the South of Portugal, or in areas predominantly urban, or they are large.

Table 4: Descriptive statistics of the environmental pressure composite indicator scores (both overall and grouped by operating context variables). The scores are obtained implementing the deterministic and unconditional analysis.

	N	Mean	SD	Min.	Q1	Median	Q3	Max.
Deterministic unconditional	149	0.7398	0.1156	0.5819	0.6614	0.6993	0.7914	1.0000
Geographical location								
North	39	0.7281	0.1192	0.5819	0.6500	0.6914	0.7600	1.0000
Centre	60	0.7288	0.1098	0.5897	0.6495	0.6956	0.7758	1.0000
South	50	0.7622	0.1186	0.5915	0.6790	0.7030	0.8652	1.0000
Intervention area								
Rural	112	0.7281	0.1133	0.5819	0.6563	0.6921	0.7587	1.0000
Semiurban	29	0.7680	0.1158	0.6223	0.6652	0.7594	0.8156	1.0000
Urban	8	0.8018	0.1235	0.6364	0.7416	0.7864	0.8626	1.0000
Volume of activity								
Small	53	0.6996	0.0954	0.5819	0.6463	0.6786	0.7022	1.0000
Medium	48	0.7272	0.1094	0.5915	0.6604	0.6942	0.7540	1.0000
Large	48	0.7969	0.1214	0.6223	0.6890	0.7719	0.8986	1.0000

4.2. The environmental pressure index in an optimistic and pessimistic scenario comparison

To provide a more comprehensive picture of the environmental pressure exerted by the Portuguese utilities jointly operating in the three sectors, we complement the results obtained using the traditional/optimistic BoD approach with a pessimistic one. In the former we give more emphasis on the areas where utilities are exerting a relatively low pressure level compared to the other utilities, highlighting the best scenario. In the latter we obtain information on how well they are performing despite the least favorable evaluation, outlying the worst scenario.

Figure 2a shows in a synthetic way the results obtained in these two opposite scenarios (we report in Appendix B the descriptive statistics of the pessimistic environmental pressure composite indicator scores). Utilities with a CI lower than 1 are the ones displaying a low performance, despite being evaluated in their most favorable scenario. Utilities with a CI_P equal to one are the ones performing weakly in the majority or even all the dimensions.

Following Rogge (2012), we can distinguish three groups of utilities based on their CI and CI_P. The first group is characterized by an *overall good* environmental pressure level, with

$CI = 1$ and $CI_P > 1$. The utilities in this group perform well both under the optimistic and the pessimistic scenario. Thus, they don't show a peculiar specialization on a particular area, but they perform relatively strongly compared to the other utilities, in all, or almost all, the environmental pressure sub-indicators considered. From a policy-making perspective, these operators (CM de Ansião, CM de Évora, CM de Ferreira do Zêzere, CM de Melgaço, CM de Óbidos, CM de Ponte de Lima, CM de Póvoa de Varzim, CM de Santiago do Cacém, INFRAQUINTA, INFRATRÓIA, SM de Castelo Branco) are the best practices that the other utilities should look at to reduce their environmental pressure or that show extremely outstanding performance. This is for example the case of the operator INFRAQUINTA, that reports one of the highest CI_P scores, $CI_P = 5.3368$, and $CI = 1$. This operator has been already identified in other studies as one of the Portuguese top performing utilities (see Molinos-Senante et al. 2016 and Henriques et al. 2020). This can be seen as an example of utility that promotes environmental sustainability and tackles environmental pressure in water supply, wastewater sanitation and urban waste management sectors as a public commitment (see `https://www.infraquinta.pt/en/empresa/activities-plan`). The exceptional good performance of this unit can be partly explained by its recent re-organisation and the relatively modern infrastructures, which create also the expectation of future investment return (Henriques et al., 2020).

The second group is characterized by an *overall mediocre* performance, with $CI < 1$ and $CI_P > 1$. The utilities in this group do not perform as good as the best practices, but also not as bad as to be considered the worst performing units. In this sense, they might have focused their effort on a specific sector or only on a few ones to deal with the environmental pressure they exert. Regarding this group, policymakers should pay attention to the dimensions mostly left behind and provide incentives for their improvements. The third group is characterized by an *overall poor* performance, with $CI < 1$ and $CI_P = 1$. In an environmental perspective, these utilities should be the first ones to be looked at, since they exert the highest level of environmental pressure. From a policy perspective, the goal is not to 'name and shame' these utilities (Cabus and De Witte, 2012), but rather to identify them and support them, as they are the detected most harmful ones for the environment. As a last remark we point out that there are no utilities with $CI = 1$ and $CI_P = 1$, ruling out the presence of extreme scenarios with excellent performance on one dimension and very poor performance on another one at the same time.

Figure 2b-d show the distribution of the utilities by the operating context variables and by their performance level. Most of the utilities belong to the *overall mediocre* performance group. From the analysis of the weights and the contribution of each element to the composite indicator, we can observe that in the optimistic case relatively good performance is related to the water loss and gas emission indicators, while in the pessimistic case the most critical component is the level of recycled waste, confirming the intuition we get from Figure 1. While for the geographical location and the intervention area there is no clear evidence of best practices, the volume of activity suggests that the small utilities are overall the worst performing ones. These utilities face huge costs to reduce their environmental impact in the three sectors and diseconomies of scale and scope arise. Policy makers should monitor more closely their activity and generate an incentive scheme to reduce their environmental footprint.

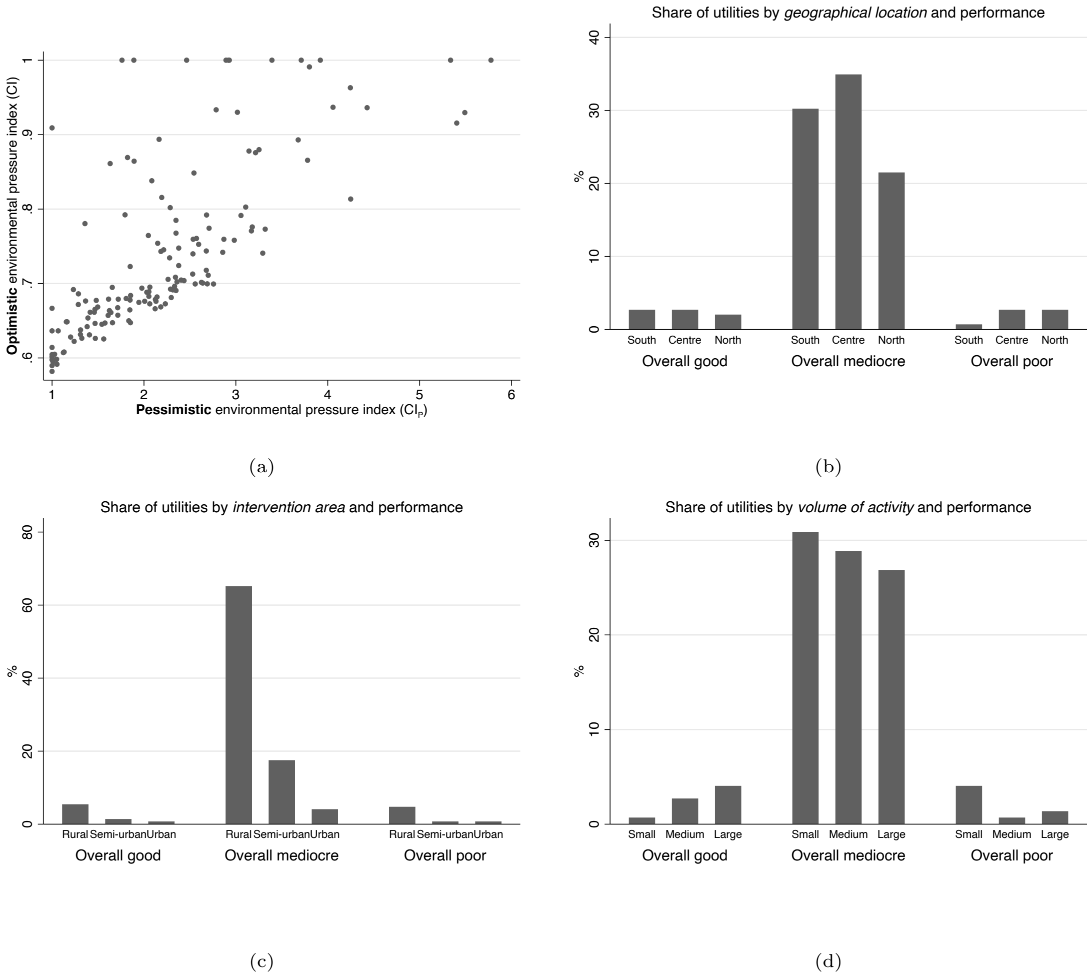

Figure 2: Comparison of environmental pressure performance in an optimistic and pessimistic scenario (Note: Overall good performance for $CI = 1$ and $CI_P > 1$; Overall mediocre performance for $CI < 1$ and $CI_P > 1$; Overall poor performance for $CI < 1$ and $CI_P = 1$).

Source: Authors' own elaboration.

4.3. The environmental pressure index accounting for outliers and exogenous characteristics

To account for the possible presence of atypical observations and to properly detect the influence of the exogenous characteristics, we estimated the traditional model in its robust unconditional and conditional version, so that unit's performance was assessed in a fairer way. Table 5 shows the descriptive statistics of the environmental pressure Composite Indicator (CI) scores for these two cases together with the deterministic case, for comparison purposes. To implement the order-m directional distance BoD model, a value for m must be chosen. A recipe to choose the suitable value of m does not exists, however we followed the procedure suggested by Daraio and Simar (2007), p. 78 - 81. Accordingly, a value of m equal to 65 seemed the most appropriate choice. Both the robust unconditional and conditional estimates display higher CI scores with respect to the deterministic case in all the summary statistics.

These estimations yielded CI scores greater than one in the upper part of the score distribution. This denotes the presence of super-performing units, i.e., units performing better than the average units they are compared with. Moreover, from the comparison of the median and the mean values, we also notice that the distribution of the conditional scores has a fatter right tale than the robust unconditional one, suggesting that the majority of the units are working in an unfavorable context. Nevertheless, there are units that are still very poorly performing as pointed out by the minimum value of 0.5930 for the robust unconditional case and 0.6034 for the conditional one.

The three estimated environmental pressure indexes are quite highly correlated (0.9688, 0.8259, 0.8623) and the distribution of the units among the observed background characteristics display a pattern similar to the one described for the deterministic unconditional case.

Figure 3 shows the geographical distribution of the estimated efficiency scores. The three CI scores display a similar pattern, confirming that the potential presence of outliers or different operating contexts do not significantly affect the outlined trend. An interesting feature of the environmental pressure Composite Indicator suggested in this paper is that it allows to aggregate different dimensions in a fully data driven way. Therefore, with a single glance we are able to identify the most critical areas, beyond the partial view offered separately by each sub-indicator as presented in section 2.

Table 5: Descriptive Statistics of the environmental pressure Composite Indicators scores for different model specifications.

	N	Mean	SD	Min.	Q1	Median	Q3	Max.
Deterministic unconditional	149	0.7398	0.1156	0.5819	0.6614	0.6993	0.7914	1.0000
Robust unconditional	149	0.8319	0.2272	0.5930	0.7029	0.7445	0.8749	1.8229
Robust conditional	149	0.8486	0.1604	0.6034	0.7371	0.8098	0.9867	1.7203

The comparison between the robust unconditional and conditional CI scores allowed us to detect the influence of the background characteristics on the estimated level of environmental pressure. Preliminarily, the *Kolmogorov - Smirnov* test was implemented to test if the difference among the conditional and the robust CI scores is statistically significant. The obtained p-value (0.0003338) provided a strong evidence in favor of this hypothesis. We focused on the partial regression plots reported in Figure 4 to investigate the source of this difference. The background

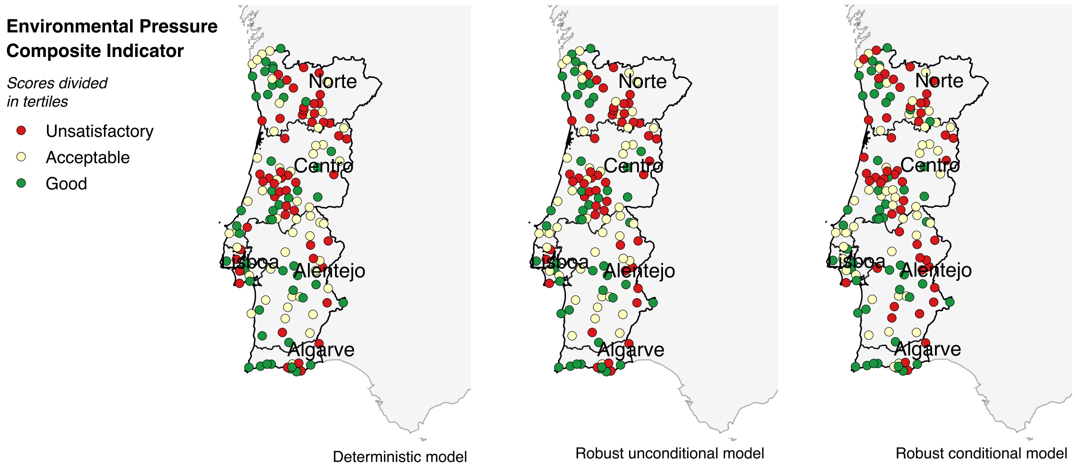

Figure 3: Geographical distribution of the environmental pressure composite indicator.

Source: Authors' own elaboration.

variables were regressed on the ratio between the robust unconditional and conditional, following the insights provided by Daraio and Simar (2007). If the ratio is increasing along the background variable, it means that this variable has a positive influence on the performance of the utilities we are measuring and the opposite holds otherwise.

We observed that the size and the area of intervention display a statistically significant relationship with the score ratio. Specifically, we observed a reversed U-shaped relation between the size and the estimated environmental pressure composite indicator, suggesting the potential presence of an optimal size. Previous literature on economies of scale and scope investigated the possible existence of an optimal size both for the water and waste sectors, concluding that a wide range of optimal scales can be detected and diseconomies in larger utilities can be found (see for example Simões et al., 2013; Carvalho and Marques, 2014, 2016; Caldas et al., 2019, and the references therein). Evidence from our empirical analysis suggests that whenever the utility is either too small or too large, it becomes difficult to contain the release of pollutants and high investments should be done to ameliorate the existing infrastructures or to increase their production capacity.

About the intervention area, the *predominantly rural areas* have a positive relationship with the environmental pressure management, while the opposite holds for *medium urban areas* and *predominantly urban areas*. To reconcile this evidence with the one stemming from the descriptive statistics, we consider that the higher scores of the urban utilities might be mostly driven by the size (as the urban utilities are also the larger). Besides, it could be noticed from Figure 3 that units located in the mountainous areas (therefore in the north or in the Serra da Estrela), on average, performed worse. A possible mechanism to explain this is that a steep terrain causes higher maintenance costs, therefore higher water losses, and higher transportation costs, therefore higher gas emissions and less waste recycling (see also Gaeta et al. (2017); Sarra et al. (2017)).

Finally, the geographical location did not display any particularly statistically significant association, even if it is still worth to be accounted for in the conditional estimation. Two factors mostly offset this evidence. From a territorial perspective, northern regions are mostly characterised by mountainous areas, while southern regions suffer particularly of seasonal imbalances and drought (Ferreira da Cruz et al., 2012; European Commission, 2014), causing again higher costs and difficulties in the process of collecting waste and treating water. From a regulatory perspective, recent changes of the social tariff regime also played a role in jeopardizing the equity, sustainability and territorial cohesion of this regulated sector (Martins et al., 2020).

The evidence stemming from the statistical inference offers an informative picture about the environmental pressure management in the Portuguese waste and water sectors. This can be considered as a starting point for further discussion to raise awareness among all the involved stakeholders on the environmental impact of these services' activity and to take action toward a more environmentally sustainable development (Molinos-Senante et al., 2016).

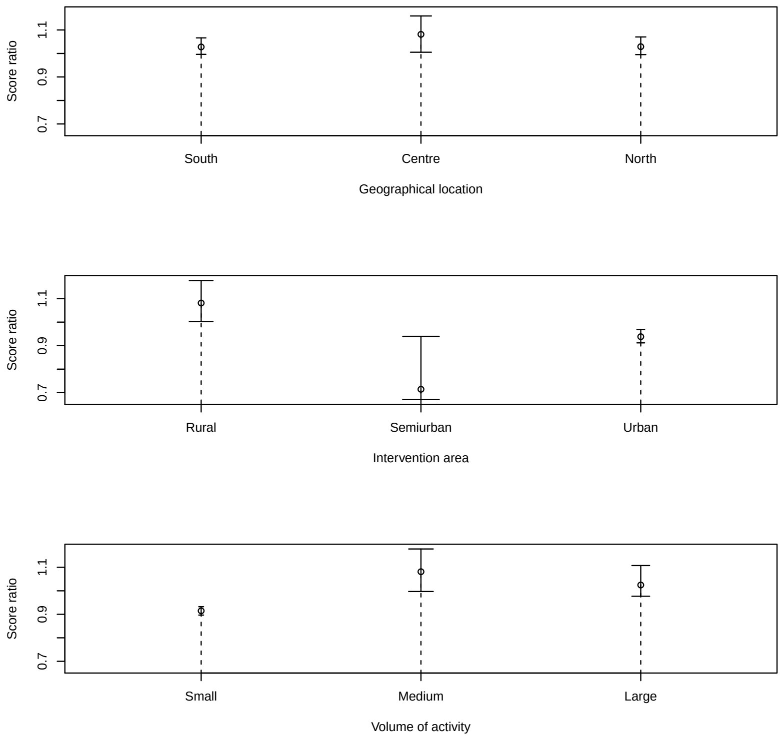

Figure 4: Visualization of the partial regression plots with confidence intervals for the operating context variables. A positive slope denotes a favorable influence on the environmental pressure composite indicator level, while the opposite holds for a negative slope.

Source: Authors' own elaboration.

5. Conclusions

The urgent need for an environmentally sustainable development calls for practical actions by managers and policy makers. However, finding good practices and selecting intervention areas are hard tasks for utilities that jointly cover services in different sectors, such as the water supply, the water collection and the waste management. This is because, for each service, they exert different levels of pressure on the environment, either in terms of substance release or in terms of resource use. As they all are potentially harmful for the environment, assigning an order of importance in an objective way becomes a tricky challenge.

We contribute to the literature by proposing a novel pressure composite indicator to measure and benchmark utilities active in different sectors and exerting different forms of environmental pressure. Specifically, we complement the use of a traditional directional distance Benefit of the Doubt composite indicator with its pessimistic version so to take into account the most harmful impact in the worst environmental scenario. In addition, we integrate the composite indicator with a robust and conditional approach so to account for the potential presence of atypical observations and the influence of contextual variables.

We test the proposed evaluation framework by evaluating 149 Portuguese utilities jointly active in the water supply, water collection and waste management sectors. In the annual reports, the Portuguese regulator (ERSAR) identifies room for improvement in any of the sub-indicators accounted in the proposed environmental pressure composite indicator, even suggesting potential ways to pursue it. With this respect, the beneficial feature of the proposed composite indicator is twofold. First, it detects the operators that exert the highest negative pressure on the environment encompassing the three services as a whole, granting them the most favorable assessment. Second, it suggests possible role models by looking at the best practices that emerge from the benchmarking exercise. On average, we find that there is room to alleviate the exerted environmental pressure and 11 utilities are detected as the best practises under both the optimistic and the pessimistic scenario. Most importantly, we are able to identify the utilities that are poorly performing in all the environmental dimensions. Ideally, all the utilities exerting a sizable pressure on the environment should be pushed to improve their pressure. However, in a context where there are limited resources and the measures to be taken are on a large (national) scale, the policy makers should start intervening in the areas where the environmental pressure is very critical, suggesting how to alleviate it by looking at the best practices observed from the analysis. Accordingly, we draw the national regulator's attention on the utilities with lower scores in both scenarios and whose background characteristics represent the most unfavorable environment. Certainly the background characteristics are variables that cannot be changed by the managers, especially in the short run, but empirical findings can direct the effort at national level and within the region. To this extent, we remark that the volume of activity plays a very significant role with respect to the environmental pressure. Specifically, small utilities are the most critical ones and the regulator should encourage shared service arrangement to seek increasing returns to scale and to invest more on their environmental sustainability. Furthermore, environmental best practices can be stimulated emphasizing the good performance signalling role of certifications, that currently are still not widely acknowledged (Molinos-Senante et al., 2016), as well as promoting more public commitment and transparency (Henriques et al., 2020).

The present analysis focuses on a cross-sectional dataset. Further research might explore the time component to check whether poor performing utilities are catching up with the best ones narrowing the gap and alleviating the overall environmental impact (Horta and Camanho, 2015; Henriques et al., 2020). Moreover, the main results point at the most critical areas, but additional analysis might follow to further explain the mechanisms and the hidden synergies behind the joint management of the three sectors (Caldas et al., 2019).

In this paper, the case of the Portugal has been presented to measure the environmental pressure of water supply, wastewater collection and urban waste management sectors. Given the worldwide relevance of the environmental sustainability and pressure, the proposed approach can be interestingly used for other countries and/or for other indicators to get useful insights where to intervene first. This will raise awareness of critical areas among the involved stakeholders and promote greater transparency in the environmental impact of the activities under scrutiny, to grant a sustainable development not only for the present generations but also for the future ones.

Acknowledgments

The authors gratefully acknowledge Ana Camanho, Kristof De Witte, Maria Molinos-Senante, Giulia Romano and Fritz Schiltz for their useful comments and suggestions on a previous version of this paper. Giovanna D'Inverno acknowledges financial support from FWO Postdoctoral Fellowship (Grant 12U0219N). Laura Carosi acknowledges financial support from University of Pisa (PRA 2018 Project).

References

Allesch, A., Brunner, P.H., 2014. Assessment methods for solid waste management: A literature review. Waste Management & Research 32, 461–473.

Bel, G., Warner, M., 2008. Does privatization of solid waste and water services reduce costs? a review of empirical studies. Resources, Conservation and Recycling 52, 1337–1348.

Cabus, S.J., De Witte, K., 2012. Naming and shaming in a 'fair'way. on disentangling the influence of policy in observed outcomes. Journal of Policy Modeling 34, 767–787.

Caldas, P., Ferreira, D., Dollery, B., Marques, R., 2019. Are there scale economies in urban waste and wastewater municipal services? a non-radial input-oriented model applied to the portuguese local government. Journal of Cleaner Production 219, 531–539.

Carvalho, P., Marques, R.C., 2014. Economies of size and density in municipal solid waste recycling in portugal. Waste management 34, 12–20.

Carvalho, P., Marques, R.C., 2016. Computing economies of scope using robust partial frontier nonparametric methods. Water 8(3), 82, 1–23.

Cazals, C., Florens, J., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1 – 25.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429 – 444.

Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T., 2007. An introduction to 'benefit of the doubt' composite indicators. Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement 82, 111–145.

Chung, Y., Färe, R., Grosskopf, S., 1997. Productivity and Undesirable Outputs: A Directional Distance Function Approach. Journal of Environmental Management 51, 229 – 240.

Correia, T., Marques, R.C., 2010. Performance of Portuguese water utilities: how do ownership, size, diversification and vertical integration relate to efficiency? Water Policy 13, 343–361.

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.

Dahl, A.L., 2012. Achievements and gaps in indicators for sustainability. Ecological indicators 17, 14–19.

D'Amato, D., Droste, N., Allen, B., Kettunen, M., Lähtinen, K., Korhonen, J., Leskinen, P., Matthies, B.D., Toppinen, A., 2017. Green, circular, bio economy: A comparative analysis of sustainability avenues. Journal of Cleaner Production 168, 716–734.

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: Methodology and applications. Springer Science & Business Media.

Dardha, E., Rogge, N., 2020. How's life in your region? measuring regional material living conditions, quality of life and subjective well-being in oecd countries using a robust, conditional benefit-of-the-doubt model. Social Indicators Research 151, 1015–1073.

Das, S., Lee, S.H., Kumar, P., Kim, K.H., Lee, S.S., Bhattacharya, S.S., 2019. Solid waste management: Scope and the challenge of sustainability. Journal of Cleaner Production 228, 658–678.

De Witte, K., Cabus, S., Thyssen, G., Groot, W., Maassen van den Brink, H., 2013. A critical review of the literature on school dropout. Educational Research Review 10, 13 – 28.

De Witte, K., Rogge, N., 2011. Accounting for exogenous influences in performance evaluations of teachers. Economics of Education Review 30, 641–653.

Degli Antoni, G., Marzetti, G.V., 2019. Recycling and waste generation: an estimate of the source reduction effect of recycling programs. Ecological Economics 161, 321–329.

Dong, Y., Hauschild, M.Z., 2017. Indicators for environmental sustainability. Procedia CIRP 61, 697–702.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C., Shale, E., 2001. Pitfalls and protocols in DEA. European Journal of Operational Research 132, 245 – 259.

D'Inverno, G., Carosi, L., Romano, G., 2020. Environmental sustainability and service quality beyond economic and financial indicators: A performance evaluation of italian water utilities. Socio-Economic Planning Sciences , 100852.

ERSAR, 2018. Dados de base. `http://www.ersar.pt/pt/setor/factos-e-numeros/dados-de-base`. Accessed: 2021-03-12.

ERSAR, 2020. Rasarp volume 1. `http://www.ersar.pt/pt/site-publicacoes/Paginas/edicoes-anuais-do-RASARP.aspx`. Accessed: 2021-05-11.

EurEau Association, 2021. Water matters. `http://www.eureau.org/`. Accessed: 2021-03-12.

European Commission, 2014. Potential for growth & job creation through the protection of water resources, with a special focus on the further implementation of the WFD and FD, issued under the EC freshwater policy framework ENV/D.1/FRA/2012/0014. MS fiche, Portugal.

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.

Gaeta, G.L., Ghinoi, S., Silvestri, F., 2017. Municipal performance in waste recycling: an empirical analysis based on data from the lombardy region (italy). Letters in Spatial and Resource Sciences 10, 337–352.

García, I., Mendoza, R., Pomar, M.C., 2008. Deficit and excess of soil water impact on plant growth of lotus tenuis by affecting nutrient uptake and arbuscular mycorrhizal symbiosis. Plant and soil 304, 117–131.

Geelen, L.M., Huijbregts, M.A., den Hollander, H., Ragas, A.M., van Jaarsveld, H.A., de Zwart, D., 2009. Confronting environmental pressure, environmental quality and human health impact indicators of priority air emissions. Atmospheric Environment 43, 1613–1621.

Giannouli, M., Samaras, Z., Keller, M., DeHaan, P., Kallivoda, M., Sorenson, S., Georgakaki, A., 2006. Development of a database system for the calculation of indicators of environmental pressure caused by transport. Science of the Total Environment 357, 247–270.

Gonçalves, I., Alves, D., Robalo, G., 2014. Social tariffs for water and waste services in mainland portugal: An impact analysis. Water Science and Technology: Water Supply 14, 513–521.

González-Benito, J., González-Benito, Ó., 2010. A study of determinant factors of stakeholder environmental pressure perceived by industrial companies. Business Strategy and the Environment 19, 164–181.

Goodland, R., 1995. The concept of environmental sustainability. Annual review of ecology and systematics 26, 1–24.

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.

Horta, I.M., Camanho, A.S., 2015. A nonparametric methodology for evaluating convergence in a multi-input multi-output setting. European journal of operational research 246, 554–561.

Juwana, I., Muttil, N., Perera, B., 2012. Indicator-based water sustainability assessment—a review. Science of the Total Environment 438, 357–371.

Kaza, S., Yao, L., Bhada-Tata, P., Van Woerden, F., 2018. What a waste 2.0: a global snapshot of solid waste management to 2050. World Bank Publications.

Lavigne, C., Jaeger, S.D., Rogge, N., 2019. Identifying the most relevant peers for benchmarking waste management performance: A conditional directional distance Benefit-of-the-Doubt approach. Waste Management 89, 418–429.

Li, Q., Racine, J., 2003. Non-parametric estimation of distributions with categorical and continuous data. Journal of Multivariate Analysis 86, 266 – 292.

Liang, S., Liu, Z., Crawford-Brown, D., Wang, Y., Xu, M., 2014. Decoupling analysis and socioeconomic drivers of environmental pressure in china. Environmental science & technology 48, 1103–1113.

Lombardi, G., Stefani, G., Paci, A., Becagli, C., Miliacca, M., Gastaldi, M., Giannetti, B., Almeida, C., 2019. The sustainability of the Italian water sector: An empirical analysis by DEA. Journal of Cleaner Production 227, 1035–1043.

Margallo, M., Taddei, M.B.M., Hernández-Pellón, A., Aldaco, R., Irabien, A., 2015. Environmental sustainability assessment of the management of municipal solid waste incineration residues: a review of the current situation. Clean Technologies and Environmental Policy 17, 1333–1353.

Marques, R.C., da Cruz, N.F., Pires, J., 2015. Measuring the sustainability of urban water services. Environmental Science & Policy 54, 142–151.

Marques, R.C., Simões, P., 2020. Revisiting the comparison of public and private water service provision: An empirical study in portugal. Water 12, 1477.

Marques, R.C., Simões, P., Pinto, F.S., 2018. Tariff regulation in the waste sector: An unavoidable future. Waste management 78, 292–300.

Martins, R., Antunes, M., Fortunato, A., 2020. Regulatory changes to portugal's social tariffs: Carrying water in a sieve? Utilities Policy 64, 101038.

Melyn, W., Moesen, W., 1991. Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. Public economics research papers , 1–24.

Moldan, B., Janoušková, S., Hák, T., 2012. How to understand and measure environmental sustainability: Indicators and targets. Ecological Indicators 17, 4–13.

Molinos-Senante, M., Gómez, T., Caballero, R., Sala-Garrido, R., 2017. Assessing the quality of service to customers provided by water utilities: A synthetic index approach. Ecological Indicators 78, 214 – 220.

Molinos-Senante, M., Marques, R.C., Pérez, F., Gómez, T., Sala-Garrido, R., Caballero, R., 2016. Assessing the sustainability of water companies: A synthetic indicator approach. Ecological indicators 61, 577–587.

Munksgaard, J., Wier, M., Lenzen, M., Dey, C., 2005. Using input-output analysis to measure the environmental pressure of consumption at different spatial levels. Journal of Industrial Ecology 9, 169–185.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2005. Handbook on Constructing Composite Indicators. OECD Statistics Working Papers 03.

Nikolaou, K., 2001. Planning of an environmental pressure indicators system in a metropolitan area. Journal of Environmental Protection and Ecology 2, 595–602.

OECD, 2020. Environment at a Glance Indicators. URL: `https://www.oecd-ilibrary.org/content/publication/ac4b8b89-en`, doi:`https://doi.org/https://doi.org/10.1787/ac4b8b89-en`.

OECD Environment Directorate, 2008. OECD Key Environmental Indicators. Paris, France: Organisation for Economic Development and Cooperation .

Pérez, F., Delgado-Antequera, L., Gómez, T., 2019. A Two-Phase Method to Assess the Sustainability of Water Companies. Energies 12, 2638.

Pérez, F., Molinos-Senante, M., Gómez, T., Caballero, R., Sala-Garrido, R., 2018. Dynamic goal programming synthetic indicator: an application for water companies sustainability assessment. Urban Water Journal 15, 592–600.

Pinto, F.S., Costa, A.S., Figueira, J., Marques, R.C., 2017. The quality of service: An overall performance assessment for water utilities. Omega 69, 115–125.

Purvis, B., Mao, Y., Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. Sustainability science 14, 681–695.

Rogge, N., 2012. Undesirable specialization in the construction of composite policy indicators: The environmental performance index. Ecological indicators 23, 143–154.

Rogge, N., De Jaeger, S., Lavigne, C., 2017. Waste Performance of NUTS 2-regions in the EU: A Conditional Directional Distance Benefit-of-the-Doubt Model. Ecological Economics 139, 19 – 32.

Romano, G., Guerrini, A., 2011. Measuring and comparing the efficiency of water utility companies: A data envelopment analysis approach. Utilities Policy 19, 202 – 209.

Sala, S., Ciuffo, B., Nijkamp, P., 2015. A systemic framework for sustainability assessment. Ecological Economics 119, 314–325.

Sarra, A., Mazzocchitti, M., Rapposelli, A., 2017. Evaluating joint environmental and cost performance in municipal waste management systems through data envelopment analysis: Scale effects and policy implications. Ecological Indicators 73, 756–771.

Serra, J., Cordovil, C.M., Cruz, S., Cameira, M.R., Hutchings, N.J., 2019. Challenges and solutions in identifying agricultural pollution hotspots using gross nitrogen balances. Agriculture, Ecosystems & Environment 283, 106568.

Serra, J., do Rosário Cameira, M., Cordovil, C.M., Hutchings, N.J., 2021. Development of a groundwater contamination index based on the agricultural hazard and aquifer vulnerability: Application to portugal. Science of The Total Environment 772, 145032.

Silva, C., Rosa, M.J., 2020. Performance assessment of 23 wastewater treatment plants-a case study. Urban Water Journal 17, 78–85.

da Silva, J., Fernandes, V., Limont, M., Dziedzic, M., Andreoli, C., Rauen, W., 2020. Water sustainability assessment from the perspective of sustainable development capitals: Conceptual model and index based on literature review. Journal of environmental management 254, 109750.

Simões, P., Cavalho, P., Marques, R.C., 2013. The market structure of urban solid waste services: how different models lead to different results. Local Government Studies 39, 396–413.

Simões, P., Marques, R.C., 2012. On the economic performance of the waste sector. A literature review. Journal of environmental management 106, 40–47.

Smeets, E., Weterings, R., 1999. Environmental indicators: Typology and overviewe. Technical report 25, 411 – 432.

Sutton, P., 2004. A perspective on environmental sustainability. Paper on the Victorian Commissioner for Environmental Sustainability , 1–32.

UN General Assembly, 2015. Transforming our world: the 2030 Agenda for Sustainable Development. New York: United Nations .

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.

Walter, M., Cullmann, A., von Hirschhausen, C., Wand, R., Zschille, M., 2009. Quo vadis efficiency analysis of water distribution? a comparative literature review. Utilities Policy 17, 225–232.

Worthington, A.C., 2014. A review of frontier approaches to efficiency and productivity measurement in urban water utilities. Urban Water Journal 11, 55–73.

Zanella, A., Camanho, A.S., Dias, T.G., 2015. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. European Journal of Operational Research 245, 517 – 530.

Zeller, V., Lavigne, C., D'Ans, P., Towa, E., Achten, W., 2020. Assessing the environmental performance for more local and more circular biowaste management options at city-region level. Science of The Total Environment 745, 140690.

Zhou, P., Ang, B., Poh, K., 2007. A mathematical programming approach to constructing composite indicators. Ecological economics 62, 291–297.

Zhu, G., Tang, Z., Shangguan, Z., Peng, C., Deng, L., 2019. Factors affecting the spatial and temporal variations in soil erodibility of china. Journal of Geophysical Research: Earth Surface 124, 737–749.

Appendix A. ‘Optimistic’ VS ‘Pessimistic’ version of the BoD model

In the following we present the primal and multiplier formulation of the model introduced by Zanella et al. 2015, along with its pessimistic counterpart.

Zanella et al. (2015, mod.7-8 p.523)
OPTIMISTIC VERSION
Primal formulation

$$\max \beta$$

$$s.t. \sum_{j=1}^{n} b_{kj}\lambda_j \leq b_{kj_0} - \beta g_b \quad \text{for } k = 1,\ldots,l$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rj_0} + \beta g_y \quad \text{for } r = 1,\ldots,s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0 \quad \text{for } j = 1,\ldots,j_0,\ldots,n$$

Multiplier formulation

$$\beta_{j_0} = \quad \min -\sum_{r=1}^{s} y_{rj_0}u_{rj_0} + \sum_{k=1}^{l} b_{kj_0}p_{kj_0} + v_{j_0}$$

$$s.t. \quad \sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1$$

$$-\sum_{r=1}^{s} y_{rj}u_{rj_0} + \sum_{k=1}^{l} b_{kj}p_{kj_0} + v_{j_0} \geq 0 \quad \text{for } j = 1,\ldots,j_0,\ldots,n$$

$$u_{rj_0} \geq 0 \quad \text{for } r = 1,\ldots,s$$

$$p_{kj_0} \geq 0 \quad \text{for } k = 1,\ldots,l$$

$$v_{j_0} \in \Re$$

where $CI_{j_0} = 1/(1+\beta_{j_0}) \quad \in (0,1]$

Based on Zanella et al. (2015)
PESSIMISTIC VERSION
Primal formulation

$$\min \beta^P$$

$$s.t. \sum_{j=1}^{n} b_{kj}\lambda_j \geq b_{kj_0} - \beta g_b \quad \text{for } k = 1,\ldots,l$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j \leq y_{rj_0} + \beta g_y \quad \text{for } r = 1,\ldots,s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0 \quad \text{for } j = 1,\ldots,j_0,\ldots,n$$

Multiplier formulation

$$\beta^P_{j_0} = \quad \max -\sum_{r=1}^{s} y_{rj_0}u_{rj_0} + \sum_{k=1}^{l} b_{kj_0}p_{kj_0} + v_{j_0}$$

$$s.t. \quad \sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1$$

$$-\sum_{r=1}^{s} y_{rj}u_{rj_0} + \sum_{k=1}^{l} b_{kj}p_{kj_0} + v_{j_0} \leq 0 \quad \text{for } j = 1,\ldots,j_0,\ldots,n$$

$$u_{rj_0} \geq 0 \quad \text{for } r = 1,\ldots,s$$

$$p_{kj_0} \geq 0 \quad \text{for } k = 1,\ldots,l$$

$$v_{j_0} \in \Re$$

where $CI^P_{j_0} = 1/(1+\beta^P_{j_0}) \quad \in [1,+\infty)$

y_{rj_0} and b_{kj_0} respectively refer to the observed r desirable and k undesirable indicator of the evaluated DMU j_0. u_{rj_0} and p_{kj_0} are the BoD weights corresponding to the r desirable and k undesirable indicator for the evaluated DMU j_0. In the optimistic model they represent the most favorable weights for the unit under evaluation, in the pessimistic model the least favorable. y_{rj} and b_{kj} respectively refer to the r desirable and k undesirable indicator of every DMU j in the dataset; n is the number of DMU under analysis; s and l respectively denote the number of desirable and undesirable indicators considered in the application.

Appendix B. Descriptive statistics of the pessimistic BoD

In the following we present the descriptive statistics for the pessimistic version of the proposed environmental pressure index, both overall and grouped by operating context variables.

Table B.1: Descriptive statistics of the *pessimistic* environmental pressure composite indicator scores (both overall and grouped by operating context variables). The scores are obtained implementing the deterministic and unconditional analysis.

	N	Mean	SD	Min.	Q1	Median	Q3	Max.
Deterministic unconditional	149	2.213	0.9602	1.000	1.481	2.122	2.679	5.776
Geographical location								
North	39	2.104	0.9171	1.000	1.317	2.131	2.642	4.429
Centre	60	2.169	0.9861	1.000	1.478	2.059	2.547	5.776
South	50	2.352	0.9644	1.000	1.731	2.204	2.695	5.404
Intervention area								
Rural	112	2.145	0.8928	1.000	1.490	2.058	2.577	5.776
Semi-urban	29	2.465	1.1425	1.000	1.481	2.194	3.017	5.404
Urban	8	2.254	1.1368	1.000	1.558	1.940	2.610	4.429
Volume of activity								
Small	53	1.944	0.7049	1.000	1.382	2.053	2.378	4.057
Medium	48	2.079	0.8729	1.000	1.442	1.874	2.531	5.776
Large	48	2.646	1.1403	1.000	1.756	2.377	3.234	5.492