Operational Research (2026) 26:71
https://doi.org/10.1007/s12351-026-01061-0

ORIGINAL PAPER

Comparative analysis of additive and multiplicative BoD models in healthcare performance evaluation

Guilherme Vara[1] · Marta Castilho Gomes[2] · Paulo Caldas[1,3,4] · Miguel Varela[3,5] · Diogo Cunha Ferreira[1]

Received: 20 May 2024 / Revised: 28 February 2025 / Accepted: 30 April 2026

Abstract
Composite indicators play a pivotal role across various fields, serving as powerful tools to condense information into a single, comprehensible metric. They function as a link between complex data and practical conclusions, which makes them valuable assets in diverse fields, assisting decision-makers. Since objectivity is vital for policy- and decision making, many researchers base their studies on the Benefit-of-the-Doubt (BoD) methodology, which originated from the widely recognised Data Envelopment Analysis. This study provides a detailed comparison between the linear and multiplicative BoD approaches, incorporating both optimistic and pessimistic viewpoints to assess the performance of the Portuguese public hospitals. The linear approach is an additive linear programming model that constructs a piecewise linear and convex efficient frontier with the benchmarks. In contrast, the multiplicative approach constructs logarithmic curves instead of linear segments to define the efficient frontier allowing for greater flexibility by accommodating non-convex and nonlinear shapes that better reflect the data distribution. As a result, the multiplicative approach achieves a tighter fit to the data, ultimately yielding higher overall scores. The results show that multiplicative scores in the pessimistic approach are lower than linear scores. Conversely, in the optimistic multiplicative approach, while expected to yield superior scores, certain entities lag due to non-compensatory elements. The implementation of the multiplicative BoD is remarkably simple, requiring only minimal changes when compared to the linear BoD. This raises questions about its low adoption and utilization compared to linear methods, despite its apparent advantages.

Keywords Healthcare performance · COVID-19 pandemic · Composite indicators · Multiplicative benefit-of-the-doubt · Data envelopment analysis

Extended author information available on the last page of the article

Published online: 25 May 2026

1 Introduction

Composite indicators (CIs) have become a popular tool in many fields, such as healthcare, education, and economics (Greco et al. 2019; Nardo et al. 2005). They serve as a valuable statistical tool that combines multiple indicators into a single measure with the goal of summarizing complex information in a simple and intuitive way, facilitating policymaking and decision-making (Freudenberg 2003).

Two major techniques commonly used in CI construction are Data Envelopment Analysis (DEA) and Multiple Criteria Decision Analysis (MCDA) (Hatefi & Torabi 2010). DEA, on the one hand, is a non-parametric method used to measure the relative efficiency of decision-making units (DMUs) based on their input–output relationships (Cooper et al. 2000). Some MCDA models, on the other hand, determine the weights of each criterion and combine them into a comprehensive CI. Notably, MCDA methods take into account the preferences and priorities of decision-makers to assign weights and aggregate the criteria (Dias et al. 2023). It's important to highlight that these weights are uniformly applied across all entities under evaluation (Juwana et al. 2012). However, this approach is generally not suitable for most public administration scenarios, as each entity has its own policies. To ensure a more accurate representation of performance and support well-informed decision-making, the composite indicator must be tailored to the specific context of each entity (Greco et al. 2019).

Furthermore, despite the advantages of MCDA in CI construction, one of the key challenges is the determination of the set of weights for the KPIs (Zhou et al. 2010), which stands out as one of the strengths of DEA. Assigning appropriate weights to each KPI is crucial as it influences the final composite score and the relative importance given to different factors. However, the process of assigning weights is subjective and may vary depending on the preferences and priorities of decision-makers. Therefore, careful consideration and expert judgment are necessary to ensure the weights accurately reflect the significance of each KPI in the overall composite indicator (El Gibari et al. 2019). Robust methodologies and transparent decision-making processes are essential to enhance the credibility and reliability of the constructed composite indicators.

Nonetheless, the construction and interpretation of composite indicators is not a straightforward task, and there are many methodological challenges that need to be addressed (Dialga & Giang 2017). One of the main challenges is the choice of aggregation method (Barclay et al. 2019), which determines how the individual key performance indicators (KPIs) are combined into a single measure. The most common aggregation method is the linear approach, which multiplies each indicator by a coefficient and adds them up. However, the linear approach has several limitations, such as the inability to capture nonlinear relationships between the indicators, the oversimplification of complex systems, and its compensatory nature (Ferreira et al. 2023). Many real-world problems involve interdependencies, non-linearities, and feedback loops, which cannot be adequately captured by a simple linear aggregation method, as also noted by Emrouznejad and Cabanda (2010).

In a previous study, the linear Benefit-of-the-Doubt (BoD) approach was employed, an extension of DEA used to construct linear composite indicators of hospital performance before and during the COVID-19 pandemic. While the linear approach provided useful insights, it does have certain limitations, as outlined in the literature review (Sect. 2.2), including the failure to account for increasing marginal products and the potential underestimation of efficiency for non-dominated DMUs, among others. Although some problems were addressed in the previous study, here, a multiplicative approach is proposed, which offers improved theoretical properties, and more accurate results in constructing composite indicators (Ferreira et al. 2021).

The motivation for this paper is threefold. Firstly, it aims to provide a comparison between the linear and multiplicative BoD approaches, using both a theoretical example and a real-world application as a follow-up. Secondly, it offers a methodological contribution to the field of composite indicators by filling a gap in the existing literature where a comprehensive comparison between both approaches is non-existent, as typically observed in models like DEA and its equivalents (Vieira et al. 2023; Sueyoshi and Sekitani 2009; Hjalmarsson et al. 1996;). Thirdly, the methodology is then applied to assess hospital performance, particularly considering the impact of the COVID-19 pandemic, which will allow to compare the results between the multiplicative approach with the previously employed linear approach. This research topic holds significant importance, given the profound influence the pandemic has had on healthcare systems worldwide. Consequently, there exists a pressing need for tools that can assist policymakers and healthcare managers in identifying areas for improvement to effectively allocate resources.

As a result, the objectives of this paper are as follows: (i) to conduct a detailed comparison between the multiplicative approach with the linear approach, highlighting their differences (ii) to integrate different weight restrictions to the multiplicative approach, (iii) to construct composite indicators of hospital performance using the multiplicative approach, and (iv) to analyse the impact of the COVID-19 pandemic on hospital efficiency and performance using the composite indicators.

As a result, the research questions that guide this paper are: (i) How does the multiplicative approach to composite indicators differ from the linear approach and what is the impact on efficiency assessment? (ii) Is the multiplicative approach a more accurate and flexible tool for measuring hospital efficiency? (iii) How has the COVID-19 pandemic affected hospital efficiency and performance, as measured by the composite indicators? (iv) What are the policy implications of the findings?

The structure of this paper is as follows: Sect. 2 provides a review of the literature on composite indicators and the multiplicative approach. Section 3 details the methodology used. Section 4 describes the data and the case study used to construct the composite indicators of hospital efficiency and performance. Section 5 presents the results of the analysis and discusses the impact of the COVID-19 pandemic on hospital efficiency and performance. Finally, Sect. 6 concludes the paper by summarizing the main findings, discussing their implications, and suggesting directions for future research.

2 Literature review

2.1 Composite indicators and their significance in policy analysis and decision-making

Composite indicators have a significant role in policy analysis and decision-making, providing a comprehensive framework to measure complex multidimensional phenomena which provide valuable insights into various fields, including economics, sustainability, social development, and governance (Olesen et al. 2015; Booysen 2002).

Composite indicators offer numerous advantages in policy analysis by providing a comprehensive understanding that considers multiple dimensions and diverse information (Greco et al. 2016). They facilitate comparisons between regions, countries, or organizations, aiding policymakers in identifying areas for improvement and prioritizing interventions (Burgass et al. 2017). Additionally, composite indicators enable the monitoring of progress and evaluation of policy impact (Saltelli 2007).

The evolution of composite indicators can be traced back to leading contributions such as the Human Development Index (HDI) and the Gross Domestic Product (GDP) (OECD 2005). Over time, the field has witnessed advancements in methodology, data availability, and conceptual frameworks, where researchers have shifted from unidimensional indicators to more comprehensive approaches that consider multiple dimensions and aggregate various KPIs. This evolution has been driven by the recognition that policy analysis requires a full understanding of the multidimensionality of the issues at hand (Vértesy 2016).

Different methodologies have been developed to construct composite indicators, including weighting and aggregation techniques, index approaches, and data-driven methodologies (Sharpe & Andrews 2012). These methods include:

1. Weighting and aggregation methods involve assigning weights to individual indicators based on their importance and then combining them to obtain an overall score or ranking (Santeramo 2015; El Gibari et al. 2019; Zhou & Ang 2009; Gómez-Limón and Riesgo 2009).
2. Index approaches use predefined formulae to calculate the composite score based on predetermined weights (Sagar and Najam 1998; Ave and Babolsar 2010).
3. Data-driven methodologies, such as principal component analysis (Hudrliková, 2013; Li et al. 2012) or data envelopment analysis (Cherchye et al. 2008), use statistical techniques to derive composite indicators.

However, challenges arise due to data availability and quality, as well as the subjectivity and complexities associated with weighting methodologies and interpretation (Decancq and Lugo 2013). Engaging stakeholders throughout the process enhances credibility and ensures alignment with policy goals (Lindén et al. 2021). Effective communication strategies, such as clear visualizations like dashboards or scorecards, also facilitate informed decision-making by policymakers and the public (Dias et al. 2024).

The future of composite indicators in policy analysis holds immense potential for further advancements. Key areas for future research include refining methodologies by improving weighting and aggregation techniques, reducing subjectivity in indicator selection, and incorporating advanced statistical approaches (Cherchye et al. 2008). The integration of emerging technologies, such as machine learning and big data analytics, can enhance real-time data inclusion and adaptability (Jiménez-Fernández et al. 2022). Ensuring transparency through standardized guidelines and stakeholder engagement can improve legitimacy and comparability (Greco et al. 2019). Lastly, enhancing communication strategies with user-friendly visualizations and clear reporting will facilitate better understanding and application of composite indicators in decision-making (Albo et al. 2015), bridging the gap between complex technical concepts and practical decision-making processes (Lyytimäki et al. 2013).

2.2 Multiplicative BoD for constructing composite indicators

In the realm of constructing composite indicators, the OECD recommends four statistical models, including the Benefit-of-the-Doubt (BoD), data envelopment analysis (DEA), principal component/factor analysis, and the unobserved components model (Karagiannis 2017).

BoD is a method rooted in DEA, initially proposed by Melyn and Moesen (1991) to evaluate macroeconomic performance and has since been adapted for constructing composite indicators and to address scenarios involving multiple reference units, resulting in a performance score that goes beyond efficiency alone (Cherchye et al. 2007). Notably, Cherchye's paper alludes to the possibility of employing a multiplicative aggregation approach, potentially accompanied by a linearization procedure.

In light of this, numerous researchers have further explored Cherchye's reference and expanded on the so-called multiplicative BoD, which involves a multiplicative aggregation of KPIs, to overcome the limitations of the traditional linear-based BoD approach. The linear BoD approach fails to consider increasing marginal productivity (Kao 2016), leading to an inadequate representation of the production process. It also tends to underestimate the efficiency of non-dominated DMUs (Tiedemann et al. 2011), which can distort performance assessments. Furthermore, the linear BoD approach neglects the important aspect of outputs competing for inputs (Banker & Maindiratta 1986), overlooking the dynamic relationships within the production system. It also exhibits poor performance when dealing with non-concave production technologies (Emrouznejad & Cabanda 2010), limiting its applicability in cases where the production technology exhibits complex characteristics. Additionally, the linear BoD approach may encounter the zero-multiplier problem (Tofallis 2014), where certain variables are not accounted for in the performance score computation, resulting in biased outcomes.

The BoD approach has found applications in various sectors, including healthcare (Matos et al. 2021), waste management (Ferreira et al. 2023), education (Sahoo et al. 2017), economics (Lafuente et al. 2022), and banking (Gulati et al. 2020), among others. While the multiplicative approach has not received extensive attention in the literature, it has demonstrated superior theoretical properties compared to linear methods, such as scale-invariance and a lower degree of compensation between variables (Verbunt & Rogge 2018). Moreover, the multiplicative aggregation maintains a higher level of information compared to other aggregating techniques (Zhou et al. 2010).

However, despite the advantages of the multiplicative approach, there are still several challenges and limitations to be addressed. Existing studies in this field often neglect the presence of undesirable KPIs (Ferreira et al. 2023; Podinovski & Kuosmanen 2011) and fail to consider minimum and maximum thresholds imposed by regulators or stakeholders (Marques et al. 2018). Furthermore, they do not account for the existence of KPIs that should fall within certain ranges defined by regulators (Simões et al. 2010). Imperfect knowledge of data is another issue that needs to be acknowledged, as performance estimates may become stochastic rather than deterministic (Ferreira et al. 2018). To overcome the limitations of the multiplicative approach, researchers have explored various integrations with other methods and developed alternative models and methodologies. Table 3 provides an overview of selected studies on the multiplicative BoD approach.

3 Methodology

3.1 The fundamentals of the benefit-of-the-doubt approach

The BoD approach is a methodology derived from DEA to construct composite indicators in the context of uncertainty about appropriate weighting schemes. DEA, originally developed by Charnes et al. (1978), is a linear programming tool that evaluates the efficiency of DMUs based on multiple inputs and outputs.

The BoD approach extends DEA to composite indicator construction, allowing for a more objective evaluation without the need for subjective criteria weights (Cherchye et al. 2007). The core idea behind the BoD approach is that information on appropriate weighting schemes can be inferred from the data itself. It assumes that good relative performance of an entity on a specific KPI indicates the entity's relatively higher policy priority in that dimension. Conversely, poor relative performance suggests lower policy importance. The BoD approach addresses this by allowing each entity to have entity-specific weights for KPIs, which maximize its composite indicator score. The linear model from Cherchye et al. (2007) is as follows:

$$CI_t = \max_{\alpha_{t,j}} \sum_{j=1}^{n} \alpha_{t,j} \cdot y_{t,j} \tag{1}$$

$$s.t. \sum_{j=1}^{n} \alpha_{t,j} \cdot y_{i,j} \leq 1, \; m \; constraints, \; one \; for \; each \; entity \; i \tag{2}$$

$$\alpha_{t,j} \geq 0, \; n \; constraints, \; one \; for \; each \; variable \; j \tag{3}$$

Table 1 Descriptive statistics for the considered variables along with the desired direction of change

Dimension	Variables	Direc[a]	Mean	SD	Min	Max
Access	(V1) Rate of first medical appointments within time	↑	70.70%	14.45%	3.40%	100.00%
	(V2) Rate of patients enrolled in surgical list within time	↑	72.61%	15.41%	25.41%	100.00%
Clinical safety	(V3) Pressure ulcer rate per 1.000	↓	0.83‰	2.00‰	0.00‰	25.20‰
	(V4) Central venous catheter-related bloodstream infections rate per 1.000	↓	0.21‰	0.60‰	0.00‰	7.40‰
	(V5) Postoperative pulmonary embolism/deep vein thrombosis per 100.000	↓	236.74	345.06	0.00	4095.56
	(V6) Postoperative septicaemia rate per 100.000	↓	644.89	949.94	0.00	12,500.00
	(V7) Rate of instrumental vaginal deliveries with third- and fourth-degree Lacerations	↓	2.40%	4.43%	0.00%	100.00%
	(V8) Rate of non-instrumented vaginal deliveries with third- and fourth-degree Lacerations	↓	0.48%	1.10%	0.00%	16.67%
Productivity	(V9) Standard patients per FTE doctor	↑	5.70	1.49	0.00	16.86
	(V10) Standard patients per FTE nurse	↑	3.55	1.18	0.00	9.50
	(V11) Inpatient bed occupancy rate	◯	85.05%	46.47%	0.00%	1867.40%
	(V12) Waiting time before surgery	↓	0.85	0.41	0.03	4.47
Assistance Performance	(V13) Rate of outpatient surgeries on potential outpatient procedures	↑	82.31%	11.64%	0.00%	100.00%
	(V14) Rate of readmissions within 30 days after discharge	↓	7.44%	1.92%	0.00%	14.81%
	(V15) Rate of inpatients staying more than 30 days	↓	3.93%	1.56%	0.00%	37.93%
	(V16) Rate of hip fracture surgery in the first 48 h	↑	46.99%	24.24%	0.00%	100.00%

[a]Direction: ↓ The lower, the better (undesirable); ↑ The higher, the better (desirable); ◯ Neutral, optimal value between 80 and 90%

where CI_t represents the composite indicator for entity t, while $y_{t,j}$ denotes the value of entity t on variable $j(j = 1, \ldots, n)$. The multiplier $\alpha_{t,j}$ is the value of variable j for entity t. Equation (1) represents the objective function, where the aim is to select the most favourable multipliers to maximize the composite indicator CI_t. To ensure that the maximum achievable score is one, Eq. (2) applies the same set of multipliers to all entities $i(i = 1, \ldots, m)$ and sets this score's limit to one. Furthermore, Eq. (3) ensures that the multipliers are nonnegative throughout the optimization process.

3.2 Multiplicative benefit-of-the-doubt

For aggregation purposes, the multiplicative approach has demonstrated superior suitability for handling variables expressed as ratios and percentages, as evidenced by studies conducted by Petridis et al. (2023) and Puyenbroeck & Rogge (2017). This multiplicative nature contributes to a less compensatory method, as indicated by Ferreira et al. (2023), resulting in more penalizing scores even when only a single variable exhibits poor performance. This characteristic makes the multiplicative approach particularly appropriate for real-world applications, such as in healthcare, where failure can have critical consequences, as human lives are at stake. Following Cherchye's indication, the multiplicative objective function is as follows:

$$CI_t = \max_{\alpha_{t,j}} \prod_{j=1}^{n} (y_{t,j})^{\alpha_{t,j}} \tag{4}$$

$$s.t. \prod_{j=1}^{n} (y_{t,j})^{\alpha_{t,j}} \leq e, \; m \; constraints, \; one \; for \; each \; entity \; i \tag{5}$$

$$\alpha_{t,j} \geq 0, n \; constraints, \; one \; for \; each \; variable \; j \tag{6}$$

with the same notation as the linear model, but with a Π (product) instead of a Σ (summation), using the Napier number as the linear inequality constraint rather than one, to ensure feasibility when the linear transformation is applied. As discussed by Cherchye (2007), the multiplicative model needs to be linearized to achieve a similar structure as before, becoming computationally more efficient and making it feasible for practical implementation and analysis.

The linearization process involves applying logarithms to both sides of the equation. This transformation not only enhances the overall efficiency of the programs utilizing this approach but also facilitates further analysis and interpretation:

$$CI_t^{log} = log(CI_t) = \max_{\alpha_{t,j}} \sum_{j=1}^{n} \alpha_{t,j} \cdot log(y_{t,j}) \tag{7}$$

$$s.t. \sum_{j=1}^{n} \alpha_{t,j} \cdot log(y_{t,j}) \leq 1, \; m \; constraints, \; one \; for \; each \; entity \; i \tag{8}$$

$$\alpha_{t,j} \geq 0, \; n \; constraints, \; one \; for \; each \; variable \; j \tag{9}$$

with the composite indicator having now the score transformed due to the logarithm, resulting in the efficient entities being valued at 0, as $\log(1) = 0$. On the other hand, the inefficient entities are represented by negative numbers, with the magnitude of negativity indicating the extent of their inefficiency. Importantly, there is no lower bound to these negative values as $\lim_{x \to 0^+} \log x = -\infty$.

This research also focuses on comparing and analysing the distinctions between the linear and multiplicative pessimistic approaches, while also comparing the optimistic and pessimistic multiplicative approaches. In other words, it aims to find the worst-case scenario by minimizing the objective function subject to the inverted inequality constraints. The pessimistic model, after the linearization process through the logarithms is as follows:

$$CI_t^{log,pessi} = log(CI_t^{pessi}) = \min_{\alpha_{t,j}} \sum_{j=1}^{n} \alpha_{t,j} \cdot log(y_{t,j}) \tag{10}$$

$$s.t. \sum_{j=1}^{n} \alpha_{t,j} \cdot log(y_{t,j}) \geq 1, m constraints, one for each entity i \tag{11}$$

$$\alpha_{t,j} \geq 0,\ n\ constraints,\ one\ for\ each\ variable\ j \tag{12}$$

3.3 Additional restrictions

Another aim of this study was to incorporate various weight restrictions to the multiplicative approach and further investigate the differences between them, both in the linear and multiplicative models. To achieve this, the logarithmic transformation must be applied to convert the same restrictions used in the linear method (Eqs. 13 and 14), making them applicable to the multiplicative approach. As a result, new equations were derived, Eqs. (15) and (16).

$$\frac{\alpha_{t,j}}{\sum_{j=1}^{n} \alpha_{t,j}} \geq w_j,\ n\ constraints,\ one\ for\ each\ variable\ j \tag{13}$$

$$\frac{\alpha_{t,j} \cdot y_{t,j}}{\sum_{j=1}^{n} \alpha_{t,j} \cdot y_{t,j}} \geq w_j,\ n\ constraints,\ one\ for\ each\ variable\ j \tag{14}$$

$$\frac{\alpha_{t,j}}{\sum_{j=1}^{n} \alpha_{t,j}} \geq w_j,\ n\ constraints,\ one\ for\ each\ variable\ j \tag{15}$$

$$\frac{\alpha_{t,j} \cdot log(y_{t,j})}{\sum_{j=1}^{n} \alpha_{t,j} \cdot log(y_{t,j})} \geq w_j,\ n\ constraints,\ one\ for\ each\ variable\ j \tag{16}$$

with Eqs. (13) and (14) called the multipliers weight restrictions and Eqs. (15) and (16) called the pie-share restrictions.

These equations are weight restrictions, which ensure that every multiplier or every variable (depending on the case, see Vara et al. (2024)) has a weight of at least

w_j in relation to the sum of all the multipliers or variables for the variable j. This weight, w_j, should be assigned by a DMU or an expert of the field for each variable, to produce realistic and meaningful results.

3.4 Scaling between models

In the multiplicative approach, a rescaling of the scores is required due to the application of logarithmic transformations during the linearization process. As the logarithm is applied to both sides of the equation, the composite indicator becomes expressed in a logarithmic scale. Consequently, rescaling of the scores becomes necessary. The rescaling formula is as follows:

$$CI_t^{final} = e^{(CI_t^{log}-1)} \tag{17}$$

where CI_t^{final} is the multiplicative score after transformation, allowing a comparison with the linear approach. For the case of the pessimistic approach, this rescale must be done before the linear rescale, otherwise the values will be corrupted.

An additional rescaling for the pessimistic scores is required because the pessimistic scores measure the distance of each entity from the inefficient frontier, where a value of 1 represents inefficiency, and the maximum positive value denotes maximum efficiency. Equation (18) standardizes the pessimistic BoD values to the common scale, 0 to 1, facilitating a straightforward and meaningful comparison between both modelling approaches.

$$CI_t^{final,pessi} = \frac{CI_t^{log,pessi} - min(CI_t^{log,pessi})}{max\left(CI_t^{log,pessi}\right) - min(CI_t^{log,pessi})} \tag{18}$$

where $CI_t^{final,pessi}$ is the pessimistic score after transformation, allowing a comparison between the optimistic and pessimistic approaches.

4 Case study: portuguese NHS

Portugal's National Health Service (NHS) has undergone significant reforms and changes to enhance efficiency, effectiveness, and access to care. The implementation of a regionalization model in the late 1990s decentralized the healthcare system, increasing the autonomy of public hospitals and allowing them to tailor services to their populations (Barros et al. 2011). Despite challenges like a shortage of healthcare professionals, inadequate funding, and an aging population, the NHS has demonstrated resilience and adaptability, improving the quality and accessibility of healthcare services (Pereira et al. 2021; Mendes et al. 2010). During the 2009 economic slowdown, the focus shifted to reducing healthcare costs and improving efficiency and sustainability. However, since 2016, with the country's economic recovery, there have been proposals to reform the health sector's governance structure to improve access, quality, and efficiency (Nunes & Ferreira 2019).

4.1 Portuguese public hospitals

Even though primary health care is considered a political priority (Barros et al. 2011), Health Centres lack autonomy and are overshadowed by Hospitals (Baganha et al. 2002). This disparity leads to inadequate outpatient services, causing an overload of hospital emergency services and hindering access to primary care (Simões et al. 2017). The COVID-19 pandemic intensified this burden on public hospitals, resulting in increased demand for care. Hospitals had to expand capacity, deploy additional healthcare professionals, and adapt to treat COVID-19 patients (Matos et al. 2021). Non-urgent medical procedures were suspended or delayed, impacting hospital finances, and disrupting patient care, while Healthcare professionals faced heightened stress and risk, working on the frontlines during the pandemic (Conde & Arribas 2021).

Though the availability of vaccines has reduced fear, the pandemic's financial impact remains (Nunes & Ferreira 2022), with public funds having incurred debt, affecting the overall financial performance of the health sector (Dang et al 2023). Ensuring proper economic and financial functioning of hospitals is crucial for maintaining access, efficiency, productivity, financial stability, and quality in the healthcare system (Afonso et al. 2024; Kaye et al 2021). Addressing these challenges will be essential to building a resilient and effective healthcare system for the future (Afonso et al. 2023).

4.2 Entities and variables characterization

This study analysed data from the Central Administration of the Health System (ACSS) official website, focusing on public hospitals and corporate public entities (EPE) in the NHS. The analysis included 40 entities, as listed in Table 2 (see Appendix), with ACSS dividing public hospitals into five groups based on similarities and operating conditions. The entities were grouped into their respective ACSS groups and separate models were run for each group to avoid bias and ensure evaluation based on their relevant peer group.

Output variables were based on the Portuguese Ministry of Health's guidelines, particularly those established in the annual contracts between hospital management and the ministerial oversight. As a result, these variables are considered highly pertinent to the primary stakeholders and have been extensively used in the literature regarding the Portuguese case (Vara et al. 2024; Matos et al. 2021; Pederneiras et al. 2023; Castro et al. 2014; Ferreira & Marques 2021). Table 4 (see Appendix) provides a brief overview of the output variables and similar studies that employed them. In addition, Table 1 displays the 16 variables employed in this case study, as well as their desired direction.

4.3 Methodological considerations in relation to sample characteristics and data handling

The BoD approach, unless suitably modified (Färe et al. 2019; Zanella et al. 2015), does not consider undesirable variables. To overcome this limitation, a data standardization procedure was employed, transforming the data into an ascending scale ranging from 1.01 to the Napier number (e) through Eq. (19). This scale was chosen as it preserves the original scale after the logarithmic transformation, where log(1) equals 0, and log(e) equals 1. Setting the minimum threshold at 1.01 was essential to avoid the final scores turning to zero whenever there was a variable with the minimum value.

$$y_{i,j} = \frac{(e-1)\cdot(y_{i,j}^{unstd} - min(y_j^{unstd}))}{max\left(y_j^{unstd}\right) - min(y_j^{unstd})} + 1 \tag{19}$$

where $y_{i,j}$ is the final standardized value from 1 to e of variable j of entity i; $y_{i,j}^{unstd}$ is the raw data, unstandardized; and y_j^{unstd} is the vector of values for the variable j from all the entities.

Benchmarking models are highly sensitive to data quality and completeness, as imperfect data can significantly modify the rankings and the efficiency/performance results. Common issues include missing values and incorrect entries, with data gaps being among the most prevalent challenges. For a more in-depth discussion on data uncertainty and its implications, readers may refer to Ferreira et al. (2023), Ebrahimi et al. (2021), and Wang et al. (2005).

In this study, missing data can arise from multiple factors, including errors in data collection, incomplete records, and technical disruptions during transmission. Additionally, certain datasets may not yet be publicly accessible, as they undergo validation by the central authority before being officially released. Rather than discarding affected entities or using simple imputation methods, multiple imputation was applied, generating 5000 values per missing entry based on a normal distribution within a defined range. This approach preserves potential benchmarks and prevents data manipulation (Vara et al. 2024; Ferreira et al. 2023). Instead of assigning fixed performance values, the methodology represents each entity's performance as a statistical distribution over time, enhancing the robustness and reliability of the results.

Finally, because there was no opportunity or availability for an expert's opinion on the matter, the value for w_j in each model was set to 0.03125. This value was obtained using Eq. (20) and ensures that every variable contributes, even if it is only the minimum, w_j(3,125%).

$$w_j = \frac{1}{n_v \cdot 2} \tag{20}$$

with n_v being the total number of variables in the study.

5 Findings and discussion

5.1 Comparison between the linear and multiplicative approaches

The main objective of this study is to conduct a comprehensive comparison between the linear and multiplicative BoD approaches. The goal is to elucidate how each approach actually quantifies the efficiency values, considering that the frontier exhibits a distinct shape for each methodology. Figure 1 provides an illustrative example of how the efficient frontier varies in these scenarios. It was created by generating pseudo-random initial values within a predetermined range and then applying a sigmoid function and plotting the results.

Secondly, the study aims to investigate the extent to which the multiplicative approach retains its non-compensatory nature compared to the linear approach, which is one of the main reasons for its use. Due to its multiplicative aggregation, entities may have one poor-performing variable absorb the entire contribution of the remaining variables, even if they are considered benchmarks in those other variables. This characteristic may result in an over-penalization effect, warranting careful examination to understand its implications in composite indicators and efficiency assessment.

Figures 3, 4, 5, and 6 (in Appendix) provide the evolution of the CI score average for the linear and multiplicative approach for the timeframe studied, for groups B, C, D, and E, respectively. Each figure includes four graphs, with one corresponding to each model (optimistic/pessimistic) and restriction (multipliers/pie-share).

The absence of a common scale in the figures is justified by substantial variations exhibited by different models. Specifically, the results obtained from the optimistic approach are more closely clustered and concise, resulting in a shorter range of values for the scale. On the contrary, the pessimistic approach displays significant fluctuations, necessitating a larger interval to accommodate the entire range of data points.

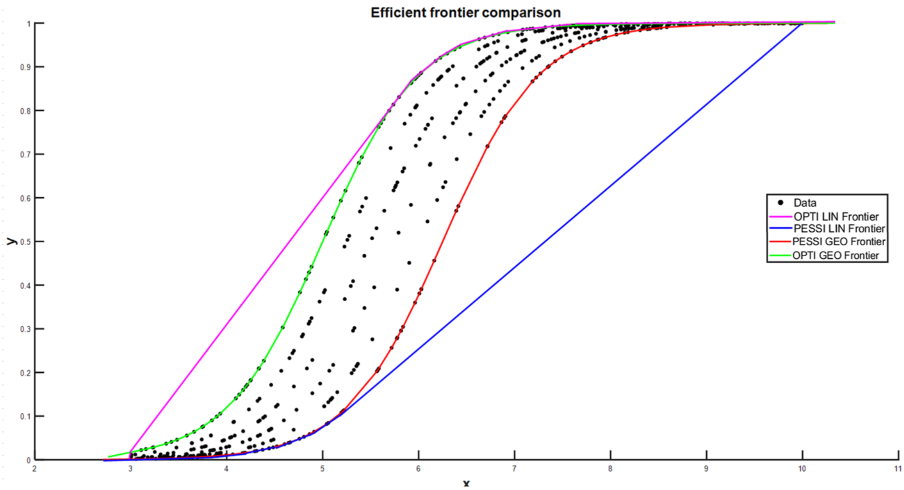

Fig. 1 Different shapes of efficient frontiers associated with a sigmoid function: y = $\frac{1}{1+e^{(-\lambda\cdot(x-x_o))}}$, $\lambda = 2, x_0 = [5{,}6.25]$

Moreover, instead of simply displaying the average in each scenario, Fig. 2 distinctly separates the top-performing entities from the weakest performers. This distinction highlights how the multiplicative approach differs from the linear approach depending on the entities' performance levels.

Firstly, the results demonstrate that, in the pessimistic approach, multiplicative scores consistently perform worse than linear scores across all groups. This observation aligns with the illustrative example presented in Fig. 1, where the linear pessimistic frontier tends to inflate the efficiency scores due to its more pronounced right-skew. Since the pessimistic approach measures the distance from the inefficient frontier, the right-skewed linear frontier places entities further left, leading to higher scores. As a result, the linear pessimistic approach produces inflated scores due to the greater distance between the inefficient frontier and most data points.

Secondly, the multiplicative approach exhibits sharper drops in performance than the linear approach due to its inherent multiplicative nature. These declines are predominantly due to a few variables exhibiting lower performance than usual, pulling down the overall efficiency score. In contrast, the linear approach compensates for low-performing variables with higher values in others, making performance drops less severe.

Furthermore, it is important to note that in the multiplicative approach, the efficient frontier is characterized by logarithmic curves instead of straight-line segments (as demonstrated by Ferreira et al 2021). These curves are not exclusively convex, and they dynamically adjust to the shape of the datapoints, reducing the distance between the frontier and the remaining entities, which leads to higher scores. Unlike the linear frontier depicted in Fig. 1, the multiplicative frontier fits the data more closely, minimizing unused space. This precision in fitting allows the multiplicative efficient frontier to accommodate a larger number of entities, thereby also contribut-

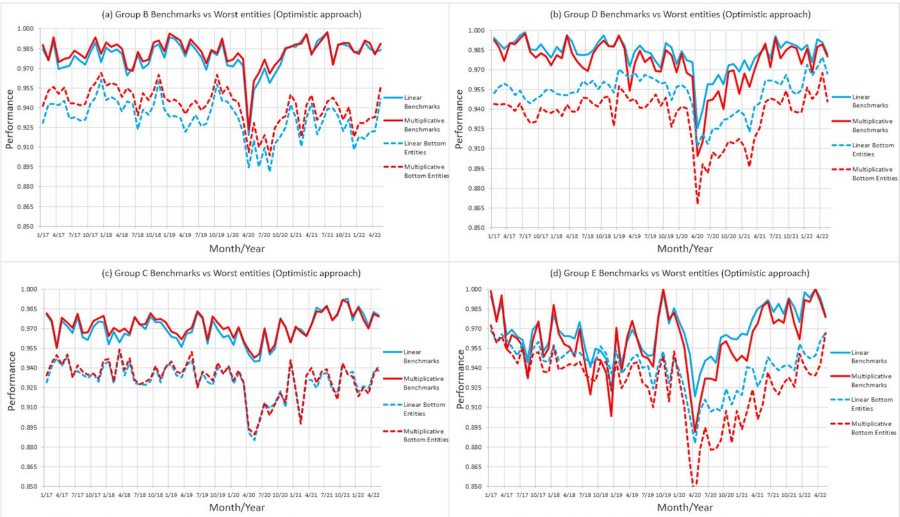

Fig. 2 Comparisons between the benchmark entities (BM) and the worst entities (WORST) within each group for both the linear (LIN) and multiplicative (GEO) approaches under the optimistic scenario

ing to higher overall scores. The multiplicative approach's ability to optimize the shape of the frontier in relation to the data points enhances the accuracy and inclusivity of entities close to the best practices.

Considering the distinctive characteristics of the multiplicative approach outlined in the previous two paragraphs, it is reasonable to anticipate that the optimistic multiplicative approach would yield higher scores than the linear approach. Its flexible frontier fits the data more precisely, making more entities efficient and positioning the frontier closer to inefficient ones, resulting in higher overall scores.

However, such consistency does not universally apply. In groups B and C (Figs. 3 and 4, respectively), the multiplicative approach generally demonstrates higher performance scores across the entire timeframe. However, there are two instances where the multiplicative scores fall below the linear scores. These correspond to declines in performance, which are notably more pronounced and penalizing in the multiplicative approach, due to the non-compensatory nature. Conversely, for groups D and E (Figs. 5 and 6, respectively), the dynamics are reversed, since the linear approach consistently achieves higher scores, except for two isolated spikes where the multiplicative approach marginally surpasses the linear approach.

To gain deeper insights into these variations, Fig. 2 was constructed specifically to distinguish the benchmark entities from the worst performers. This approach was adopted to circumvent the potential negative influence of worst performers on the multiplicative approach's average in the prior figures, considering the non-compensatory nature of the model.

Figure 2 provides a distinct perspective on the contrasting behaviours of the best and worst entities within groups B and C in comparison to groups D and E. Notably, within the latter groups (D and E), the worst entities exhibit significantly inferior performance in the multiplicative approach, as the disparity between the multiplicative and linear approaches becomes more pronounced for the worst entities. Moreover, the distance between the best and worst entities is particularly reduced. This proximity occasionally results in instances where the order is reversed, with the best entities occasionally underperforming the worst entities. Consequently, these factors contribute to an overall degradation in the multiplicative score's average which yields a lower average score than the linear approach.

Nevertheless, although the previous paragraph explained why the multiplicative approach yields lower average scores for groups D and E than the linear approach, a closer look at the raw data is needed to fully understand the cause. Notably, groups D and E encompass the largest hospitals, which handle the most complex and specialized cases. The performance of these two groups, is significantly worse than that of groups B and C, largely attributed to their substantial size and their role as the primary healthcare institutions in larger Portuguese cities. While performance scores may appear to follow a uniform scale across all groups, this is misleading. The BoD methodology evaluates efficiency within each group separately, using its own benchmark entities. As a result, the scale is relative to each group's benchmarks, creating the illusion of comparability between groups when, in reality, no direct reference exists for measuring entities across different groups.

Nonetheless, the non-compensatory nature of the multiplicative approach further amplifies this disparity (Rogge 2018), especially for groups D and E. These groups exhibit significantly poorer absolute performance levels and greater variability in their variables. As a result, the efficiency scores of the entities that do not qualify as benchmarks (thus, having values lesser than one) are highly disadvantaged within the multiplicative approach.

5.2 Weight restrictions on the multiplicative approach

In reference to the weight restrictions outlined in the earlier investigation (Vara et al. 2024), it is apparent that models incorporating pie-share restrictions, Eqs. (14) and (16), exhibit a tendency to penalize entities significantly for having very low scores in one or more variables. This effect persists even if these entities are considered benchmarks in all other variables. This outcome results from the inherent nature of these restrictions, which are non-defined at zero and are very sensitive to small numerical variations.

Consequently, this dynamic manifests as pronounced downward dashed spikes in the performance scores for entities exhibiting notably poor values in at least one variable, while entities with lower variability across all variables tend to experience minimal to negligible changes when contrasted with the alternative type of restrictions. Notably, the multiplicative model employing pie-share restrictions introduces a dual penalization effect, originating from both the inherent non-compensatory nature of the model and the heightened sensitivity of these restrictions to small numerical values. This cumulative effect accentuates the severity of the spikes in performance scores, at times approaching extremely close to zero, leaving entities with minimal margin for error. On the other hand, the weight restrictions on the multipliers, as detailed in Eqs. (13) and (15), operated as envisioned, serving as a soft constraint aimed at preventing certain variables from attaining a weight of zero.

Lastly, the models incorporating pie-share restrictions make it easier for experts to express their opinion on the importance of each variable, rather than on the importance of each multiplier. This is because variables are more easily understood and quantified than multipliers, which can be more abstract and difficult to interpret. Therefore, these restrictions can lead to more accurate and intuitive performance evaluations, which can help decision-makers make more informed choices.

5.3 Policy implications

The existing methodology employed by the Portuguese government for hospital performance analysis offers a limited perspective, lacking a comprehensive view or a ranking system for individual hospitals. This study introduces methods that enable the identification of top-performing entities, providing valuable insights into what sets them apart from others. These insights are pivotal for learning lessons and enhancing the performance of other healthcare facilities.

The pandemic forced stringent access restrictions on hospitals, resulting in a significant drop in patient treatment and a subsequent decrease in performance. However, expenses remained constant, leading to an overall reduction in hospital productivity. This impact was not uniform, as the best entities did not get so affected and were also able to recover quickly. Understanding the strategies adopted by these entities is crucial for preventing severe performance and productivity declines at a national level.

Analysis of the annual reports of these leading entities from 2020 onwards reveals a common strategy: anticipation, prediction, and preparation. These entities demonstrated proactive decision-making, staying one step ahead of the crisis. Even as early as January 2020, two months prior to Portugal's first COVID-19 case, these institutions were already implementing measures to prepare for the pandemic. This included reorganizing and expanding spaces and building additional reserves (such as medical supplies, equipment, or even financial reserves, to meet increased demands or unexpected challenges during a crisis) to ensure they were well-prepared for any contingency. These strategies not only helped them weather the storm but also serve as a model for effective crisis management and preparedness in the healthcare sector. The political implications of these findings are clear: policies should encourage and support proactive strategies in healthcare management and preparedness.

6 Conclusions, limitations, and future work

In summary, this study aimed to comprehensively compare the linear and multiplicative BoD approaches, alongside a simultaneous comparison of different weight constraints in the multiplicative BoD method, as a follow up on a previous paper.

The theoretical example showed that the multiplicative approach has two key features. The first is its efficient frontier, which consists of logarithmic curves instead of linear segments, allowing for a better fit to the data. This reduces the distance between the frontier and the remaining data points and enables the inclusion of a larger number of entities and contributing to higher overall scores. The second feature is its non-compensatory nature, where low performance in a single KPI can significantly impact the overall score, even if other KPIs perform well (Rogge 2018). This last characteristic makes the multiplicative approach particularly appropriate for real-world applications, such as in healthcare, where even a single failure can have critical consequences, as human lives are at stake (Petridis et al. 2023; Blancas et al. 2013).

The case study provided a practical comparison of both approaches, leading to three key conclusions. Firstly, the multiplicative approach exhibits sharper performance drops due to its non-compensatory nature, where a low score in a single KPI significantly reduces the overall score. Additionally, in the pessimistic approach, the multiplicative BoD consistently yields lower scores than the linear BoD. This occurs because the right-skewed nature of the linear pessimistic frontier places entities further from the frontier, leading to higher scores.

Secondly, while the optimistic multiplicative BoD generally produces higher scores than the optimistic linear BoD due to its adaptability and better data fit, this is not always the case. The non-compensatory nature of the multiplicative approach penalizes entities with inconsistent KPI performance, as strengths in some areas cannot offset weaknesses in others, unlike in the linear approach. This effect is particularly evident in larger hospitals (groups D and E), which handle more complex cases and experience greater variability. As a result, they show lower scores under the multiplicative approach.

Thirdly, the inclusion of pie-share restrictions in the multiplicative approach amplifies penalties for entities with low scores in at least one KPI, causing sharp declines in performance scores. In contrast, entities with less variability across all KPIs observe only minor changes. These restrictions create a dual penalization effect, as entities are affected both by the imposed restrictions and the non-compensatory nature of the multiplicative approach.

A question that arises from this study is why the reason for the multiplicative approach's relatively low adoption and utilization when compared to the linear approach, despite the evident advantages it offers to the field. The multiplicative BoD implementation is remarkably straightforward and effortless, requiring only minor transformations from the linear BoD.

This study's limitations stem from its methodology, particularly its sensitivity to sample size and variable selection. Changes in these factors can affect results and alter benchmark entities. Expert input or guidance is also crucial to set minimum weight assignments (represented as w_j) for variables and manage missing data effectively.

Future research holds great potential for enhancing these methods. Beyond improving robustness through better techniques for handling data imperfections (Abbasi-Yadkori et al. 2017), integrating machine learning and AI could enable real-time data processing and continuous refinement. By linking these models to a real-time, auto-updating database, predictions could be continuously adjusted as new data becomes available, improving accuracy over time. This dynamic system would provide valuable insights for policymakers, healthcare managers, and the public, supporting better decision-making and resource allocation.

Appendix

See Tables 2, 3, 4 and Figs. 3, 4, 5, 6.

Table 2 List of considered entities and the group of each

Group	Institution
Group B	Centro Hospitalar Póvoa de Varzim/Vila do Conde, EPE
	Centro Hospitalar do Médio Ave, EPE
	Centro Hospitalar do Oeste, EPE
	Hospital Distrital da Figueira da Foz, EPE
	Hospital Santa Maria Maior, EPE
	Unidade Local de Saúde da Guarda, EPE
	Unidade Local de Saúde de Castelo Branco, EPE
	Unidade Local de Saúde do Litoral Alentejano, EPE
	Unidade Local de Saúde do Nordeste, EPE
Group C	Centro Hospitalar Barreiro/Montijo, EPE
	Centro Hospitalar Entre Douro e Vouga, EPE
	Centro Hospitalar Médio Tejo, EPE
	Centro Hospitalar Tâmega e Sousa, EPE
	Centro Hospitalar Universitário Cova da Beira, EPE
	Centro Hospitalar de Leiria, EPE
	Centro Hospitalar de Setúbal, EPE
	Centro Hospitalar do Baixo Vouga, EPE
	Hospital Distrital de Santarém, EPE
	Hospital da Senhora da Oliveira, Guimarães, EPE
	Hospital de Cascais, PPP
	Hospital de Loures, EPE
	Hospital de Vila Franca de Xira, EPE
	Unidade Local de Saúde de Matosinhos, EPE
	Unidade Local de Saúde do Alto Minho, EPE
	Unidade Local de Saúde do Baixo Alentejo, EPE
	Unidade Local de Saúde do Norte Alentejano, EPE
Group D	Centro Hospitalar Tondela-Viseu, EPE
	Centro Hospitalar Trás-os-Montes e Alto Douro, EPE
	Centro Hospitalar Universitário do Algarve, EPE
	Centro Hospitalar Vila Nova de Gaia/Espinho, EPE
	Hospital Espírito Santo de Évora, EPE
	Hospital Fernando Fonseca, EPE
	Hospital Garcia de Orta, EPE
	Hospital de Braga, EPE
Group E	Centro Hospitalar Universitário Lisboa Norte, EPE
	Centro Hospitalar Universitário de Lisboa Central, EPE
	Centro Hospitalar Universitário de São João, EPE
	Centro Hospitalar Universitário do Porto, EPE
	Centro Hospitalar de Lisboa Ocidental, EPE
	Centro Hospitalar e Universitário de Coimbra, EPE

Table 3 Literature review on multiplicative BoD for constructing composite indicators

Author(s) and study	Brief description of the study
Cherchye et al. (2007). An introduction to 'benefit of the doubt' composite indicators	In their study, Cherchye et al. (2007) introduce the linear BoD approach as a method for aggregating KPIs into a single performance score. The unique aspect of this approach is that it assigns individual weights to each entity, which are determined based on computations rather than relying on subjective preferences of stakeholders and DMUs, so that the composite score is maximized. In the concluding remarks, the authors highlight the potential of multiplicative aggregation as a promising direction for future research. They emphasize that this multiplicative approach offers the flexibility to be linearized through the use of logarithms, resulting in a similar structure to the linear BoD approach. This observation suggests that the multiplicative approach can leverage the advantages of the linear BoD while incorporating additional features that address its limitations. Thus, further exploration and investigation of the multiplicative aggregation method within the context of composite indicator construction are warranted to uncover its full potential and enhance the understanding of its applicability in practical settings
Emrouznejad & Cabanda (2010). An aggregate measure of financial ratios using a multiplicative DEA model	Emrouznejad & Cabanda (2010) show that when arithmetic operations between ratios are considered, two main problems arise, which render the results meaningless. The first one, the convexity issue, appears because the weighted sum of ratios is not the same as the ratio between the weighted sum of nominators and the weighted sum of denominators. The second is the proportionality issue which arises when data are in ratios, because if the nominator and the denominator changed proportionally, the ratio will not be changed. Therefore, the authors propose that the multiplicative model should be used instead, with a log-linear transformation
Zhou et al. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach	In this study, Zhou et al. (2010) propose a multiplicative optimization approach that can be turned into a linear optimization through the natural logarithmic similarly to Emrouznejad & Cabanda (2010). The distinction lies on the right-hand side of the constraints; instead of having 'h' as the constraint, the authors replace it with Euler's number, making the model invariant to any changes made on the right-hand side. Then, the authors provide two multiplicative optimization models, one seeking the best set of weights for aggregation and the other seeking the worst set of weights. Finally, both models, best and worst, are combined into one measure of performance, based on a parameter that weights how much each model contributes to the final score. Their proposed multiplicative optimization approach is not invariant to the measurement units of sub-indicators
Puyenbroeck & Rogge (2017). Multiplicative mean quantity index numbers with Benefit-of-the-Doubt weights	In this paper, unlike the previous works where direct multiplicative approaches are applied, Van Puyenbroeck & Rogge (2017) propose a two-step method to avoid a rescaling of the data and the loss of its information. This indirect approach first solves a Linear BoD model and secondly, with the weights generated from it, a weighted multiplicative mean is done. The authors, by doing this, assign shadow prices to each variable and give bias to the model. Nevertheless, by initially using linear BoD, the model ends up being compensatory
Verbunt & Rogge (2018). Multiplicative composite indicators with compromise Benefit-of-the-Doubt weights	This work builds on the suggestions made by Van Puyenbroeck & Rogge (2017), by aggregating optimistic and pessimistic weights in their two-step method. Alternatively, instead of using entity specific benchmarks, Verbunt & Rogge (2018) use a fixed, entity as the benchmark. However, in a fixed benchmark setting, the weights are sensitive to the variability of the sub-indicators. Verbunt and Rogge therefore apply a max–min normalisation such that the normalized sub-indicator values fully reflect DMUs' comparative performances

Table 3 (continued)

Author(s) and study	Brief description of the study
Rogge (2018). Composite indicators as generalized benefit-of-the-doubt weighted averages	Rogge (2018) extends the two-step approach developed by Puyenbroeck and Rogge (2017) by incorporating the Hölder order parameter which introduces a spectrum of decision-making perspectives. When the Hölder order approaches $-\infty$, the composite indicator prioritizes the worst KPIs for each DMU, leading to a more cautious evaluation. Conversely, for larger orders, the model imposes lower penalties on poor performance in certain dimensions, resulting in a less conservative approach that focuses primarily on KPIs with favourable outcomes. This innovative methodology offers enhanced flexibility in constructing composite indicators and addresses decision-making uncertainty in complex policy contexts. However, further empirical evaluation is necessary to assess its robustness and applicability across various domains
Dominguez-Gil et al. (2022). A multiplicative composite indicator to evaluate educational systems in OECD countries	Dominguez-Gil et al. (2022) built a multidimensional composite indicator which uses a multiplicative aggregation for two periods allowing for inter-temporal comparisons. The proposed approach offers a dynamic analysis of entities over time, based on Verbunt and Rogge (2018), enabling the detection of changes in performance, baseline influence, and weighting impact. Although the methodology facilitates a first comparative analysis, it emphasizes that final policy decisions should not solely focus on improving the composite indicator rank but instead be informed by a comprehensive analysis of each entity's situation
Mariani & Ciommi (2022). Aggregating Composite Indicators through the Multiplicative Mean: A Penalization Approach	The authors addressed the issue of indicator aggregation using a non-compensative approach through a penalization factor that captures the unbalance among indicators. The penalized multiplicative mean is computed as the product of the multiplicative mean and a penalization factor, which represents a reliability measure for the multiplicative mean and accounts for the unbalance among indicators. The authors apply this approach to construct the penalized Human Development Index (pHDI) and compare it to the traditional HDI for various countries. The results show that the pHDI preserves the ranking provided by the multiplicative mean while offering more accurate distinctions for countries with poor performances. The proposed method provides a promising way to address the compensability issue in composite indicators and offers potential for generalization to other aggregation functions
Petridis et al. (2023). Ranking econometric techniques using multiplicativeal Benefit of Doubt	Petridis et al. focus on the evaluation and ranking of econometric techniques used for forecasting stock prices, considering various error measures to capture the accuracy of each technique. The authors propose a multiplicativeal BoD model, resembling a Weighted Multi-Objective Product Model that can handle various error measures, and additional constraints can be introduced to further customize the ranking. The authors argue that the multiplicative approach must be used when dealing with rates, growth rates, or ratios, although it requires all values to be positive (or non-zero)
Ferreira et al. (2023). A multiplicative aggregation of performance indicators considering regulatory constraints: An application to the urban solid waste management	Ferreira et al. (2023) propose a multiplicative BoD approach which accounts for: regulatory constraints, through the use of targets and thresholds to assure that the minimum requirements are met; deals with missing data, by using Monte-Carlo simulation; works with distributions of performance, rather than fixed values; and provides a correct treatment of undesirable variables. The authors construct a frontier such that all targets with respect to it obey all constraints, and it can be straightforwardly adapted to any performance analysis featuring imperfect data knowledge. The paper finishes by applying the model on data of the Portuguese urban solid waste management utilities and giving possible improvements. However, the authors acknowledge the potential infeasibility of the model for some entities due to restrictive regulations, which warrants further investigation

Table 4 List of considered variables: description and authors that have used the same variable

Variables	Description	Authors
(V1) Rate of first medical appointments within time	Number of first medical appointments within the maximum legal time of response for a first appointment in hospitals divided by the number of first medical appointments in percentage	Matos et al. (2021); Pederneiras et al. (2023); Castro et al. (2014); Ferreira & Marques (2021)
(V2) Rate of enrolled in surgical list within time	Number of patients enrolled in surgical list within time divided by the total number of patients enrolled in the surgical list in percentage	Ferreira & Marques (2021)
(V3) Pressure ulcer rate per 1.000 inpatients	Number of episodes with pressure ulcers divided by the number of episodes with exclusions for episodes with ulcers per thousand inpatients	Ferreira & Marques (2021)
(V4) Central venous catheter-related bloodstream infections rate	Number of central venous catheter-related bloodstream infections episodes divided by the number of episodes with exclusions for bloodstream infections episodes per thousand inpatients	Ferreira & Marques (2021)
(V5) Postoperative pulmonary embolism/deep vein thrombosis per 100.000 inpatients	Number of postoperative pulmonary embolism or deep vein thrombosis episodes divided by the number of postoperative pulmonary embolism or deep vein thrombosis exclusion episodes per 100.000 surgical procedures	Matos et al. (2021); Ferreira & Marques (2021)
(V6) Postoperative septicaemia rate per 100.000 inpatients	Number of episodes with postoperative sepsis divided by the number of episodes with exclusions for episodes with postoperative sepsis per 100.000 surgical procedures	Matos et al. (2021); Ferreira & Marques (2021)
(V7) Rate of instrumental vaginal deliveries with 3rd and 4th degree lacerations	Number of episodes of obstetric trauma during instrumental vaginal delivery divided by the total number of episodes with exclusions for episodes with obstetric trauma during instrumental vaginal delivery in percentage	Matos et al. (2021); Pederneiras et al.2023
(V8) Rate of non-instrumented vaginal deliveries with 3rd and 4th degree lacerations	Number of Episodes of Obstetric Trauma During Non-Instrumental Vaginal Delivery divided by the total number of Episodes with Exclusions for Episodes with Obstetric Trauma During Non-Instrumental Vaginal Delivery in percentage	Matos et al. (2021); Pederneiras et al. (2023)
(V9) Standard patients per FTE doctor	Standard patients[1] divided by the Full-time Equivalent (FTE) doctors	Pederneiras et al. (2023); Castro et al. (2014); Ferreira & Nunes (2018)
(V10) Standard patients per FTE nurse	Standard patients divided by the FTE nurses	Pederneiras et al. (2023); Castro et al. (2014); Ferreira & Nunes (2018)
(V11) Inpatient bed occupancy rate	Number of acute admission days divided by the number of acute beds multiplied by 30.4375[2] and by the accumulated number of months in percentage	Pederneiras et al. (2023)
(V12) Waiting time before surgery	Total number of days until surgery in homogeneous diagnostic groups scheduled inpatient surgical episodes divided by the total number of homogeneous diagnostic groups Scheduled Inpatient Surgical episodes	Matos et al. (2021); Pederneiras et al. (2023); Ferreira & Marques (2021)
(V13) Rate of outpatient surgeries on potential outpatient procedures	Number of outpatient surgical episodes with outpatient procedures divided by the number of inpatient and outpatient surgical episodes with outpatient procedures in percentage	Matos et al. (2021); Pederneiras et al. (2023); Castro et al. (2014); Ferreira & Marques (2021)

Table 4 (continued)

Variables	Description	Authors
(V14) Rate of readmissions within 30 days after discharge	Number of readmissions within 30 days of discharge divided by the total number of admissions with discharge in the period in percentage	Matos et al. (2021); Chowdhury and Zelenyuk (2016); Dahl and Kongstad (2017); Ferreira & Marques (2021)
(V15) Rate of inpatients staying more than 30 days	Number of admissions with stay of over 30 days divided by the total number of admissions with discharge in the period in percentage	Matos et al. (2021); Baek et al. (2018); Ferreira & Marques (2021)
(V16) Rate of hip fracture surgery in the first 48 h	Total number of patient episodes aged > = 65 years, with main diagnosis 820, with surgery performed within the first 48 h after admission divided by the total number of patient episodes aged > = 65 years, with main diagnosis 820, with surgery performed in percentage	Matos et al. (2021); Gutacker et al. (2016); Lee and Elfar (2014); Ferreira & Marques (2021)

Standard patients[1] → Hypothetical patient with a specific set of medical conditions and characteristics that are used to standardize the measurement and allow comparison of healthcare outcomes

30.4375[2] → Number of days in each month, 365.25/12 = 30.4375

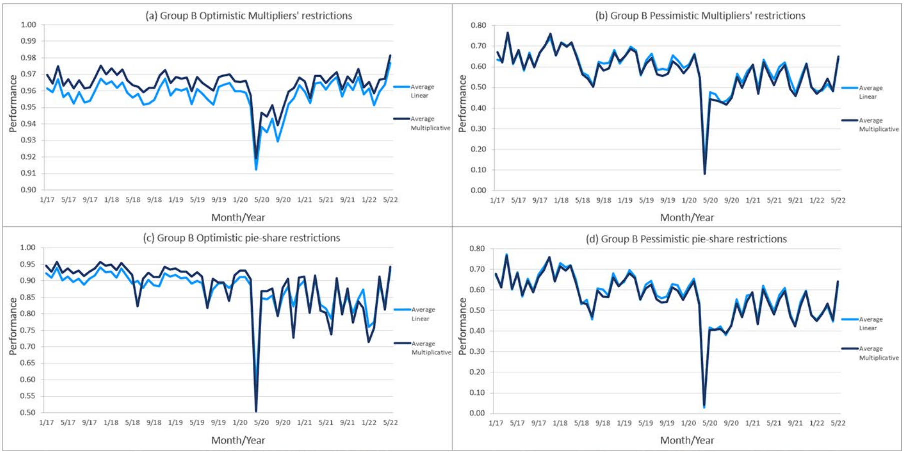

Fig. 3 Comparisons between the Linear (AVG LIN) and Multiplicative (AVG GEO) approaches for group B

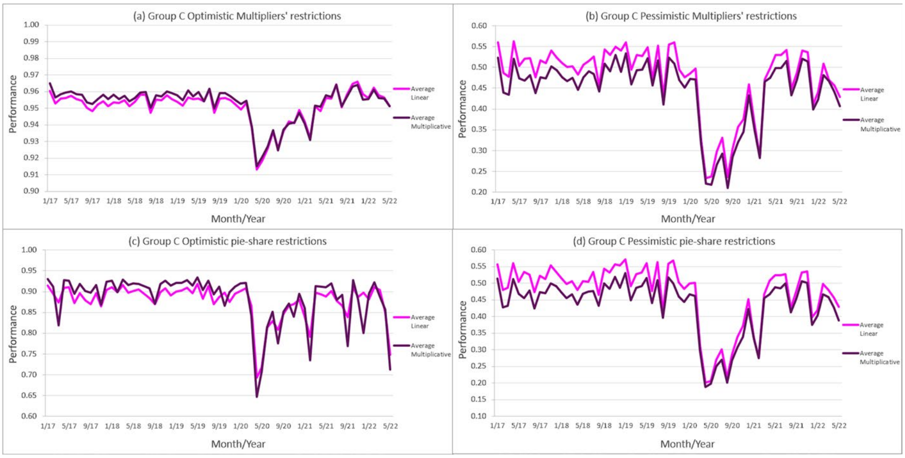

Fig. 4 Comparisons between the Linear (AVG LIN) and Multiplicative (AVG GEO) approaches for group C

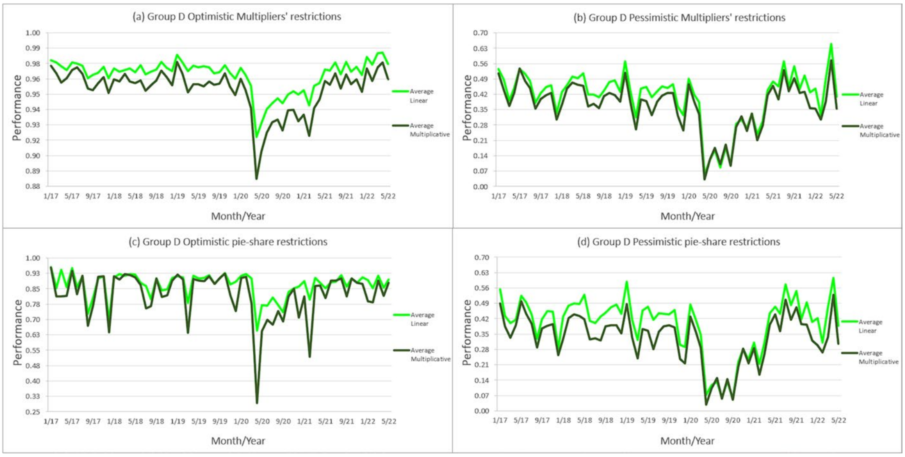

Fig. 5 Comparisons between the Linear (AVG LIN) and Multiplicative (AVG GEO) approaches for group D

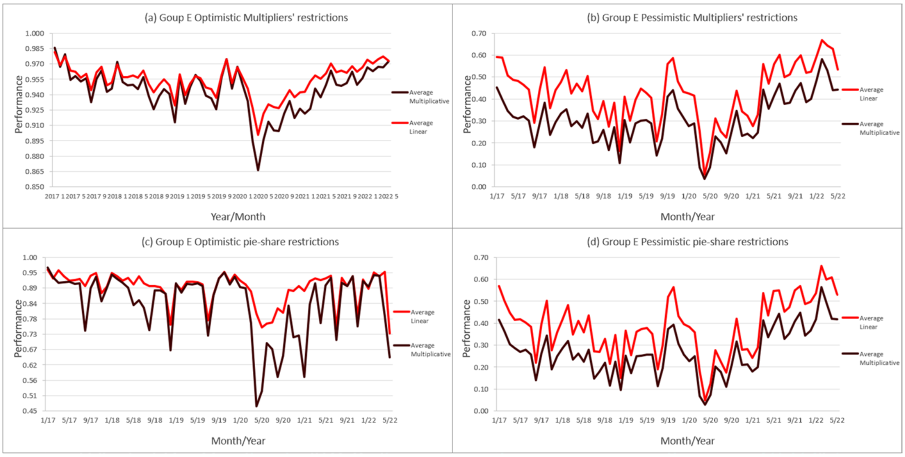

Fig. 6 Comparisons between the Linear (AVG LIN) and Multiplicative (AVG GEO) approaches for group E

Acknowledgements This work is financed by Portuguese funds through the FCT - Foundation for Science and Technology, I.P., under the project UID/97/2025 (CEGIST).

Funding Open access funding provided by FCT|FCCN (b-on). Guilherme Vara recognizes the support from the Portuguese National Foundation for Science and Technology (FCT) through his Ph.D. scholarship (2025.01376.BD).

Declarations

Competing interests The authors declare that they have no competing interests.

References

Abbasi-Yadkori Y, Bartlett P, Gabillon V, Malek A (2017) Hit-and-run for sampling and planning in non-convex spaces. In: Artificial intelligence and statistics. PMLR, pp 888–895

Afonso GP, Ferreira DC, Figueira JR (2023) A Network-DEA model to evaluate the impact of quality and access on hospital performance. Ann Oper Res. https://doi.org/10.1007/s10479-023-05362-x

Afonso GP, Figueira JR, Ferreira DC (2024) Dealing with uncertainty in healthcare performance assessment: a fuzzy network-DEA approach with undesirable outputs. Int Trans Oper Res. https://doi.org/10.1111/itor.13490

Albo Y, Lanir J, Bak P, Rafaeli S (2015) Off the radar: comparative evaluation of radial visualization solutions for composite indicators. IEEE Trans vis Comput Graph 22(1):569–578

Ave P, Babolsar I (2010) Environmental performance index and economic growth: evidence from some developing countries. Aust J Basic Appl Sci 4(8):3098–3102

Baek H, Cho M, Kim S, Hwang H, Song M, Yo S (2018) Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. PloS one 13(4):e0195901.

Baganha MI, Ribeiro JS, Pires S (2002) O sector da saúde em Portugal: funcionamento do sistema e caracterização sócio-profissional. Oficina Do CES 182:1–33

Banker RD, Maindiratta A (1986) Piecewise loglinear estimation of efficient production surfaces. Manag Sci 32(1):126–135

Barclay M, Dixon-Woods M, Lyratzopoulos G (2019) The problem with composite indicators. BMJ Qual Saf 28(4):338–344

Barros PP, Machado SR, Simões JDA, World Health Organization (2011) Portugal: health system review

Blancas FJ, Contreras I, Ramírez-Hurtado JM (2013) Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. J Oper Res Soc 64(5):668–678

Booysen F (2002) An overview and evaluation of composite indices of development. Soc Indic Res 59:115–151

Burgass MJ, Halpern BS, Nicholson E, Milner-Gulland EJ (2017) Navigating uncertainty in environmental composite indicators. Ecol Ind 75:268–278

Castro, R., Portela, M., Camanho, A. (2014). Benchmarking dos serviços dos hospitais portugueses: uma aplicação de data envelopment analysis. In: Investigação operacional em ação: casos de aplicação. Imprensa da Universidade de Coimbra, pp 703–740

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. European journal of operational research 2(6):429-444.

Cherchye L, Moesen W, Rogge N, Puyenbroeck TV (2007) An introduction to 'benefit of the doubt'composite indicators. Soc Indic Res 82:111–145

Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T, Saisana M, Saltelli A, Tarantola S (2008) Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. J Oper Res Soc 59(2):239–251

Chowdhury H, Zelenyuk V (2016) Performance of hospital services in Ontario: DEA with truncated regression approach. Omega, 63:111-122.

Condes, E., & Arribas, J. R. (2021). Impact of COVID-19 on Madrid hospital system.

Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software, vol 2. Springer, New York, p 489

Dahl CM, Kongstad LP (2017) The costs of acute readmissions to a different hospital–does the effect vary across provider types? Social Science & Medicine 183:116-125.

Dang HAH, Nguyen CV, Carletto C (2023) Did a successful fight against COVID-19 come at a cost? Impacts of the pandemic on employment outcomes in Vietnam. World Dev 161:106129

Decancq K, Lugo MA (2013) Weights in multidimensional indices of wellbeing: an overview. Econom Rev 32(1):7–34

Dialga I, Thi Hang Giang L (2017) Highlighting methodological limitations in the steps of composite indicators construction. Soc Indic Res 131:441–465

Dias RN, Filipe RM, Matos HA (2023) Sustainability analysis of a solar-driven calcium looping plant for thermochemical energy storage. J Clean Prod 429:139551

Dias RN, Filipe RM, Matos HA (2024) Decision-making based on sustainability analysis using GREENSCOPE. Clean Technol Environ Policy 26(3):755–770

Dominguez-Gil C, Segovia-Gonzalez MM, Contreras I (2022) A multiplicative composite indicator to evaluate educational systems in OECD countries. Compare 52(8):1296–1313

Ebrahimi B, Tavana M, Charles V (2021) A note and new extensions on "interval efficiency measures in data envelopment analysis with imprecise data." Oper Res 21(4):2719–2737

El Gibari S, Gómez T, Ruiz F (2019) Building composite indicators using multicriteria methods: a review. J Bus Econ 89(1):1–24

Emrouznejad A, Cabanda E (2010) An aggregate measure of financial ratios using a multiplicative DEA model. Int J Financ Serv Manag 4(2):114–126

Ferreira DC, Caldas P, Varela M, Marques RC (2023) A multiplicative aggregation of performance indicators considering regulatory constraints: an application to the urban solid waste management. Expert Syst Appl 218:119540

Ferreira DC, Marques RC (2021) Public-private partnerships in health care services: do they outperform public hospitals regarding quality and access? Evidence from Portugal. Socio-Econ Plan Sci 73:100798

Ferreira DC, Marques RC, Nunes AM (2021) Pay for performance in health care: a new best practice tariff-based tool using a log-linear piecewise frontier function and a dual–primal approach for unique solutions. Oper Res 21:2101–2146

Ferreira DC, Marques RC, Pedro MI (2018) Explanatory variables driving the technical efficiency of European seaports: an order-α approach dealing with imperfect knowledge. Transp Res E Logist Transp Rev 119:41–62

Ferreira DC, Nunes AM (2019) Technical efficiency of Portuguese public hospitals: A comparative analysis across the five regions of Portugal. The International journal of health planning and management, 34(1):e411-e422.

Färe R, Karagiannis G, Hasannasab M, Margaritis D (2019) A benefit-of-the-doubt model with reverse indicators. Eur J Oper Res 278(2):394–400

Freudenberg M (2003) Composite indicators of country performance: a critical assessment.

Gómez-Limón JA, Riesgo L (2009) Alternative approaches to the construction of a composite indicator of agricultural sustainability: an application to irrigated agriculture in the Duero Basin in Spain. J Environ Manag 90(11):3345–3362

Greco S, Ehrgott M, Figueira J (2016) Multiple criteria decision analysis, 2nd ed. Springer, New York

Greco S, Ishizaka A, Tasiou M, Torrisi G (2019) On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. Soc Indic Res 141:61–94

Gulati R, Kattumuri R, Kumar S (2020) A non-parametric index of corporate governance in the banking industry: an application to Indian data. Socio-Econ Plan Sci 70:100702

Gutacker N, Siciliani L, Moscelli G, Gravelle H (2016) Choice of hospital: which type of quality matters? J Health Econ 50:230–246

Hatefi SM, Torabi SA (2010) A common weight MCDA–DEA approach to construct composite indicators. Ecol Econ 70(1):114–120

Hjalmarsson L, Kumbhakar SC, Heshmati A (1996) DEA, DFA and SFA: a comparison. J Prod Anal 7:303–327

Hudrliková L (2013) Composite indicators as a useful tool for international comparison: the Europe 2020 example. Prague Econ Pap 22(4):459–473

Jiménez-Fernández E, Sánchez A, Pérez ES (2022) Unsupervised machine learning approach for building composite indicators with fuzzy metrics. Expert Syst Appl 200:116927

Juwana I, Muttil N, Perera BJC (2012) Indicator-based water sustainability assessment—a review. Sci Total Environ 438:357–371

Kao C (2016) Network data envelopment analysis: Foundations and extensions, vol 240. Springer, Cham

Karagiannis G (2017) On aggregate composite indicators. J Oper Res Soc 68(7):741–746

Kaye AD, Okeagu CN, Pham AD, Silva RA, Hurley JJ, Arron BL, Cornett EM (2021) Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives. Best Pract Res Clin Anaesthesiol 35(3):293–306

Lafuente E, Araya M, Leiva JC (2022) Assessment of local competitiveness: a composite indicator analysis of Costa Rican counties using the 'Benefit of the Doubt'model. Socio-Econ Plan Sci 81:100864

Lee DJ, Elfar JC (2014) Timing of hip fracture surgery in the elderly. Geriatr Orthop Surg Rehabil 5(3):138–140

Lindén D, Cinelli M, Spada M, Becker W, Gasser P, Burgherr P (2021) A framework based on statistical analysis and stakeholders' preferences to inform weighting in composite indicators. Environ Model Softw 145:105208

Li T, Zhang H, Yuan C, Liu Z, Fan C (2012) A PCA-based method for construction of composite sustainability indicators. Int J Life Cycle Assess 17:593–603

Lyytimäki J, Tapio P, Varho V, Söderman T (2013) The use, non-use and misuse of indicators in sustainability assessment and communication. Int J Sustain Dev World Ecol 20(5):385–393

Mariani F, Ciommi M (2022) Aggregating composite indicators through the multiplicative mean: a penalization approach. Computation 10(4):64

Marques RC, Simões P, Pinto FS (2018) Tariff regulation in the waste sector: an unavoidable future. Waste Manag 78:292–300

Matos R, Ferreira D, Pedro MI (2021) Economic analysis of Portuguese public hospitals through the construction of quality, efficiency, access, and financial related composite indicators. Soc Indic Res 157:361–392

Melyn W, Moesen W (1991) Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. Public Econ Res papers, 1–24
Mendes M, Costa J, Torres J (2010) Continued care in Portugal: a comprehensive overview. Health Policy 94(2):98–105
Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for composite indicators building. Eur Comm, Ispra 15(1):19–20
Nunes AM, Ferreira DC (2019) Reforms in the Portuguese health care sector: challenges and proposals. Int J Health Plann Manage 34(1):e21–e33
Nunes AM, Ferreira DFDC (2022) Evaluating Portuguese public hospitals performance: any difference before and during COVID-19? Sustainability 15(1):294
OECD (2005) Statistics, knowledge and policy: Key indicators to inform decision making. OECD Publishing, Paris. https://doi.org/10.1787/9789264009011-en
Olesen OB, Petersen NC, Podinovski VV (2015) Efficiency analysis with ratio measures. Eur J Oper Res 245(2):446–462
Pederneiras YM, Pereira MA, Figueira JR (2023) Are the Portuguese public hospitals sustainable? a triple bottom line hybrid data envelopment analysis approach. Int Trans Oper Res 30(1):453–475
Pereira MA, Ferreira DC, Figueira JR, Marques RC (2021) Measuring the efficiency of the Portuguese public hospitals: a value modelled network data envelopment analysis with simulation. Expert Syst Appl 181:115169
Petridis K, Petridis NE, Abdelaziz FB, Masri H (2023) Ranking econometric techniques using geometrical benefit of doubt. Ann Oper Res 330(1):411–430
Podinovski VV, Kuosmanen T (2011) Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions. Eur J Oper Res 211(3):577–585
Rogge N (2018) Composite indicators as generalized benefit-of-the-doubt weighted averages. Eur J Oper Res 267(1):381–392
Sagar AD, Najam A (1998) The human development index: a critical review. Ecol Econ 25(3):249–264
Sahoo BK, Singh R, Mishra B, Sankaran K (2017) Research productivity in management schools of India during 1968–2015: a directional benefit-of-doubt model analysis. Omega 66:118–139
Saltelli A (2007) Composite indicators between analysis and advocacy. Soc Indic Res 81:65–77
Santeramo FG (2015) On the composite indicators for food security: decisions matter! Food Rev Int 31(1):63–73
Sharpe A, Andrews B (2012) An assessment of weighting methodologies for composite indicators: The case of the index of economic well-being. CSLS.
Simões J, Augusto GF, Fronteira I (2017) Introduction of freedom of choice for hospital outpatient care in Portugal: implications and results of the 2016 reform. Health Policy 121(12):1203–1207
Simões P, De Witte K, Marques RC (2010) Regulatory structures and operational environment in the Portuguese waste sector. Waste Manag 30(6):1130–1137
Sueyoshi T, Sekitani K (2009) An occurrence of multiple projections in DEA-based measurement of technical efficiency: theoretical comparison among DEA models from desirable properties. Eur J Oper Res 196(2):764–794
Tiedemann T, Francksen T, Latacz-Lohmann U (2011) Assessing the performance of German Bundesliga football players: a non-parametric metafrontier approach. Cent Eur J Oper Res 19:571–587
Tofallis C (2014) On constructing a composite indicator with multiplicative aggregation and the avoidance of zero weights in DEA. J Oper Res Soc 65:791–792
Van Puyenbroeck T, Rogge N (2017) Multiplicative mean quantity index numbers with benefit-of-the-doubt weights. Eur J Oper Res 256(3):1004–1014
Vara GM, Gomes MC, Ferreira DC (2024) Assessing the performance of Portuguese public hospitals before and during COVID-19 outbreak, with optimistic and pessimistic benchmarking approaches. Health Care Manage Sci. https://doi.org/10.1007/s10729-024-09693-4
Verbunt P, Rogge N (2018) Multiplicative composite indicators with compromise benefit-of-the-doubt weights. Eur J Oper Res 264(1):388–401
Vieira I, Ferreira D, Pedro MI (2023) The satisfaction of healthcare consumers: analysis and comparison of different methodologies. Int Trans Oper Res 30(1):545–571
Vértesy D (2016) A Critical assessment of the quality and validity of composite indicators of innovation. In: 21st international conference on science and technology indicators-STI 2016. Book of Proceedings.
Wang YM, Greatbanks R, Yang JB (2005) Interval efficiency assessment using data envelopment analysis. Fuzzy Sets Syst 153(3):347–370

Zanella A, Camanho AS, Dias TG (2015) Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. Eur J Oper Res 245(2):517–530
Zhou P, Ang BW (2009) Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. Soc Indic Res 94:83–96
Zhou P, Ang BW, Zhou DQ (2010) Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. Soc Indic Res 96:169–181

Authors and Affiliations

Guilherme Vara[1] · Marta Castilho Gomes[2] · Paulo Caldas[1,3,4] · Miguel Varela[3,5] · Diogo Cunha Ferreira[1]

✉ Guilherme Vara
guilherme.vara@tecnico.ulisboa.pt

Marta Castilho Gomes
marta.gomes@tecnico.ulisboa.pt

Paulo Caldas
paulo.caldas@isg.pt

Miguel Varela
miguel.varela@isg.pt

Diogo Cunha Ferreira
diogo.cunha.ferreira@tecnico.ulisboa.pt

1 CEGIST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

2 CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

3 Business and Economics School, Instituto Superior de Gestão, Av. Mal. Craveiro Lopes 2A, 1700-284 Lisbon, Portugal

4 University of New England, Armidale, NSW 2350, Australia

5 CEFAGE, Faculdade de Economia, Universidade do Algarve, Campus of Gambelas, 8005-139 Faro, Portugal