OR Spectrum
https://doi.org/10.1007/s00291-025-00832-z

ORIGINAL ARTICLE

Robust data envelopment analysis models for efficiency evaluation with new uncertainty sets

Aliasghar Arabmaldar[1] · Adel Hatami-Marbini[2]

Received: 8 January 2025 / Accepted: 21 July 2025

Abstract

The integration of robust optimisation techniques and data envelopment analysis (DEA) models results in a methodology called *robust DEA*. This methodology aims to tackle uncertain data and ensure robust and reliable efficiency measures. In applying robust optimisation approaches, the selection of the uncertainty set plays a pivotal role since it determines the trade-off between achieving optimal objective and ensuring a high probability of constraint feasibility, a concept well-known as the *price of robustness*. This trade-off can be adjusted using a robust parameter based on managers' risk preferences. Similar to robust optimisation, robust DEA aims to protect the deterministic DEA models against data uncertainty within a user-specified uncertainty set, providing a probability bound on constraint feasibility. Despite recent advancements in robust optimisation approaches, robust DEA models are still in their early stages of development, accentuating the need for further research, especially in the application of new types of uncertainty sets. To address the identified research gap, this study aims to develop two novel robust DEA models considering recently introduced uncertainty sets—namely, variable budgeted and order statistic uncertainty sets—to improve the flexibility and generality of the existing robust DEA models. We discuss in depth how the existing robust DEA models under budgeted uncertainty sets represent a special case of the proposed robust DEA models in this paper when the robust parameter is appropriately selected. Finally, we present a case study on EU banks to illustrate the efficacy and applicability of the proposed models, which show a robust evaluation strategy for management in uncertain environments.

Keywords Robust data envelopment analysis (DEA) · Robust optimisation · Uncertainty sets · Efficiency measurement · EU banking

Extended author information available on the last page of the article

1 Introduction

Data envelopment analysis (DEA) is a non-parametric frontier analysis method widely used to assess the performance of decision-making units (DMUs) that use multiple inputs to produce multiple outputs. Within the DEA framework, envelopment and multiplier models are two various approaches for evaluating the efficiency of DMUs, with envelopment models constructing an efficient frontier from observed data and multiplier models considering the relationship between inputs and outputs by applying their weights (Cooper et al. 2007). Given that DEA heavily relies on data, the precision and robustness of efficiency measurements are significantly affected by the quality of the data (Hatami-Marbini and Arabmaldar 2021). The original CCR (Charnes et al. 1978) and BCC (Banker et al. 1984) models, along with their subsequent developments, have been widely applied to evaluate relative efficiency across various real-world problems such as healthcare, education, banking, and manufacturing. These models commonly assume deterministic input and output data, thus overlooking uncertainties inherent in practical applications. The DEA literature provides substantial evidence of the critical role of uncertainty in performance assessment applications (see e.g., Hatami-Marbini & Arabmaldar; 2021; Olesen and Petersen 2016; Zhu 2003). Therefore, neglecting uncertainty can lead to doubts about the application of the DEA method for performance benchmarking, as decision-makers may question the accuracy of DEA's efficiency estimates. This issue is particularly pronounced in situations marked by high levels of uncertainty due to frequent and substantial disruptions such as in financial markets (Zervopoulos et al. 2023). Such conditions make it difficult to accurately and reliably assess performance and make informed managerial decisions to enhance the efficiency of underperforming units. Consequently, enhancing the robustness of classical DEA models to handle significant data uncertainty is essential. Addressing uncertainty has long been a core focus in DEA applications, leading to extensive research aimed at developing solutions within the uncertain DEA framework. Key approaches include chance-constrained and stochastic DEA (Olesen and Petersen 1995, 2016), bootstrap DEA (Pham et al. 2023; Simar and Wilson 1998), imprecise and interval DEA (Despotis and Smirlis 2002; Mostafaee and Saljooghi 2010; Akbarian 2020; Toloo et al. 2021), fuzzy DEA (Hatami-Marbini et al. 2011; Hatami-Marbini 2019), and robust DEA (Shokouhi et al. 2010; Hatami-Marbini et al. 2022a, b; Arabmaldar et al. 2023).

It should be noted that robust DEA, while sharing the common goal of tackling data uncertainty with other methodologies, such as the statistical-based robust non-parametric estimation techniques (Daraio and Simar 2007) and the chance-constrained DEA approach (Olesen and Petersen 2016), differs significantly. Robust DEA immunises against data uncertainty through the use of uncertainty sets, offering flexibility and robustness without relying on predefined probabilistic assumptions, unlike stochastic production models (e.g., Chambers and Quiggin 2000; Olesen and Petersen 2016; Li et al. 2024), which assume specific error distributions. This distinction is particularly significant in DEA applications with small sample sizes, where the use of stochastic assumptions may be problematic due to limited empirical support (Olesen and Petersen 2016; Sengupta 1992). Moreover, robust DEA integrates robust optimisation techniques with classical DEA models to deal effectively with data uncertainty, all while maintaining similar axiomatic assumptions (Arabmaldar et al. 2023, 2024).

Robust DEA, which is the main methodological framework in this study, has emerged as a leading approach in recent literature for addressing uncertain data in DEA (Arabmaldar et al. 2024; Li et al. 2024; Hatami-Marbini et al. 2022a, b). Robust DEA has several advantages, making it highly suitable for tackling uncertainty in measuring efficiency. Unlike chance-constrained DEA models (Olesen and Petersen 1995), robust DEA does not require precise distribution functions of uncertain parameters, which are often unknown and must be estimated from historical data that may be biased or unavailable. Furthermore, robust DEA does not rely on probability distributions and statistical properties of input and output data, which are frequently missing in uncertain situations but are necessary for methods such as bootstrap DEA (e.g., Daraio and Simar 2007). In robust DEA, results are not sensitive to the precise identification of lower and upper bounds of relative efficiencies, which is necessary in interval DEA models (Despotis and Smirlis 2002). Moreover, in contrast to fuzzy DEA (Emrouznejad et al. 2014; Hatami-Marbini et al. 2011), robust DEA is widely applicable to real-world situations without needing standardised rules for assuming membership functions of inputs and outputs, which can be complex. Since robust DEA employs robust optimisation techniques to cope with uncertainty, it is crucial and valuable to discuss and review the main robust optimisation approaches in this study.

1.1 Robust optimisation

Robust optimisation is one of the most popular techniques in the field of optimisation, gaining increasing attention in recent years. Robust optimisation aims to find an optimal (and robust) solution that optimises the objective function value along with exhibiting the least sensitivity to possible perturbations (e.g., Ben-Tal and Nemirovski 2000; Bertsimas and Sim 2004). It often suggests a deterministic robust counterpart problem while ensuring feasibility across all possible realisations within a predefined uncertainty set. Robust optimisation is particularly useful when defining uncertainty through probability distributions is impracticable. For a comprehensive overview of robust optimisation, readers are referred to Ben-Tal et al. (2009).

Selecting the *uncertainty set* is a crucial factor in effectively applying the robust optimisation approach. The uncertainty set in robust optimisation determines the trade-off between two competing objectives: (*i*) achieving the optimal objective function value, and (*ii*) ensuring constraint feasibility with a high probability. This trade-off is often referred to as the *price of robustness* (PoR) by Bertsimas and Sim (2004). The balance between these objectives depends on two key features of the selected uncertainty set (Ben-Tal et al. 2010; Gregory et al. 2011). First, the size of the uncertainty set is vital and is typically determined by managers based on their level of conservatism. A smaller uncertainty set commonly leads to an improved objective function value but reduces the probability of constraint feasibility, demonstrating that an improvement in one aspect often results in a decline in the other. Second, the geometric flexibility of the uncertainty set plays a crucial role (Dehghani Filabadi and Mahmoudzadeh 2022; Zhang and Gupta 2023). Improving both the objective function value and constraint feasibility against data perturbation can be achieved if the uncertainty set includes regions with higher probabilities of uncertain situations while excluding those that are exceedingly improbable. Therefore, a more geometrically flexible uncertainty set, shaped to encompass high-probability regions, can bring about this balance. In summary, the primary goal of robust optimisation is to find ways to reduce the PoR.

Here, we provide a brief overview of popular uncertainty sets widely used in the robust optimisation literature. The *interval uncertainty* set, also known as the box uncertainty set, was discussed by Soyster (1973). This set furnishes substantial protection, though it is often considered conservative, as it identifies the optimal solution under the worst-case scenario of the unknown parameters. The *ellipsoidal uncertainty* set, introduced by Ben-Tal and Nemirovski (1998), is based on the standard deviation formula and results in a quadratic form. This set provides a more refined approach compared to the interval set, accounting for correlations between uncertainties. The *budgeted uncertainty* set, proposed by Bertsimas and Sim (2004), is the first polyhedral uncertainty set. It effectively balances protection and conservativeness in robust optimisation models by imposing a budget constraint on the sum of all random variables, ensuring they do not all take the extreme value of 1. The *demand uncertainty* set is based on the generalised central limit theorem (Bandi et al. 2015; Bandi and Gupta 2020). The *discrete uncertainty* set consists of a finite collection of distinct scenarios, each representing a potential realisation of uncertain parameters (Goerigk et al. 2022; Goerigk and Khosravi 2023).

Bertsimas and Sim (2004)'s approach has been extensively used for various operational research problems due to its ability to address over-conservatism in robust optimisation. However, despite its popularity, this approach is criticised for its inherent hidden conservatism, which can potentially limit practical applicability and efficiency (Poss 2013). To impede this issue, Poss (2013, 2014) introduced *variable budgeted uncertainty* as a more flexible generalisation that reduces the PoR and balances protection with practical performance by being less conservative. Goerigk et al. (2022) expanded upon the concept of two-stage robust optimisation problems by introducing the notion of *two-stage budgeted uncertainty*, considering both discrete and continuous cases. Their model involves an initial decision stage, an adversarial scenario selection, and a final decision stage. They also added an extra adversarial stage, resulting in min–max-min–max problems and extending the model to general multi-stage scenarios. Of late, Zhang and Gupta (2023) proposed a new uncertainty set by placing constraints on the order statistics of random variables and employing the probability integral transformation for robust linear optimisation models. They utilised quantiles of random variables to represent uncertainties and adapted the assignment problem framework to develop a tractable formulation for the *order statistic uncertainty* set. They also showed that this set generalises the interval, budgeted, and demand uncertainty sets.

Beyond the aforementioned linear and quadratic uncertainty sets, various data-driven approaches for designing uncertainty sets have also been explored (e.g., Bertsimas et al. 2018). Having outlined these robust optimisation concepts, we will now turn to a detailed overview of robust DEA in the next subsection.

1.2 Robust optimisation in DEA

Robust DEA is a conservative approach developed to deal with uncertainties in input and/or output data of DMUs. Similar to robust optimisation models, robust DEA aims to protect the input and output parameters against uncertainty within a user-specified uncertainty set, providing a probability bound on constraint feasibility and leading to a more reliable performance assessment. Recently, robust DEA has rapidly evolved, with many models based on the methodologies of Bertsimas and Sim (2004) and Ben-Tal and Nemirovski (2000) to handle uncertain data in DEA.

Robust DEA in the multiplier form was first introduced by Sadjadi and Omrani (2008), who assumed that output data contained inherent uncertainties and applied the robust optimisation approaches developed by Ben-Tal and Nemirovski (2000) and Bertsimas and Sim (2004) to measure the robust efficiency of the DMUs. Following this, a well-cited work by Shokouhi et al. (2010, 2014) presented a robust optimisation-based DEA method for a multiplier model, based on Bertsimas and Sim (2004), that addressed data uncertainties more effectively than the interval DEA approach and with lower complexity than the fuzzy DEA approach. Since these initial models, both theoretical and practical advancements in robust DEA have emerged.

Arabmaldar et al. (2017) and Toloo and Mensah (2019) developed relaxed robust [multiplier] DEA models using the budgeted uncertainty set under constant returns to scale (CRS) and variable returns to scale (VRS) technologies, respectively, to reduce the computational burden of the model proposed by Sadjadi and Omrani (2008).

Focusing on envelopment forms with budgeted uncertainty sets, Hatami-Marbini and Arabmaldar (2021) extended robust DEA to estimate Farrell's cost efficiency, incorporating endogenous uncertainty in input and/or output data along with exogenous uncertainty in input prices, while Salahi et al. (2021) aimed to find robust common weights under norm-1. Toloo et al. (2022) proposed a robust fractional DEA model using a budgeted uncertainty set and employed linearised models to explore duality relations from both pessimistic and optimistic perspectives on the data. They demonstrated that the primal worst form of the multiplier model is equivalent to the dual best form of the envelopment model.

A generalised robust DEA model was developed by Arabmaldar et al. (2023), using robust optimisation with a budgeted uncertainty set. This model incorporates the directional distance function approach along with predefined direction vectors. Li et al. (2024) extended a robust two-stage multiplier DEA model based on a budgeted uncertainty set to analyse bank efficiency, where the impact of the structure and uncertainty of nonperforming loans on bank performance is considered. Arabmaldar et al. (2024) is among the latest developments, first proposing a novel robust DEA model focused on the multiplier form with variable budgeted uncertainty, which is less conservative than existing models, and then presenting a method for determining probabilistic bounds for constraint violations of uncertain parameters. A detailed discussion of the study's contributions is provided in the following section.

1.3 Research gap and contributions

Despite recent advancements in robust optimisation approaches, robust DEA models remain relatively underdeveloped and require further research, particularly through the application of alternative uncertainty sets, which are the cornerstone of the robust optimisation framework. While robust DEA effectively addresses uncertainty, it often suffers from over-conservatism due to the use of traditional uncertainty sets, such as interval, ellipsoidal, and budgeted sets, which are frequently overlooked in the literature. Although various choices for uncertainty sets exist, this research mainly focuses on comparing the proposed robust DEA models with existing ones that use budgeted uncertainty sets, as introduced by Bertsimas and Sim (2004), due to their prominence in the literature.

While Arabmaldar et al. (2024) focused on the multiplier form of deterministic DEA, the envelopment form proposed in the present study is equally important, as it directly models production frontiers and holds significant practical relevance for efficiency measurement. This study contributes to the literature by developing two robust DEA models in the envelopment form, incorporating both recently developed variable budgeted and order statistic-based uncertainty sets, thereby improving the flexibility and generality of existing approaches as well as assessing the impact of sample size on shaping the robustness of results obtained from DEA. More precisely, the study introduces robust DEA models based on variable budgeted and order statistic uncertainty sets to address the over-conservatism often observed in traditional robust DEA models that use budgeted uncertainty sets. The development of these new models involves three key steps: (i) constructing uncertainty sets for input and output data, (ii) formulating the robust counterpart of the deterministic model to compute robust efficiency scores under varying uncertainty sets, and (iii) specifying probability bounds for constraint violations of uncertain parameters. It is important to point out that these new models do not directly compute the probability of constraint violations. Instead, they rely on the structure of the uncertainty sets to implicitly control the level of protection against data uncertainty. This approach aligns with the principles of robust optimisation, where the focus is on ensuring constraint feasibility under worst-case scenarios within a predefined uncertainty set, rather than explicitly estimating probabilistic outcomes.

The proposed models in this study are evaluated for their theoretical merits, including tractability and flexibility, and are complemented by an empirical analysis using data from the European banking sector. This empirical investigation assesses the performance of banking institutions under various input and output uncertainty scenarios, providing insights into their practical relevance. Moreover, it explores the impact of sample size on the robustness and efficiency scores produced by both existing and newly developed robust DEA models, highlighting the advantages of the proposed approach in improving decision-making under uncertainty.

1.4 Structure

The remainder of the paper is structured as follows: Section 2 provides an overview of the basic and robust DEA models, including specific notations and extensions of robust DEA model properties. Section 3 delineates the mathematical details of two new robust DEA models with variable budgeted uncertainty and order statistic uncertainty sets, complemented by a simplified numerical example to clarify their practical application. Section 4 demonstrates the validity, applicability, and effectiveness of these models, using a real-life dataset from the European banking sector. Finally, Section 5 presents concluding remarks and future research directions.

2 Preliminaries

This section presents a brief overview of two existing DEA models that lay the groundwork for the new developments of robust DEA introduced in this paper. The first subsection reviews the envelopment form of the traditional DEA model under the CRS assumption. The second subsection introduces the existing robust DEA models proposed by Hatami-Marbini and Arabmaldar (2021) and Salahi et al. (2021).

2.1 DEA models

DEA models are non-parametric estimators in frontier analysis used to measure the relative efficiencies of homogeneous DMUs that utilise multiple inputs to produce multiple outputs. Consider n DMUs denoted by $\mathrm{DMU}_j; j \in J = \{1, \dots, n\}$, where each DMU consumes m semi-positive inputs $\boldsymbol{x}_j = (\dots, x_{ij}, \dots); i \in I = \{1, \dots, m\}$ to produce s semi-positive outputs $\boldsymbol{y}_j = (\dots, y_{rj}, \dots); r \in R = \{1, \dots, s\}$. The production possibility set (PPS), or technology T, can be defined as $T = \{(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{y} \text{ can be produced from } \boldsymbol{x}\}$. Following standard DEA literature (Cooper et al. 2007; Kerstens et al. 2022), the construction of this set is based on a set of foundational assumptions imposed on the observed input–output data: (19) inclusion of all observations, (20) input and output monotonicity (free disposability), (21) convexity, and (22) ray unboundedness. These assumptions ensure that the reference technology is both economically interpretable and theoretically consistent with production principles.

Following Farrell's idea of measuring technical efficiency based on the relative distance between a DMU and the efficient frontier, the input-oriented technical efficiency of a given $\mathrm{DMU_o}$ is calculated using the following linear programming problem (Charnes et al. 1978):

$$
\begin{aligned}
& \theta_o^{CCR} = \min_{\lambda_j, \theta_o} \theta_o \\
& \text{s.t.} \\
& \sum_{j \in J} \boldsymbol{x}_j \lambda_j \le \boldsymbol{x}_o \theta_o, \\
& \sum_{j \in J} \boldsymbol{y}_j \lambda_j \ge \boldsymbol{y}_o, \\
& \lambda_j \ge 0, j \in J,
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\lambda} = (\dots, \lambda_j, \dots), j \in J$ is the nonnegative intensity decision variable. While this envelopment DEA model represents a formulation under the assumption of

CRS, including the convexity constraint $\sum_{j \in J} \lambda_j = 1$ converts it to a VRS DEA model. It is evident that since $(\lambda_o = 1, \lambda_j = 0; j \neq o, \theta_o = 1)$ is a feasible solution of model (1), the model is both feasible and bounded ($\theta_o^{CCR*} \in (0,1]$). Furthermore, if $\left(\theta_o^{CCR*}, \lambda_o^*\right)$ is the optimal solution of model (1), then $\lambda_o^* \leq \theta_o^{CCR*}$, and consequently $\lambda_o^* \leq 1$. This confirms that the required condition for developing a feasible robust DEA model is satisfied, namely, the non-negativity of all decision variables corresponding to each uncertain parameter.

Let us introduce model (2), which is equivalent reformulation to the CCR model (1). This model is obtained by explicitly incorporating the reference unit o into the convex combination and rearranging the constraints accordingly. Such a transformation is standard in the DEA literature (see e.g., Hatami-Marbini and Arabmaldar 2021; Toloo et al. 2022), and does not alter the feasible region and the optimal value of θ_o^{CCR}. Therefore, both models (1) and (2) yield the same efficiency score. This equivalence is important because all uncertainties in the constraints are on the right-hand side of the CCR model (1), and this arrangement will be necessary for the following discussions in this paper.

$$
\begin{aligned}
&\theta_o^{CCR} = \min_{\lambda_j, \theta_o} \theta_o \\
&\text{s.t.} \\
&\sum_{j \in J (j \neq o)} \boldsymbol{x}_j \lambda_j + (\lambda_o - \theta_o)\, \boldsymbol{x}_o \leq 0, \\
&\sum_{j \in J (j \neq o)} \boldsymbol{y}_j \lambda_j + (\lambda_o - 1)\, \boldsymbol{y}_o \geq 0, \\
&\lambda_j \geq 0, j \in J.
\end{aligned}
\tag{2}
$$

2.2 Robust DEA models and their extensions

In classic DEA models, input and output data are assumed to be deterministic. However, real-world data often involve imprecision, such as bounded, ordinal, or ratio-bounded data, introducing uncertainty. Various techniques have been developed in the DEA literature to address these uncertainties. This section reviews the robust counterpart of the deterministic DEA model (2), as initially introduced by Hatami-Marbini and Arabmaldar (2021) and Salahi et al. (2021) in the robust DEA literature. This robust DEA model is based on the budgeted uncertainty set developed by Bertsimas and Sim (2004)[1].

Consider the i^{th} input constraint and the r^{th} output constraint in model (2). Let $J_i^x = \{j | \widetilde{x}_{ij} \geq 0\}$ and $J_r^y = \{j | \widetilde{y}_{rj} \geq 0\}$ be the index sets associated with uncertain inputs $\widetilde{x}_{ij}$ and uncertain outputs $\widetilde{y}_{rj}$. It should be noted that $J_i^x, J_r^y \subseteq J$. The cardinalities of these sets are denoted as $|J_i^x|$ and $|J_r^y|$, respectively. The objective is to determine an efficiency measure for the DMUs that not only achieves the highest possible value but also ensures the feasibility of the input and output constraints with a specified probability, which can be expressed

[1] Note that, for simplicity, the input (output) constraint index i (r) is occasionally omitted throughout this paper.

by the chance constraints $\text{Prob}\left(\sum_{j \in J} \lambda_j \widetilde{x}_{ij} \leq \theta_o \widetilde{x}_{io}\right) \geq p_i, \forall i \in I$ and $\text{Prob}\left(\sum_{j \in J} \lambda_j \widetilde{y}_{rj} \geq \widetilde{y}_{ro}\right) \geq p_r, \forall r \in R$, respectively.

Assume that the random variables $\widetilde{x}_{ij}$ and $\widetilde{y}_{rj}$ follow an unknown but symmetric distribution[2, 3]. These variables can vary within the ranges $[x_{ij} - \widehat{x}_{ij}, x_{ij} + \widehat{x}_{ij}]$ and $[y_{rj} - \widehat{y}_{rj}, y_{rj} + \widehat{y}_{rj}]$, where $\widehat{x}_{ij}$ and $\widehat{y}_{rj}$ represent the maximum perturbations of inputs and outputs, respectively. The variables $\widetilde{x}_{ij}$ and $\widetilde{y}_{rj}$ are transformed into Z_{ij}^x and Z_{rj}^y, defined as $Z_{ij}^x = |\widetilde{x}_{ij} - x_{ij}| / \widehat{x}_{ij}$ and $Z_{rj}^y = |\widetilde{y}_{rj} - y_{rj}| / \widehat{y}_{rj}$, which lie within the interval $[0, 1]$. The vectors of these variables are denoted as $\boldsymbol{Z}_i^x = (Z_{i1}^x, \ldots, Z_{i|J_i^x|}^x)$ and $\boldsymbol{Z}_r^y = (Z_{r1}^y, \ldots, Z_{r|J_r^y|}^y)$. Henceforth, whenever random variables are mentioned, we mean the random variables Z_{ij}^x and Z_{rj}^y.

We emphasise that our proposed robust DEA models are firmly anchored in the foundational axioms of DEA, as laid out in the literature (e.g., Arabmaldar et al. 2023). In particular, the robust technology set $T^R = \{(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \,|\, \widetilde{\boldsymbol{y}} \text{ can be produced from } \widetilde{\boldsymbol{x}}\}$ constructed in this study satisfies the following axioms:

($\Lambda 1$) Inclusion of observations: $(\widetilde{x}_j, \widetilde{y}_j) \in T^R$ for all $j = 1, \ldots, n$.

($\Lambda 2$) Monotonicity (Free disposability): If $(\widetilde{x}, \widetilde{y}) \in T^R$ and $\widetilde{x}' \geq \widetilde{x}, \widetilde{y}' \leq \widetilde{y}$, then $(\widetilde{x}', \widetilde{y}') \in T^R$.

($\Lambda 3$) Convexity: If $(\widetilde{x}, \widetilde{y}), (\widetilde{x}', \widetilde{y}') \in T^R$, then for all $\mu \in [0,1]$, $\mu (\widetilde{x}, \widetilde{y}) + (1 - \mu) (\widetilde{x}', \widetilde{y}') \in T^R$.

($\Lambda 4$) Ray unboundedness: If $(\widetilde{x}, \widetilde{y}) \in T^R$, then for all $\mu > 0$, $(\mu \widetilde{x}, \mu \widetilde{y}) \in T^R$.

These axioms shape the structure of the robust technology set and ensure that the incorporation of uncertainty—through various uncertainty sets—is built upon a rigorous theoretical foundation in DEA.

The general robust DEA model, which ensures the feasibility of the input constraint i and the output constraint r for any realisation of $\boldsymbol{Z}_i^x$ and $\boldsymbol{Z}_r^y$ within the uncertainty sets $\mathcal{U}_i$ and $\mathcal{U}_r$, can be formulated as follows:

$$
\begin{aligned}
& \theta_o^{Robust} = \min \theta_o \\
& \text{s.t.} \\
& \sum_{j \in J (j \neq o)} \lambda_j x_{ij} + (\lambda_o - \theta_o) x_{io} + \beta_i (\boldsymbol{\lambda}^*, \mathcal{U}_i) \leq 0, \quad \forall i \in I, \\
& \sum_{j \in J (j \neq o)} \lambda_j y_{rj} + (\lambda_o - 1) y_{ro} - \beta_r (\boldsymbol{\lambda}^*, \mathcal{U}_r) \geq 0, \quad \forall r \in R, \\
& \lambda_j \geq 0, \quad \forall j \in J,
\end{aligned}
\tag{3}
$$

[2] A symmetric distribution is one in which the values are distributed evenly around the central point, such that deviations from the nominal value are equally likely in both positive and negative directions. In robust optimisation, this assumption is a common simplification (e.g., Ben-Tal et al. 2009; Bertsimas et al. 2011), as it enables tractable formulations while still capturing the essential aspects of uncertainty. Although it may not accurately reflect all real-world situations, especially where data exhibit skewness, it provides a practical compromise between model realism and mathematical solvability, particularly when detailed distributional information is unavailable.

[3] Henceforth, any reference to random variables pertains to Z_{ij}^x and Z_{rj}^y.

where $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}_i) = \max_{\boldsymbol{Z}_i^x \in \mathcal{U}_i} \sum_{j \in J^x (j \neq o)} |\lambda_j| \, z_{ij}^x \widehat{x}_{ij} + |\lambda_o - \theta_o| \, z_{io}^x \widehat{x}_o$ and $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}_r) = \max_{\boldsymbol{Z}_r^y \in \mathcal{U}_r} \sum_{j \in J^y (j \neq o)} |\lambda_j| \, z_{rj}^y \widehat{y}_{rj} + |\lambda_o - 1| \, z_{ro}^y \widehat{y}_o$, which are known as protection functions. These functions correspond to the uncertain input i and uncertain output r of DMU_j, respectively, and are defined to protect the input and output constraints against data uncertainty, thereby ensuring the feasibility of the constraints. It should be noted that the protection functions have their own objective functions, constraints, parameters, and decision variables. However, some decision variables from the outer model (3) are treated as constants in the inner problems. Specifically, the optimal objective values of the protection functions become part of the constraints in model (3). In other words, within the protection functions $|\lambda_j| \, (\forall j \in J, j \neq o)$, $|\lambda_o - \theta_o|$, and $|\lambda_o - 1|$, which are decision variables for model (3), are considered constant parameters, while z_{ij}^x and z_{rj}^y are treated as decision variables.

It is worth noting that the uncertainty set remains the cornerstone of the robust optimisation approach. In model (3), $\mathcal{U}_i$ and $\mathcal{U}_r$ can represent various uncertainty sets, commonly used in the robust optimisation literature, including but not limited to interval, budgeted, and ellipsoidal uncertainty sets. It is important to note that within the robust optimisation framework, and without loss of generality, the uncertainty sets in the models are assumed to be *constraint-wise*[4].

Following Bertsimas and Sim (2004), the total perturbations of z_{ij}^x and z_{rj}^y for all inputs and outputs are given by $\sum_{j \in J^x} z_{ij}^x$ and $\sum_{j \in J^y} z_{rj}^y$, respectively. These are constrained by the level of the uncertainty parameters $\boldsymbol{\Gamma}^x = (\Gamma_1^x, \ldots, \Gamma_m^x)$ and $\boldsymbol{\Gamma}^y = (\Gamma_1^y, \ldots, \Gamma_s^y)$, viz. $\sum_{j \in J_i^x} z_{ij}^x \leq \Gamma_i^x$ and $\sum_{j \in J_r^y} z_{rj}^y \leq \Gamma_r^y$, which vary within the interval $[0, n]$. To adjust the level of conservatism in a robust solution, the parameters $\Gamma_i^x \in [0, |J_i^x|]$ and $\Gamma_r^y \in [0, |J_r^y|]$, known as the *robust parameters* or the *budgets of uncertainty*, represent the maximum number of uncertain parameters allowed in the model's constraints. In view of this, the budgeted uncertainty sets can be expressed as follows:

$$\mathcal{U}_i^B(\Gamma_i^x) = \{\widetilde{x}_{ij} | \widetilde{x}_{ij} = x_{ij} + z_{ij}^x \widehat{x}_{ij}, 0 \leq z_{ij}^x \leq 1, \sum_{j \in J_i^x} z_{ij}^x \leq \Gamma_i^x, \forall i \in I, \forall j \in J_i^x\}, \tag{4}$$

$$\mathcal{U}_r^B(\Gamma_r^y) = \{\widetilde{y}_{rj} | \widetilde{y}_{rj} = y_{rj} + z_{rj}^y \widehat{y}_{rj}, 0 \leq z_{rj}^y \leq 1, \sum_{j \in J_r^y} z_{rj}^y \leq \Gamma_r^y, \forall r \in R, \forall j \in J_r^y\}.$$

The uncertainty sets, $\mathcal{U}^B(\Gamma^x)$ and $\mathcal{U}^B(\Gamma^y)$, defined in (4), can first be used within the robust DEA model (3). In other words, we have the protection functions $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}^B(\Gamma^x))$ and $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}^B(\Gamma^y))$ for the input and output constraints, respectively. Then, the linearised robust counterpart of model (3) using these protection

[4] This is because a joint uncertainty set $\mathcal{U}$ across constraints can always be reformulated into a constraint-wise format (see Sect. 1.2.1 in Ben-Tal et al. 2009). As a result, for simplicity, the constraint index $i(r)$ will be omitted, focusing instead on a representative constraint wherever needed.

functions, based on Bertsimas and Sim (2004)'s approach, can be formulated as follows[5]:

$$
\begin{aligned}
&\theta_o^B = \min \theta_o \\
&\text{s.t.} \\
&\sum_{j \in J} \boldsymbol{x}_j \lambda_j + \sum_{j \in J_i^x} \boldsymbol{q}_j^x + \boldsymbol{\Gamma}^x \boldsymbol{p}^x \le \boldsymbol{x}_o \theta_o, \\
&\sum_{j \in J} \boldsymbol{y}_j \lambda_j - \sum_{j \in J_r^y} \boldsymbol{q}_j^y - \boldsymbol{\Gamma}^y \boldsymbol{p}^y \ge \boldsymbol{y}_o, \\
&\boldsymbol{q}_j^x + \boldsymbol{p}^x \ge \widehat{\boldsymbol{x}}_j \lambda_j, \quad \forall j \in J_i^x, j \ne o, \\
&q_o^x + \boldsymbol{p}^x \ge \widehat{\boldsymbol{x}}_o (\theta_o - \lambda_o), \quad o \in J_i^x, \\
&\boldsymbol{q}_j^y + \boldsymbol{p}^y \ge \widehat{\boldsymbol{y}}_j \lambda_j, \quad \forall j \in J_r^y, j \ne o, \\
&q_o^y + \boldsymbol{p}^y \ge \widehat{\boldsymbol{y}}_o (1 - \lambda_o), \quad o \in J_r^y, \\
&\boldsymbol{q}_j^x, \boldsymbol{p}^x \ge 0, \quad \forall j \in J_i^x, \\
&\boldsymbol{q}_j^y, \boldsymbol{p}^y \ge 0, \quad \forall j \in J_r^y, \\
&\lambda_j \ge 0, \quad \forall j \in J,
\end{aligned} \tag{5}
$$

where $\boldsymbol{p}^x = (p_1^x, \ldots, p_m^x), \boldsymbol{p}^y = (p_1^y, \ldots, p_s^y)$, $\boldsymbol{q}_j^x = (q_{1j}^x, .., q_{mj}^x)$, and $\boldsymbol{q}_j^y = (q_{1j}^y, .., q_{sj}^y)$ are variables used to measure the robustness of model (5) when the level of uncertainty budgeting is varied by an infinitesimally small amount. The terms $\sum_{j \in J_i^x} \boldsymbol{q}_j^x + \boldsymbol{\Gamma}^x \boldsymbol{p}^x$ and $-\sum_{j \in J_r^y} \boldsymbol{q}_j^y - \boldsymbol{\Gamma}^y \boldsymbol{p}^y$ demonstrate the worst-case deviations of the uncertain inputs and outputs from their nominal values for the budgeted uncertainty and are included in the robust model (5) to immunise the model against the violation of the input and output constraints, respectively. Furthermore, the pre-defined robust parameters $\boldsymbol{\Gamma}^x$ and $\boldsymbol{\Gamma}^y$ indicate the maximum number of uncertain inputs and outputs, respectively, that are protected against perturbations. As proved by Hatami-Marbini and Arabmaldar (2021), the optimal objective function value of model (5), θ_o^{B*}, is greater than or equal to that of model (1) or (2), $\theta_o^{\text{CCR}*}$.

The following proposition provides an equivalent formulation of $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}^x(\boldsymbol{\Gamma}^x))$ and $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}^y(\boldsymbol{\Gamma}^y))$, which will be essential for developing new robust DEA models in this paper.

Proposition 1 *The protection functions* $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}^x(\boldsymbol{\Gamma}^x))$ *and* $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}^y(\boldsymbol{\Gamma}^y))$ *can be equivalently expressed as follows:*

$$
\beta_i(\lambda^*, \mathcal{U}^x(\boldsymbol{\Gamma}^x)) = \max_{\{S_i^x \cup \{\alpha_i^x\} | S_i^x \subseteq J_i^x, |S_i^x| = [\Gamma_i^x], \alpha_i^x \in J_i^x \setminus S_i^x\}} \left\{ \begin{array}{c} \sum_{j \in S_i^x (j \ne o)} \lambda_j \widehat{x}_{ij} + (\Gamma_i^x - [\Gamma_i^x]) \lambda_{\alpha_i^x} \widehat{x}_{i\alpha_i^x} + \\ |\lambda_o - \theta_o| \widehat{x}_{io} + (\Gamma_i^x - [\Gamma_i^x]) |\lambda_o - \theta_o| \widehat{x}_{io} \end{array} \right\}, \tag{6}
$$

$$
\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}^y(\boldsymbol{\Gamma}^y)) = \max_{\{S_r^y \cup \{\alpha_r^y\} | S_r^y \subseteq J_r^y, |S_r^y| = [\Gamma_r^y], \alpha_r^y \in J_r^y \setminus S_r^y\}} \left\{ \begin{array}{c} \sum_{j \in S_r^y (j \ne o)} \lambda_j \widehat{y}_{rj} + (\Gamma_r^y - [\Gamma_r^y]) \lambda_{\alpha_r^y} \widehat{y}_{r\alpha_r^y} + \\ |\lambda_o - 1| \widehat{y}_{ro} + (\Gamma_r^y - [\Gamma_r^y]) |\lambda_o - 1| \widehat{y}_{ro} \end{array} \right\}
$$

[5] For more details, see Hatami-Marbini and Arabmaldar (2021) and Salahi et al. (2021).

Proof See Appendix.

Bertsimas and Sim (2004) discussed that when only a subset of the uncertain parameters is allowed to change, a bound for the robust counterpart model is essential to ascertain that the robust solution remains feasible with high probability. They demonstrated that the probability of constraint violation, $\text{prob}(\sum_j \widetilde{a}_{ij}\boldsymbol{x}_j^* > b_i)$ is bounded above by $P = \exp(-\frac{\Gamma^2}{2|J_i|})$ where $|J_i|$ denotes the cardinality of the set of uncertain parameters associated with the i^{th} constraint. Here, we adopt the bounds proposed by Bertsimas and Sim (2004) for the existing robust DEA model (5) under the budgeted uncertainty set. Since, in the i^{th} and r^{th} constraints of model (5), a fixed number $\boldsymbol{\Gamma}^x = \boldsymbol{\Gamma}^y = \boldsymbol{\Gamma}$ of input and output data are allowed to deviate from their deterministic values, the constraint feasibility of this model is guaranteed with the probability bound for both input and output constraint sets as follows[6]

$$p^B = \text{Prob}(\sum_{j\in J}\lambda_j\tilde{x}_{ij} \le \theta_o\tilde{x}_{io}) \text{ or Prob } (\sum_{j\in J}\lambda_j\tilde{y}_{rj} \ge \tilde{y}_{ro}) \ge \exp(-\frac{\Gamma^2}{2\,|J_i^x|\,|J_r^y|}). \quad (7)$$

3 Proposed robust DEA models

The motivation for this study arises from the observation that, despite the widespread adoption of Bertsimas and Sim (2004)'s approach among scholars, there is a hidden over-conservatism in robust optimisation problems, particularly when the uncertainty sets are polyhedral (Thiele 2010; Liu et al. 2016). More precisely, in some cases, the decision variables become overly conservative in response to the uncertain parameters, leading to solutions that are more conservative than originally intended by the decision-maker. Therefore, greater care is needed when interpreting the budget of uncertainty as the maximum number of parameters that can vary.

The first and foremost objective of robust optimisation and, consequently, robust DEA is to balance the trade-off between the optimality of the objective function value and the probability of constraint feasibility, namely, the PoR. This objective can be more effectively achieved by incorporating more flexible uncertainty sets into deterministic models. To extend the existing robust DEA models in alignment with robust optimisation approaches, we propose two new robust DEA models in this section. These models utilise two different uncertainty sets, which can be viewed as generalisations of the budgeted uncertainty set. Section 3.1 presents a novel robust DEA model that incorporates the variable budgeted uncertainty set, as proposed by Poss (2013, 2014). Unlike the traditional budgeted uncertainty set introduced by Bertsimas and Sim (2004), the variable budgeted uncertainty set does not constrain the amount of uncertainty to a pre-specified number. Instead, it employs a non-negative function defined within the feasibility region of decision variables, providing greater flexibility in handling data uncertainty. Furthermore, Section 3.2 proposes a new

[6]Note that $\boldsymbol{x}_j^*$ represents a vector of decision variables at optimality, and $\widetilde{a}_{ij} = [\widetilde{a}_{1j}, \ldots, \widetilde{a}_{nj}]$ is the technological coefficients

robust DEA model employing the recently developed order statistic uncertainty set by Zhang and Gupta (2023). The order statistic uncertainty set has greater geometric flexibility compared to the budgeted uncertainty set and is considered to include interval, budgeted, and demand uncertainty sets as special cases within the robust optimisation model. To complement the theoretical development, Section 3.3 presents a simple illustrative numerical example that demonstrates the proposed robust DEA models under different uncertainty sets, highlighting their practical implications and the trade-off between robustness and performance.

3.1 Robust DEA with variable budgeted uncertainty

Assume that the true values of uncertain input and output data are introduced as $\widetilde{x}_{ij} = x_{ij} + z^x_{ij}\widehat{x}_{ij}\,(\forall i \in I, \forall j \in J^x_i)$, and $\widetilde{y}_{rj} = y_{rj} + z^y_{rj}\widehat{y}_{rj}\,(\forall r \in R, \forall j \in J^y_r)$, respectively. It is also assumed that $\sum_{j\in J^x_i} z^x_{ij} \le \gamma^x(\boldsymbol{\lambda})$, and $\sum_{j\in J^y_r} z^y_{rj} \le \gamma^y(\boldsymbol{\lambda})$, where $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$ are given non-negative functions defined on the decision variable $\boldsymbol{\lambda}$, which is in vector form, that limit the amount of uncertainty. By adopting the variable budgeted uncertainty proposed by Poss (2013, 2014) as a generalisation of the budgeted uncertainty defined by Bertsimas and Sim (2004), the following variable budgeted uncertainty sets can be obtained for the uncertain input and output data:

$$\mathcal{U}^{VB}_i(\boldsymbol{\lambda}) = \{(\widetilde{\boldsymbol{x}}_{\boldsymbol{i}})\,|\widetilde{x}_{ij} = x_{ij} + z^x_{ij}\widehat{x}_{ij}, 0 \le z^x_{ij} \le 1, \sum_{j\in J^x_i} z^x_{ij} \le \gamma^x(\boldsymbol{\lambda}), \forall i \in I, \forall j \in J^x_i\}, \tag{8}$$

$$\mathcal{U}^{VB}_r(\boldsymbol{\lambda}) = \{(\widetilde{\boldsymbol{y}}_{\boldsymbol{r}})\,|\widetilde{y}_{rj} = y_{rj} + z^y_{rj}\widehat{y}_{rj}, 0 \le z^y_{rj} \le 1, \sum_{j\in J^y_r} z^y_{rj} \le \gamma^y(\boldsymbol{\lambda}), \forall r \in R, \forall j \in J^y_r\}.$$

$\mathcal{U}^{VB}_i(\boldsymbol{\lambda})$ and $\mathcal{U}^{VB}_r(\boldsymbol{\lambda})$ are multi-functions of the decision variable $\boldsymbol{\lambda}$ and act as alternatives to the uncertainty sets $\mathcal{U}^B_i(\boldsymbol{\Gamma}^x)$ and $\mathcal{U}^B_r(\boldsymbol{\Gamma}^y)$, which are bounded by $\boldsymbol{\Gamma}^x$ and $\boldsymbol{\Gamma}^y$, respectively. In other words, given the decision variable $\boldsymbol{\lambda}$, the uncertainty sets $\mathcal{U}^{VB}_i(\boldsymbol{\lambda})$ and $\mathcal{U}^{VB}_r(\boldsymbol{\lambda})$ include all feasible values for the uncertain input and output parameters, respectively. Furthermore, if $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$ are set constantly equal to $\boldsymbol{\Gamma}^x$ and $\boldsymbol{\Gamma}^y$, then it is evident that $\mathcal{U}^{VB}_i(\boldsymbol{\lambda})$ and $\mathcal{U}^{VB}_r(\boldsymbol{\lambda})$ coincide with $\mathcal{U}^B_i(\boldsymbol{\Gamma}^x)$ and $\mathcal{U}^B_r(\boldsymbol{\Gamma}^y)$, respectively, for any $\boldsymbol{\lambda}$. In general, utilising $\mathcal{U}^{VB}_i(\boldsymbol{\lambda})$ and $\mathcal{U}^{VB}_r(\boldsymbol{\lambda})$ helps to avoid the issue of over-conservatism that can arise when decision-variable vectors $\boldsymbol{\lambda}$ contain few components in each constraint. This approach suggests a new framework that is less conservative compared to $\mathcal{U}^B_i(\boldsymbol{\Gamma}^x)$ and $\mathcal{U}^B_r(\boldsymbol{\Gamma}^y)$.

As discussed by Poss (2013, 2014), it is necessary that the functions $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$, which are involved in defining $\mathcal{U}^{\gamma^x}(\boldsymbol{\lambda})$ and $\mathcal{U}^{\gamma^y}(\boldsymbol{\lambda})$, to be affine functions of $\boldsymbol{\lambda}$ so as to satisfy the probability bounds proposed by Bertsimas and Sim (2004). Following Poss (2013, 2014), we consider the case where $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$ are affine functions of λ, specified as $\gamma^x(\boldsymbol{\lambda}) = \gamma^x_0 + \sum_{j\in J^x_i} \gamma^x_j \lambda_j$ and $\gamma^y(\boldsymbol{\lambda}) = \gamma^y_0 + \sum_{j\in J^y_r} \gamma^y_j \lambda_j$, respectively. Consequently, the robust counterpart of model (1), based on the uncertainty sets $\mathcal{U}_i^{VB}(\boldsymbol{\lambda})$ and $\mathcal{U}_r^{VB}(\boldsymbol{\lambda})$ defined in (8), is as follows:

$$
\begin{aligned}
&\theta_o^{VB} = \min\ \theta_0 \\
&\text{s.t.} \\
&\quad \sum_{j \in J(j \neq o)} \lambda_j x_{ij} + (\lambda_o - \theta_0)\, x_{io} + \beta_i\left(\boldsymbol{\lambda}^*, \mathcal{U}_i^{VB}(\boldsymbol{\lambda})\right) \leq 0, \quad \forall i \in I, \\
&- \sum_{j \in J(j \neq o)} \lambda_j y_{rj} - (\lambda_o - 1)\, y_{ro} + \beta_r\left(\boldsymbol{\lambda}^*, \mathcal{U}_r^{VB}(\boldsymbol{\lambda})\right) \leq 0, \quad \forall r \in R, \\
&\lambda_j \geq 0, \quad \forall j \in J.
\end{aligned}
\tag{9}
$$

The following theorem shows how to solve model (9) as a mixed-integer linear programming when $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$ are appropriately selected.

Theorem 1 *The robust counterpart of model (2) based on the variable budgeted uncertainty set is equivalent to the following robust DEA model:*

$$
\begin{aligned}
&\theta_o^{VB} = \min \theta_0 \\
&s.t. \\
&\sum_{j \in J} \lambda_j x_{ij} + p_i^{x'} \gamma_0^x + \sum_{j \in J_i^x} \gamma_j^x w_{ij}^x + \sum_{j \in J_i^x} q_{ij}^{x'} \leq \theta_0 x_{io}, \quad \forall i \in I, \\
&\sum_{j \in J} \lambda_j y_{rj} - p_r^{y'} \gamma_0^y - \sum_{j \in J_r^y} \gamma_j^y w_{rj}^y - \sum_{j \in J_r^y} q_{rj}^{y'} \geq y_{ro}, \quad \forall r \in R, \\
&p_i^{x'} + q_{ij}^{x'} \geq \lambda_j \hat{x}_{ij}, \quad \forall i \in I, \forall j \in J_i^x, j \neq o, \\
&p_i^{x'} + q_{io}^{x'} \geq (\theta_0 - \lambda_o)\, \hat{x}_{io}, \quad \forall i \in I, o \in J_i^x, \\
&w_{ij}^x - p_i^{x'} \geq -\max_j (\hat{x}_{ij})\,(1 - \lambda_j), \forall i \in I, \forall j \in J, \\
&p_r^{y'} + q_{rj}^{y'} \geq \lambda_j \hat{y}_{rj}, \quad \forall r \in R, \forall j \in J_r^y, j \neq o, \\
&p_r^{y'} + q_{ro}^{y'} \geq (1 - \lambda_o)\, \hat{y}_{ro}, \quad \forall r \in R, o \in J_r^y, \\
&w_{rj}^y - p_r^{y'} \geq -\max_j (\hat{y}_{rj})\,(1 - \lambda_j), \quad \forall r \in R, j \in J, \\
&p_i^{x'}, q_{ij}^{x'}, p_r^{y'}, q_{rj}^{y'}, w_{ij}^x, w_{rj}^y \geq 0, \quad \forall i \in I, \forall r \in R, \forall j \in J_i^x, \forall j \in J_r^y, \\
&\lambda_j \leq h_j^x, \quad \forall j \in J, \\
&\lambda_j \leq h_j^y, \quad \forall j \in J, \\
&h_j^y, h_j^x \in \{0,1\}^n, \quad \forall j \in J, \\
&\lambda_j \geq 0, \quad \forall j \in J.
\end{aligned}
\tag{10}
$$

Proof See Appendix.

To demonstrate that the proposed robust DEA model (10) adheres to the properties of traditional DEA models, we present the following theorem.

Theorem 2 *(i) Model (10) is always feasible, and (ii)* $0 < \theta_o^{VB*} \leq 1$.

Proof See Appendix.

The following theorem, which is important for gaining insights into the proposed approach, compares the optimal objective values between the traditional robust DEA model (5) and the proposed robust DEA model (10).

Theorem 3 *The optimal objective function value of model (10) is greater than or equal to that of model (5), i.e.,* $\theta_o^{VB*} \geq \theta_o^{B*}$.

Proof See Appendix.

Finally, to discuss the probability bound for the constraints in the proposed robust DEA model (10), we apply the probabilistic bounds proposed by Bertsimas and Sim (2004) alongside the robust DEA model (5) to derive the same bound. Let $\boldsymbol{\lambda}^*$ be the intensity vectors that satisfy the robust input and output constraints in model (10) for $\mathcal{U}_i^{VB}(\boldsymbol{\lambda})$ and $\mathcal{U}_r^{VB}(\boldsymbol{\lambda})$. It is trivial that if $\|\boldsymbol{\lambda}^*\| \leq \boldsymbol{\Gamma}$, then $\text{Prob}(\sum_{j\in J} \lambda_j \widetilde{x}_{ij} > \theta_o \widetilde{x}_{io}) = 0$. In addition, if $\|\boldsymbol{\lambda}^*\| > \boldsymbol{\Gamma}$ such that $\boldsymbol{\lambda}^*$ satisfies the robust constraint $\sum_{j\in J} \lambda_j \widetilde{x}_{ij} \leq \theta_o \widetilde{x}_{io}$, for $\widetilde{x}_{ij} \in \mathcal{U}_i^{VB}(\boldsymbol{\lambda})$, then according to Proposition 2 and Theorem 2 in Bertsimas and Sim (2004), we can easily verify the following probabilistic bound:

$$P^{VB} = \text{Prob}(\sum_{j\in J} \lambda_j \tilde{x}_{ij} > \theta_o \tilde{x}_{io}) \text{ or } \text{Prob}(\sum_{j\in J} \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{ro}) \geq \exp(-\frac{\boldsymbol{\Gamma}^2}{2\|\boldsymbol{\lambda}\|^*}) \quad (11)$$

The above inequality shows that $\mathcal{U}_i^{VB}(\boldsymbol{\lambda})$ and $\mathcal{U}_r^{VB}(\boldsymbol{\lambda})$ allow decision-makers to ensure the same or an even higher level of protection for the feasibility of constraints in robust counterpart models. This means that by utilising these uncertainty sets, decision-makers can achieve a more reliable and resilient solution, effectively immunising against variations and uncertainties in the data. As a result, the proposed robust DEA model not only maintains feasibility under adverse conditions but may also provide better performance guarantees compared to existing approaches, thereby enhancing the robustness of the decision-making process.

3.2 Robust DEA with the order statistic uncertainty set

We here build on the order statistic uncertainty set proposed by Zhang and Gupta (2023) to develop a new robust DEA model aimed at reducing the cost associated with uncertainty. The relationship between this model, the proposed robust DEA model (10), and the robust DEA model (5) with the budgeted uncertainty set is explored.

For simplicity, we omit the constraint indices for the input and output constraints and focus on an arbitrary constraint.

Suppose the random variables z_j^x and z_j^y are continuous and independently distributed in the range [0,1], each following an arbitrary continuous distribution with an unknown cumulative distribution function F_j^x and F_j^y, respectively. Let $U_j^x = F_j^x(z_j^x), \forall j \in J^x$ and $U_j^y = F_j^y(z_j^y), \forall j \in J^y$ be the random variables, where each U_j^x and U_j^y is uniformly distributed over [0,1]. Here, $U_{(1)}, \ldots, U_{(|J^x|)}$ denotes the order statistics of U_j^x (and similarly $U_{(1)}, \ldots, U_{(|J^y|)}$ for U_j^y), which represents the rearranged sequence of $U_j^x(U_j^y)$ with $U_{(k)}$ being the kth smallest value. Unlike the original random variables U_j^x and U_j^y, the order statistic $U_{(k)}$ has a *Beta* distribution with parameters $(k, |J^x| + 1 - k)$. Let $I_t\,(k, |J^x| + 1 - k)$ denote the cumulative distribution function for $Beta\,(k, |J^x| + 1 - k)$ distribution, and let Q_k^t be the quantile function defined as $Q_k^t = \inf\{\tau : I_\tau\,(k, |J^x| + 1 - k) = t\}$. The order statistic uncertainty set can thus be defined as follows (Zhang and Gupta 2023, p. 1026):

$$\mathcal{U}_i^{OS}(\boldsymbol{\varepsilon}) = \left\{\eta^x \middle|\ F_j^x\left(z_j^x\right) = U_j^x, \forall j \in J^x, \text{and } U_{(k)}^x \leq Q_k^{(1-\varepsilon_k)}, \forall k \in J\ \right\}, \tag{12}$$

$$\mathcal{U}_r^{OS}(\boldsymbol{\varepsilon}) = \left\{\eta^y \middle|\ F_j^y\left(z_j^y\right) = U_j^y, \forall j \in J^y, \text{and } U_{(k)}^y \leq Q_k^{(1-\varepsilon_k)}, \forall k \in J\ \right\},$$

where $Q_k^{(1-\varepsilon_k)}$ is the upper limit for $U_{(k)}^x(U_{(k)}^y)$ such that $\mathrm{Prob}(U_{(k)}^x \leq Q_k^{(1-\varepsilon_k)}) = 1 - \varepsilon_k$ and $\boldsymbol{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{|J|}\right)$, where $\varepsilon_j \in [0, 1]$.

As discussed by Zhang and Gupta (2023), the order statistic uncertainty set (12) is intractable for obtaining robust counterpart models.[7] To deal with this difficulty in reformulating protection functions associated with the order statistic uncertainty set, Zhang and Gupta (2023) proposed an assignment formulation that provides a tractable solution for these problems. In this study, we adopt the same method to develop a suitable formulation associate with the order statistic uncertainty sets in the DEA context. We therefore propose the following proposition to provide a tractable formulation for input and output protection functions, i.e., $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}_i^{OS}(\boldsymbol{\varepsilon}))$ and $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}_r^{OS}(\boldsymbol{\varepsilon}))$. Let $\rho_{jk}^x(\rho_{jk}^y)$ be the quantile of order $Q_k^{(1-\varepsilon_k)}$ for $z_j^x(z_j^y)$, i.e., $\rho_{jk}^x = \inf\{\boldsymbol{\lambda} : F_j^x(\boldsymbol{\lambda}) \geq Q_k^{(1-\varepsilon_k)}\}, \forall j, k \in J_x (\rho_{jk}^y = \inf\{\boldsymbol{\lambda} : F_j^y(\boldsymbol{\lambda}) \geq Q_k^{(1-\varepsilon_k)}\}, \forall j, k \in J_y)$.

Proposition 2 *For a fixed* λ*, the optimal objective value for* $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}_i^{OS}(\boldsymbol{\varepsilon}))$ *and* $\beta_r(\boldsymbol{\lambda}^*, \mathcal{U}_r^{OS}(\boldsymbol{\varepsilon}))$*, corresponding to input and output constraints, are equal to the optimal objective values for the following linear optimisation problems, respectively:*

[7] The reasons for the intractability are as follows: (i) the uncertainty set $\mathcal{U}^{OS}(\varepsilon)$ is defined using constraints on the cumulative distribution functions of variable $z_j^x(z_j^y)$, rather than being directly based on the random variable $z_j^x(z_j^y)$; (ii) there are $|J^x|\,(|J^y|)!$ permutations of $F_j^x\,(\lambda)\,(F_j^y\,(\lambda))$ for all possible outcomes of $U_{(k)}^x(U_{(k)}^x)$, which makes reformulating $\beta_i\left(\boldsymbol{\lambda}^*, \mathcal{U}_i^{OS}(\varepsilon)\right)$ challenging; and (iii) the non-convexity of the order statistic uncertainty set $\mathcal{U}^{OS}(\varepsilon)$ (for more details see (Zhang and Gupta 2023).

$$
\begin{aligned}
&\max_{\eta} \sum_{j\in J(j\neq o)} \widehat{x}_{ij} \left|\lambda_j\right| .(\sum_{k\in J} \rho^x_{jk}\eta^x_{jk}) + \widehat{x}_{io} \left|\lambda_o - \theta_o\right| . \left(\sum_{k\in J} \rho^x_{ok}\eta^x_{ok}\right)\\
&s.t.\\
&\sum_k \eta^x_{jk} = 1, \quad \forall j \in J^x,\\
&\sum_j \eta^x_{jk} = 1, \quad \forall k \in J^x,\\
&\eta^x_{jk} \geq 0, \quad \forall j,k \in J^x.
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
&\max_{\eta} \sum_{j\in J(j\neq o)} \widehat{y}_{ij} \left|\lambda_j\right| .(\sum_{k\in J} \rho^y_{jk}\eta^y_{jk}) + \widehat{y}_{io} \left|\lambda_o - 1\right| . \left(\sum_{k\in J} \rho^y_{ok}\eta^y_{ok}\right)\\
&s.t.\\
&\sum_k \eta^y_{jk} = 1, \quad \forall j \in J^y,\\
&\sum_j \eta^y_{jk} = 1, \quad \forall k \in J^y,\\
&\eta^y_{jk} \geq 0, \quad \forall j,k \in J^y.
\end{aligned}
\tag{14}
$$

Proof The proof is omitted, as it follows a similar method to that used in Zhang and Gupta (2023).

Hereafter, we denote the protection functions defined in (13) and (14) as $\beta_i(\boldsymbol{\lambda}^*,\mathcal{U}_i^{OS}(\boldsymbol{\rho^x}))$ and $\beta_r(\boldsymbol{\lambda}^*,\mathcal{U}_r^{OS}(\boldsymbol{\rho^y}))$, respectively. Now, by incorporating the defined uncertainty set $\mathcal{U}^{OS}(\boldsymbol{\varepsilon})$ into the DEA model (3), we obtain the following robust counterpart model:

$$
\begin{aligned}
&\theta_o^{OS} = \min \theta_o\\
&\text{s.t.}\\
&\sum_{j\in J(j\neq o)} \lambda_j x_{ij} + (\lambda_o - \theta_o)\, x_{io} + \beta_i\left(\boldsymbol{\lambda}^*,\mathcal{U}_i^{OS}(\boldsymbol{\rho^x})\right) \leq 0, \quad \forall i \in I,\\
&\sum_{j\in J(j\neq o)} \lambda_j y_{rj} + (\lambda_o - 1)\, y_{ro} - \beta_r\left(\boldsymbol{\lambda}^*,\mathcal{U}_r^{OS}(\boldsymbol{\rho^y})\right) \geq 0, \quad \forall r \in R,\\
&\lambda_j \geq 0, \quad \forall j \in J.
\end{aligned}
\tag{15}
$$

The following theorem presents the robust formulation of the DEA model (2) incorporating uncertain input and output data defined by the uncertainty sets $\mathcal{U}_i^{OS}(\boldsymbol{\varepsilon})$ and $\mathcal{U}_r^{OS}(\boldsymbol{\varepsilon})$.

Theorem 4 *Model (15) is equivalent to the following linear programming problem:*

$$
\begin{aligned}
&\theta_o^{OS} = \min \theta_o \\
&s.t. \\
&\sum_{j\in J} x_{ij}\lambda_j + \sum_{j\in J_i^x} (\psi_{ij}^x + \Phi_{ij}^x) \leq x_{io}\theta_o, \quad \forall i \in I, \\
&\sum_{j\in J} y_{rj}\lambda_j - \sum_{j\in J_r^y} (\psi_{rj}^y + \Phi_{rj}^y) \geq y_{ro}, \qquad \forall r \in R, \\
&\psi_{ij}^x + \Phi_{ik}^x \geq \hat{x}_{ij}\ \lambda_j\ \rho_{ijk}, \quad \forall j,k \in J_i^x, j \neq o, \forall i \in I, \\
&\psi_{io}^x + \Phi_{io}^x \geq \hat{x}_{io}\,(\theta_o - \lambda_o)\,\rho_{io}, \quad o \in J_i^x, \forall i \in I, \\
&\psi_{rj}^y + \Phi_{rk}^y \geq \hat{y}_{rj}\ \lambda_j\ \rho_{rjk}, \quad \forall j,k \in J_r^y, j \neq o, \forall r \in R, \\
&\psi_{ro}^y + \Phi_{ro}^y \geq \hat{y}_{ro}\,(1 - \lambda_o)\,\rho_{ro}, \quad o \in J_r^y, \forall r \in R, \\
&\psi_{ij}^x, \Phi_{ij}^x,\ free\ in\ sign \quad \forall i \in I;\ \forall j \in J_i^x, \\
&\psi_{rj}^y, \Phi_{rj}^y,\ free\ in\ sign \quad \forall r \in R;\ \forall j \in J_i^y, \\
&\lambda_j \geq 0, \quad \forall j \in J.
\end{aligned} \tag{16}
$$

Proof See Appendix.

The following theorem demonstrates that the proposed robust DEA model (16) preserves the fundamental properties of traditional DEA models.

Theorem 5 *(i) Model (16) is always feasible, and (ii)* $0 < \theta_o^{OS*} \leq 1$.

Proof See Appendix.

An interesting result from the above theorem is that the proposed robust DEA model (16) not only maintains the core properties of traditional DEA models but also improves their applicability by effectively addressing the uncertainty inherent in both input and output data.

The following theorem demonstrates that the existing robust DEA model (5), which utilises the budgeted uncertainty set with the robust parameter Γ, is equivalent to the proposed robust DEA model (16), which employs the order statistic uncertainty set, given that the values of ρ_{jk} are appropriately selected.

Theorem 6 *The existing robust DEA model (5) is equivalent to the proposed robust DEA model (16) when the values of* ρ_{jk} *are chosen as follows:*

$$
\rho_{jk} = \begin{cases} 0, & k \in [1, |J^x| - \lfloor \boldsymbol{\Gamma}^x \rfloor - 1], & \forall j \in J^x, \\ \boldsymbol{\Gamma}^x - \lfloor \boldsymbol{\Gamma}^x \rfloor, & k = |J^x| - \lfloor \boldsymbol{\Gamma}^x \rfloor, & \forall j \in J^x, \\ 1, & k \in [|J^x| - \lfloor \boldsymbol{\Gamma}^x \rfloor + 1, |J^x|], & \forall j \in J^x. \end{cases}
$$

Proof See Appendix.

Let us now follow the probabilistic bound suggested in Zhang and Gupta (2023) and propose the following probability bound for constraint violation in the robust DEA model (16):

$$p^{OS} = \text{Prob}(\sum_{j\in J} \lambda_j \tilde{x}_{ij} \leq \theta_o \tilde{x}_{io}) \geq \frac{1}{2} + \frac{1}{2}.\,|J^x|!\,\det[\Delta] \text{ or}$$
$$\text{Prob}(\sum_{j\in J} \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{ro}) \geq \frac{1}{2} + \frac{1}{2}.\,|J^y|!\,\det[\Delta]\,, \tag{17}$$

where Δ is the $|J| \times |J|$ matrix with $(i,j)^{\text{th}}$ element defined as:

$$\Delta_{ij} = \begin{cases} \left(Q_i^{(1-\varepsilon_i)}\right)^{j-i+1} / (j-i+1)! & j-i+1 \geq 0, \\ 0 & j-i+1 < 0. \end{cases} \tag{18}$$

As observed from the above and in line with the definition of uncertainty sets (12), the probabilistic guarantee for the order statistic uncertainty set can be identified by the upper limit of the cumulative distribution functions of random variables $Q_i^{(1-\varepsilon_i)}$. Note that to achieve a high probabilistic guarantee, one must use larger $Q_i^{(1-\varepsilon_i)}$ values, which can be attained by selecting a smaller ε_i.

In this section, we develop two new robust DEA models based variable budgeted uncertainty and order statistic uncertainty sets, both of which represent different generalisations of the traditional budgeted uncertainty set. To conclude this section, we provide a discussion on the specific situations in which it is most suitable to use one of the uncertainties sets and their associated robust DEA models. The budgeted uncertainty set is often most favourable when a fixed and predictable approach to managing overall risk is required (Bertsimas and Brown 2009). Its key advantages include simplicity and computational efficiency, making it particularly well-suited for scenarios where uniform management of total uncertainty across all parameters is necessary. This uncertainty set is ideal in environments where the uncertainty is relatively stable and can be anticipated, allowing for a straightforward application.

On the other hand, the variable budgeted uncertainty set is more appropriate for dynamic environments where uncertainties are not static and may fluctuate over time (Poss 2013, 2014). This uncertainty set allows for different levels of conservatism across scenarios, providing flexibility that the traditional budgeted uncertainty set cannot provide. As a result, it is more suited for situations where the decision-making context demands a higher degree of adaptability and a tailored approach to risk management (Poss 2013, 2014). Therefore, the variable budgeted uncertainty set aligns with the DEA framework by offering a flexible and practical method to deal with uncertainty in input and output data, while maintaining the interpretability and applicability of the models. Furthermore, the order statistic uncertainty set is particularly valuable in situations where the decision-making process is heavily influenced by extreme values[8] among the uncertainties. This uncertainty set is designed to prioritise and manage the extreme risks, making it the preferred choice in contexts that require robust management of tail risks (Bertsimas and Brown 2009). Thereby, the order statistic uncertainty set is well-suited to the DEA context as it captures prioritised uncertainty levels by considering specific quantiles of data distributions. This uncertainty enables robust efficiency evaluations by effectively addressing worst-case or targeted variations in inputs and outputs.

In the end, the selection of the appropriate uncertainty set and associated robust DEA model depends on the specific characteristics of the uncertainties involved and the decision-maker's priorities. Factors such as the need for predictability, flexibility, computational efficiency, and the management of extreme risks should guide this choice.

3.3 Illustrative numerical example

To further clarify the practical application of the proposed robust DEA models, we now present a simple numerical example. This example provides an intermediate step between the theoretical development and the comprehensive real-world case study, enabling readers to better understand how the models function on a small, controlled dataset and how uncertainty affects efficiency assessment. Consider five DMUs, each using two inputs to produce a single output. The nominal input–output data are presented in Table 1.

To demonstrate the effect of uncertainty, we assume that each input and output is subject to uncertainty at 5% of its nominal value. We then evaluate the performance of each DMU using the deterministic DEA model, along with the three robust DEA models proposed in this study under the budgeted, variable budgeted, and order statistic uncertainty sets.

Fig. 1 shows the efficiency scores obtained under these three different model settings. As observed, incorporating robustness generally leads to higher efficiency scores, reflecting the added caution imposed by accounting for uncertainty. The degree of the efficiency change depends on both the DMU and the uncertainty set employed, which demonstrates the varying degrees of robustness and conservatism introduced by each set. For instance, DMU5 shows a noticeable shift in efficiency across all robust models, with the largest impact under the variable budgeted set. In contrast, DMU2 and DMU3 retain full efficiency (score of 1) even when uncertainty is introduced, suggesting that these units are robustly efficient.

Table 1 Data for the numerical example

DMUs	Input 1	Input 2	Output 1
DMU1	2	2	2
DMU2	1	4	4
DMU3	4	1	6
DMU4	3	2	1
DMU5	4	6	8

[8] Extreme values refer to the most extreme possible outcomes that the uncertain parameters (random variables z_j^x and z_j^y) can assume within a specified uncertainty set.

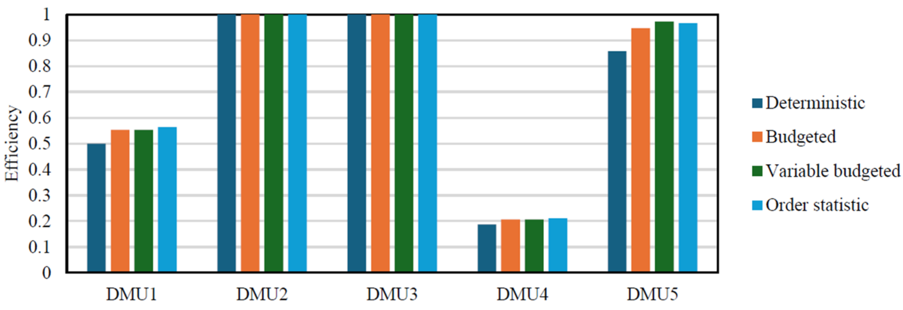

Fig. 1 Deterministic and robust efficiency scores for different uncertainty sets

Table 2 PoR (%) under different uncertainty sets

DMU	Budgeted	Variable budgeted	Order statistic
DMU1	9.52	9.52	11.32
DMU2	0.00	0.00	0.00
DMU3	0.00	0.00	0.00
DMU4	9.51	9.51	11.31
DMU5	9.53	11.91	11.32

To quantitatively assess the impact of uncertainty on efficiency scores, Table 2 reports the PoR for each DMU under the three uncertainty sets: budgeted, variable budgeted, and order statistic. The PoR is calculated as the percentage deviation of the robust efficiency score from its deterministic counterpart, i.e., $PoR_o^{(\mathcal{U})} = \frac{\left|\theta_o^{(D)} - \theta_o^{(\mathcal{U})}\right|}{\theta_o^{(D)}} \times 100\%, \forall o \in \{1,2,\ldots,n\}$, where $\theta_o^{(D)}$ and $\theta_o^{(\mathcal{U})}$ represent the efficiency score of DMU_o under the deterministic model and the model with the uncertainty set, respectively. A lower PoR indicates reduced sensitivity of the efficiency score to uncertainty and therefore reflects greater robustness and stability of the model. Here, robustness refers to the ability of a model to maintain consistent efficiency scores in the presence of data uncertainty.

As shown in Table 2, DMU2 and DMU3 maintain the PoR of 0.00% across all three uncertainty sets, confirming their strong and consistent performance even when inputs and outputs are subject to uncertainty. This suggests that these DMUs lie firmly on the efficient frontier and remain unaffected by the uncertainty levels. In contrast, DMU5 exhibits the highest sensitivity to uncertainty, with efficiency increases of 11.91% under the variable budgeted model and 11.32% under the order statistic uncertainty set. The lowest PoR for DMU5 is observed under the budgeted model (9.53%), suggesting that this setting introduces the least deviation from its deterministic score and therefore offers a more stable and appropriate robustness adjustment for this DMU. Likewise, DMU1 and DMU4 experience moderate but consistent changes in efficiency. Both show the PoR values of 9.51–9.52% under budgeted and variable budgeted sets, which increase to 11.32% under the order statistic model. This pattern reflects the more conservative nature of the order statistic approach.

Overall, the results highlight the trade-off between performance and robustness, demonstrating how the choice of uncertainty model affects efficiency scores and guides the identification of settings with more robust performance under uncertainty.

4 Case study

Uncertainty plays a critical role in the performance assessment of banking institutions when using DEA. In practical bank evaluation, data uncertainty can lead to inaccurate and fragile results when relying on traditional DEA models. Given the complex and dynamic nature of the banking sector, the accuracy of efficiency evaluations can be significantly impacted by uncertainties in input and output data. Addressing these uncertainties is essential for obtaining reliable and robust performance measures, which are crucial for informed decision-making and strategic planning in the banking industry.

In this section, we extend this line of research by using three real-world datasets derived from Zervopoulos et al. (2023). These samples consist of 50, 80, and 100 banks based in the European Union. The selection of input and output variables is motivated by the need to capture the key factors that contribute to a bank's efficiency, as per the intermediation approach (Sealey and Lindley 1977), which views banks as entities that use purchased funds to generate assets (Ayadi et al. 2016). The inputs selected for the analysis include three inputs— (x_1) Deposits & short-term funding; (x_2) Equity; and (x_3) Fixed assets —and two outputs— (y_1) Gross loans; and (y_2) Other earning assets. These variables are chosen because they reflect the core operations of a bank, though it is recognised that certain indicators may not directly capture overall performance in every context. For instance, fixed assets, typically considered stable over a year, might present uncertainty in the form of depreciation, market value fluctuations, or regulatory changes affecting the bank's operational capacity. The outputs—gross loans and other earning assets—are linked to the bank's revenue generation capacity, thus influencing its financial performance. These variables are assumed to reflect uncertainty due to factors such as market volatility, changes in regulatory requirements, and economic shifts that can affect both input and output. Descriptive statistics for the input and output measures are provided in Table 3, which details the variables used across the three samples.

4.1 Set-up

Each uncertainty set can be adjusted using a scaling parameter. We set the uncertainty level for all the input and output data ($\widehat{x}_{ij}$ and $\widehat{y}_{rj}$) at 1%, 5%, and 10% of the nominal value to capture a range of potential variations and assess the robustness of our results under different levels of uncertainty (Hatami-Marbini and Arabmaldar 2021; Toloo et al. 2022)[9]. These levels are chosen to reflect varying degrees of confidence in the data, allowing us to evaluate how sensitive the performance measures are to changes in the accuracy of input and output values. In addition, we set the probabil-

[9] For example, with a 5% uncertainty level, a nominal value of 300 varies within an interval of [295, 315].

Table 3 Descriptive statistics of inputs and outputs (in thousand USD) for three samples

Descriptive statistics	Input 1(x_1)	Input 2(x_2)	Input 3(x_3)	Output 1(y_1)	Output 2(y_2)
	Deposits & Short-term funding	Equity	Fixed assets	Gross loans	Other earning assets
Sample size: 50					
Mean	20,983,682.62	1,502,051.78	128,065.61	20,018,731.27	10,324,543.59
Min	26,892.34	3,181.62	162.74	2,977.02	4,770.21
Max	322,973,827.28	18,062,174.57	1,714,310.00	323,764,206.81	134,002,711.32
SD	64,139,219.71	3,804,040.20	387,995.41	64,439,147.23	29,939,498.33
Sample size: 80					
Mean	14,586,792.70	1,065,451.81	98,156.43	13,768,655.28	7,302,286.86
Min	26,892.30	3,181.63	162.66	2,977.04	4,770.24
Max	322,973,827.30	18,062,174.56	1,714,310.03	323,764,206.83	134,002,711.32
SD	51,359,518.20	3,062,081.60	324,484.44	51,594,069.12	24,094,636.20
Sample size: 100					
Mean	14,658,123.33	1,097,496.34	147,073.36	14,133,661.50	7,202,571.52
Min	26,892.29	3,181.61	162.75	2,977.02	4,770.19
Max	322,973,827.31	18,062,174.59	3,383,799.11	323,764,206.77	134,002,711.31
SD	48,031,165.74		483,785.42	49,287,487.14	22,404,410.84

ity of constraint violation for both inputs and outputs to less than 1%, reflecting a reasonable level of risk tolerance for decision-makers. Let us consider the following uncertainty sets:

- *Budgeted uncertainty*: To model data uncertainty for budgeted case, an appropriate level of uncertainty budget Γ can be selected based on the following equation; $\Gamma(\varepsilon) = 1 + \Omega^{-1}(1-\varepsilon)\sqrt{n}$, where Ω shows the cumulative distribution of the standard Gaussian variable, n is the number of uncertain inputs and outputs in each constraint, and ε denotes the violation probability of the constraints (Bertsimas and Sim 2004). For this case study, with sample sizes of 50, 80, and 100 banks, and a violation probability of input/output constraint sets at less than 1%, the required levels of the budget of uncertainty Γ are at least 17.40, 21.75, and 24.26, respectively. These values ensure that the model is robust against approximately 34%, 27%, and 24% of the uncertain data achieving their worst-case values.
- *Variable budgeted uncertainty*: Under variable budgeted uncertainty, we use the cardinality of the robust optimal solution ($\|\boldsymbol{\lambda}^*\| = \sum_{j=1}^{n} \lambda_j^*$) rather than the number of uncertain data, n, in each constraint. First, the optimal solution values for $\boldsymbol{\lambda}^*$ are obtained using the robust DEA model (5) for a given uncertain parameter Γ and for each input and/or output constraint. In this study, for sample sizes of 50, 80, and 100, and with a violation probability of input/output constraints below 1%, the values of $\|\boldsymbol{\lambda}^*\|$ fall within the ranges [0.0017, 12.6023], [0.0007, 12.2696], and [0.0007, 60.7252], respectively. The maximum value of $\|\boldsymbol{\lambda}^*\|$ is then used to estimate the affine functions $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$. We estimate the best over-approximating affine functions $\gamma^x(\boldsymbol{\lambda})$ and $\gamma^y(\boldsymbol{\lambda})$ by

$\max\left|\beta\left(\mathbf{\Gamma}\right)-\gamma^{x}\left(\boldsymbol{\lambda}\right)\left(\gamma^{y}\left(\boldsymbol{\lambda}\right)\right)\right|$ to ensure that $\gamma^{x}\left(\boldsymbol{\lambda}\right)$ and$\gamma^{y}\left(\boldsymbol{\lambda}\right)$ are an upper approximate of $\Gamma(\varepsilon)$. This approach guarantees that $\mathcal{U}_i^{BV}\left(\lambda\right)$ and $\mathcal{U}_r^{BV}\left(\boldsymbol{\lambda}\right)$ yields the probabilistic bounds equivalent to $\mathcal{U}_i^{B}\left(\Gamma_i^{x}\right)$ and $\mathcal{U}_r^{B}\left(\Gamma_r^{y}\right)$, respectively. In this study, the over-approximating $\Gamma(\varepsilon)$ is estimated using a linear function. In other words, for each value of n, we compute $\Gamma(\varepsilon)=\beta_{\varepsilon}(n)$ and then derive the affine functions $\gamma_{\varepsilon}^{x}(\boldsymbol{\lambda})$ and $\gamma_{\varepsilon}^{y}(\boldsymbol{\lambda})$ that overestimate β_{ε}. It should be noted that analysts may use alternative affine functions considered more appropriate for their specific problems or adjust them based on experimental results.

- *Order statistic uncertainty*: The proposed robust model (16) with the order statistic uncertainty set requires the quantiles ρ_{jk} as input parameters. In practical situations where historical data is unavailable, experts may select these parameters based on institutional knowledge and experience (Zhang and Gupta 2023). For this study, to ensure a fair comparison with the budgeted and variable budgeted uncertainty sets, we use Theorem 1 to identify the quantiles ρ_{jk}. In doing so, we set $\mathbf{\Gamma}^{x}-\left[\mathbf{\Gamma}^{x}\right]$ and $\mathbf{\Gamma}^{y}-\left[\mathbf{\Gamma}^{y}\right]$ for input and output constraints. For sample sizes of 50, 80, and 100, and with a violation probability of input/output constraints of less than 1%, the quantiles ρ_{jk} are selected to be at least 0.40, 0.75, and 0.26, respectively. While Theorem 1 suggests that the robust models (5) and (16) yield equivalent results, the lack of historical data and the simplified calculations indicate that the results are approximately the same.

To deepen the discussion of these uncertainty sets within a banking context, we elaborate on how each framework captures distinct types of operational risk and informs managerial decision-making. In the banking sector, budgeted uncertainty represents situations where only a limited number of input and output parameters are expected to deviate from their nominal values, subject to a predefined deviation budget. This is particularly realistic in diversified loan portfolios, where adverse events may affect only certain borrower segments (e.g., small businesses in a particular region) without causing system-wide disruption. The budgeted model captures this partial deviation scenario by allowing a bounded number of coefficients to change, thereby enhancing robustness without being excessively conservative. This ensures that banks maintain performance under typical volatility without unnecessarily restricting lending or capital flows. Variable budgeted uncertainty builds on this by allowing the uncertainty budget itself to vary in response to external indicators such as macroeconomic conditions, variations in credit risk assessments, and real-time stress-testing feedback. For example, a bank may tighten the uncertainty budget when facing early warnings from market stress tests or regulatory alerts. This dynamic adjustment enables banks to tailor their robustness levels to prevailing conditions, supporting agile responses and more efficient capital deployment. In contrast, order statistic uncertainty models the impact of extreme deviations, focusing on the largest observed disruptions across the input–output space. In banking, this reflects rare but high-impact scenarios such as widespread defaults during financial crises or sudden liquidity shortages from mass customer withdrawals. By concentrating on these worst-case deviations, the model prioritises resilience against tail risks. While more conservative, this framework is well-suited for stress-testing, contingency planning, and regulatory capital adequacy assessments.

Beyond the mathematical modelling, the level of conservatism embedded in each robust optimisation approach carries significant implications for banking operations. A low level of conservatism—characterised by tight uncertainty budgets—enables banks to remain competitive by maximising lending volumes and returns, but it also exposes them to higher risk in volatile environments. Conversely, high conservatism increases capital buffers and reduces exposure to uncertainty but may result in lower profitability, reduced market share, and underutilised resources.

For example, a bank using the order statistic model may choose to hold excess capital to guard against worst-case loss scenarios, while a bank relying on the budgeted model may opt to reallocate capital based on the most probable disruptions. The ability to adjust conservatism in line with the institution's risk appetite, strategic goals, and regulatory requirements is essential. The robust DEA models presented in this study offer decision-makers structured tools to navigate these trade-offs, enhancing the transparency, accountability, and stability of performance assessments under uncertainty.

In the next step, we utilise the CPLEX and Gurobi solvers within the GAMS environment to execute the models efficiently and obtain results in polynomial time. These solvers are renowned for their robustness and capability in handling large-scale optimisation problems, allowing us to solve the models effectively and manage computational complexity.

4.2 Empirical results

Table 4 reports the descriptive statistics of efficiency scores for different levels of perturbations for three different sample sizes[10]. Notably, a 0% perturbation corresponds to the deterministic case, which is identical for all robust DEA models, regardless of the uncertainty sets used. As observed in Table 4, for each robust DEA model, increasing the level of perturbations leads to higher robust efficiency scores. Comparison reveals that the proposed robust DEA model with the order statistic uncertainty set, model (16), demonstrates smaller changes compared to the robust DEA models with budgeted and variable budgeted uncertainty sets, models (5) and (10), respectively. For example, with a sample size of 100 and 10% data uncertainty, the mean efficiency scores for the robust DEA models are as follows: 0.6020 for model (5), 0.6676 for model (10), and 0.5563 for model (16). These results indicate that the proposed robust DEA model with the order statistic uncertainty set is more robust compared to the models with budgeted and variable budgeted uncertainty sets[11]. This occurs due to the greater geometric flexibility of the order statistic uncertainty set compared to budgeted and variable budgeted uncertainty sets.

As shown in Table 4 and consistent with Theorem 3, the efficiency scores obtained from the robust DEA model (10) with the variable budgeted uncertainty set are higher

[10] It should be noted that in the reporting of tables and figures, we also use the names of the uncertainty sets to better illustrate the impact of each uncertainty set on the robust DEA models.

[11] Detailed results using the existing robust DEA model (5) and the proposed robust DEA models (10) and (16), for varying levels of perturbations across all three sample sizes are provided in *Supplementary Materials*.

Table 4 Descriptive statistics of efficiency scores for different levels of perturbations across three sample sizes

Model/ uncertainty		$N = 50$				$N = 80$				$N = 100$			
		Min	Median	Mean	Max	Min	Median	Mean	Max	Min	Median	Mean	Max
Model (5) budgeted	0%	0.2256	0.4867	0.5461	1.0000	0.2085	0.4449	0.5025	1.0000	0.2085	0.4520	0.5044	1.0000
	1%	0.2302	0.4965	0.5605	1.0000	0.2127	0.4539	0.5171	1.0000	0.2127	0.4611	0.5152	1.0000
	5%	0.2494	0.5577	0.5990	1.0000	0.2304	0.4917	0.5515	1.0000	0.2304	0.4996	0.5510	1.0000
	10%	0.2758	0.6167	0.6570	1.0000	0.2548	0.5437	0.6028	1.0000	0.2548	0.5524	0.6020	1.0000
Model (10) Variable budgeted	1%	0.2338	0.5254	0.5787	1.0000	0.2159	0.4683	0.5295	1.0000	0.2159	0.4695	0.5290	1.0000
	5%	0.2699	0.6038	0.6376	1.0000	0.2475	0.5351	0.5890	1.0000	0.2475	0.5358	0.5891	1.0000
	10%	0.3235	0.7164	0.7189	1.0000	0.2620	0.5646	0.6111	1.0000	0.2901	0.6289	0.6676	1.0000
Model (16) Order statistic	1%	0.2293	0.4945	0.5530	1.0000	0.2144	0.4575	0.5140	1.0000	0.2110	0.4574	0.5094	1.0000
	5%	0.2444	0.5272	0.5816	1.0000	0.2398	0.5118	0.5630	1.0000	0.2214	0.4799	0.5297	1.0000
	10%	0.2648	0.5711	0.6196	1.0000	0.2760	0.5889	0.6282	1.0000	0.2351	0.5096	0.5563	1.0000

than those from the robust DEA model (5) with the budgeted uncertainty set. Furthermore, in accordance with Theorem 6, Table 4 indicates that the efficiency scores for the robust DEA model (5) with budgeted and robust model (16) with order statistic uncertainty set are approximately the same.

The kernel density curves of efficiencies for the existing robust DEA models (5) and the proposed robust DEA models (10) and (16) are shown in Fig. 2, illustrating the distribution of efficiencies across different perturbation levels and sample sizes of 50, 80, and 100. The curves demonstrate increased convergence with larger sample sizes, with the most notable convergence occurring between the density curves of the existing robust DEA model (5) with the budgeted uncertainty set and the proposed robust DEA model (16) with the order statistic uncertainty set. In particular, the most significant convergence is observed for the proposed robust DEA model (16) with the order statistic uncertainty set when the sample size is 100 and the perturbation level is 10%. In addition to the findings described above, Fig. 2 shows that for all three uncertainty sets, as the level of perturbations increases from deterministic (0%) to 10%, the distribution of efficiencies becomes smoother.

The three robust DEA methods may give different ranks to each unit. We utilise a non-parametric statistical test to validate the fitness between them, and the correlation with each other. To assess potential shifts in efficiency rankings across banks under robust DEA models with budgeted, variable budgeted, and order statistic sets, we conduct Spearman rank correlation analysis for three various sample sizes of 50, 80, and 100, as presented in Table 5. The analysis reveals a strong and statistically significant correlation between the estimates from the existing robust DEA model (5) and the proposed robust DEA models (10) and (16), with correlation coefficients ranging from 0.9677 to 1. Notably, the correlation between the existing robust DEA model (5) and the proposed robust DEA models (10) and (16) increases with larger sample sizes. This empirical evidence, along with the convergence of efficiency densities illustrated in Fig. 2, supports the consistency and reliability of the alternative robust DEA models, particularly the robust DEA model with the order statistic uncertainty set.

The previous analyses demonstrate that data uncertainty has a significant impact on the outcomes of efficiency assessments. We utilise the concept of the PoR to evaluate how well banks can handle data uncertainty. Fig. 3 displays the average and standard deviation of the PoR for three robust DEA models across different sample sizes and levels of data perturbation. It is evident that both the average and standard deviation of the PoR increase with higher levels of data perturbation for all three robust DEA models. Therefore, decision-makers should consider data uncertainty to avoid making overly aggressive decisions when assessing organisational performance. Among the models, the robust DEA model (10) with variable budgeted uncertainty is more sensitive to data uncertainty, exhibiting a higher PoR as data perturbation levels increase. In contrast, the robust DEA model (16) with the order statistic uncertainty set provides more robust efficiency levels across all DMUs, making it a more consistent benchmark for efficiency comparison and better at accommodating data fluctuations.

To further analyse inefficiency and the PoR, we employ K-means clustering proposed by Jain (2010) to examine the performance of European banks with a sample

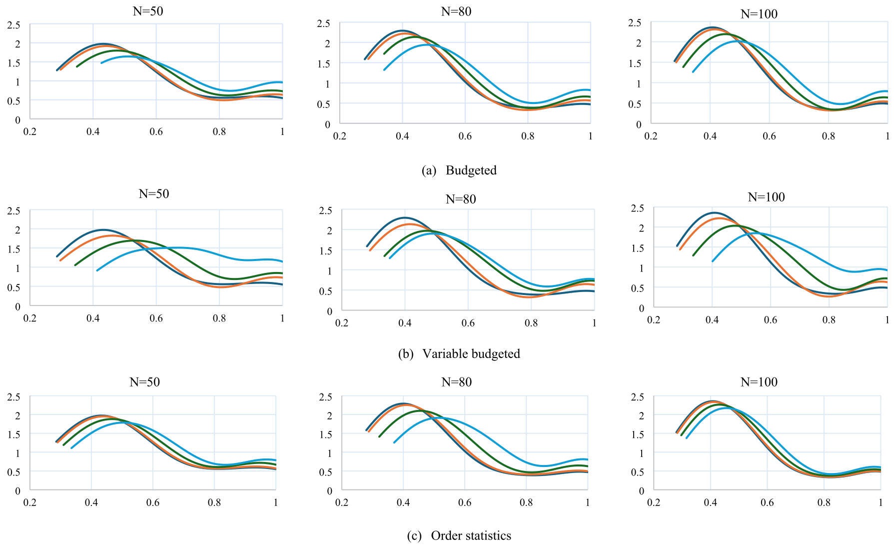

Fig. 2 Kernel density curves for robust efficiency scores (0%, 1%, 5%, 10%)

size of 100, utilising results from the robust DEA model (5) with budgeted uncertainty and the proposed robust DEA models (10) and (16) with variable budgeted and order statistic uncertainty sets, respectively. Inefficiency, defined as "1-efficiency", is used to align with the PoR, where lower values are preferable. The banks are classified into three clusters based on average inefficiency and the average PoR for deterministic and uncertain scenarios (1%, 5%, and 10%).

Fig. 4 illustrates the clustering distribution of the 100 European banks. The banks are divided into three clusters: cluster 1 (At-risk), cluster 2 (Moderate), and cluster 3 (Excellent), with each cluster defined by its centre and marked with a red circle, diamond, and triangle, respectively. The banks in cluster 3 demonstrate lower inefficiency and PoR, indicating superior (excellent) performance relative to the group average. As detailed in Supplementary Materials, banks in cluster 3 are efficient in both deterministic and uncertain conditions.

By analysing these clusters, bank managers and data analysts can gain valuable insights into the relative performance of different banks under uncertainty. This approach helps identify which banks are performing efficiently and which are lagging, providing a clear picture of how well banks handle data perturbations. The "Excellent" cluster, characterised by low inefficiency and a favourable PoR, can serve as a benchmark for best practices and high performance in managing uncertainty in both inputs and outputs.

These banks demonstrate effective resource management and operational efficiency despite data fluctuations, setting a standard that others can aim to emulate. For banks in the “Moderate” cluster, there is potential for improvement by focusing on optimising their resource allocation. This may involve reducing inputs while maintaining outputs, thus enhancing their efficiency and moving closer to the performance of the “Excellent” cluster. Banks in the “At-Risk” cluster face more sig-

Table 5 Spearman rank correlations at different levels of perturbations for three sample sizes

			Model (5) Budgeted				Model (10) Variable budgeted			Model (16) Order statistic		
Size			0	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	Model (5) Budgeted	0	1.0000									
		1%	0.9979	1.0000								
		5%	0.9895	0.9916	1.0000							
		10%	0.9677	0.9688	0.9891	1.0000						
	Model (10) Variable budgeted	1%	0.9827	0.9822	0.9741	0.9508	1.0000					
		5%	0.9783	0.9795	0.9718	0.9492	0.9961	1.0000				
		10%	0.9787	0.9799	0.9708	0.9530	0.9905	0.9938	1.0000			
	Model (16) Order statistic	1%	1.0000	0.9979	0.9895	0.9677	0.9827	0.9783	0.9787	1.0000		
		5%	1.0000	0.9979	0.9895	0.9677	0.9827	0.9783	0.9787	1.0000	1/0000	
		10%	0.9994	0.9971	0.9887	0.9682	0.9811	0.9788	0.9792	0.9994	0.9994	1.0000
80	Model (5) Budgeted	0	1.0000									
		1%	0.9981	1.0000								
		5%	0.9983	0.9990	1.0000							
		10%	0.9859	0.9874	0.9861	1.0000						
	Model (10) Variable budgeted	1%	0.9926	0.9919	0.9939	0.9772	1.0000					
		5%	0.9885	0.9869	0.9890	0.9744	0.9972	1.0000				
		10%	0.9882	0.9863	0.9888	0.9742	0.9973	0.9997	1.0000			
	Model (16) Order statistic	1%	0.9999	0.9981	0.9983	0.9859	0.9926	0.9884	0.9882	1.0000		
		5%	0.9997	0.9978	0.9980	0.9861	0.9920	0.9887	0.9885	0.9997	1.0000	
		10%	0.9978	0.9976	0.9979	0.9860	0.9911	0.9873	0.9871	0.9978	0.9980	1.0000

Table 5 (continued)

			Model (5) Budgeted				Model (10) Variable budgeted			Model (16) Order statistic		
Size			0	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	Model (5) Budgeted	0	1.0000									
		1%	0.9993	1.0000								
		5%	0.9984	0.9990	1.0000							
		10%	0.9938	0.9942	0.9951	1.0000						
	Model (10) Variable budgeted	1%	0.9936	0.9937	0.9960	0.9917	1.0000					
		5%	0.9903	0.9900	0.9926	0.9913	0.9977	1.0000				
		10%	0.9890	0.9893	0.9922	0.9913	0.9968	0.9982	1.0000			
	Model (16) Order statistic	1%	0.9999	0.9993	0.9984	0.9938	0.9936	0.9903	0.9890	1.0000		
		5%	0.9998	0.9994	0.9986	0.9940	0.9930	0.9904	0.9891	0.9998	1.0000	
		10%	0.9998	0.9994	0.9986	0.9940	0.9930	0.9904	0.9891	0.9998	1.0000	1.0000

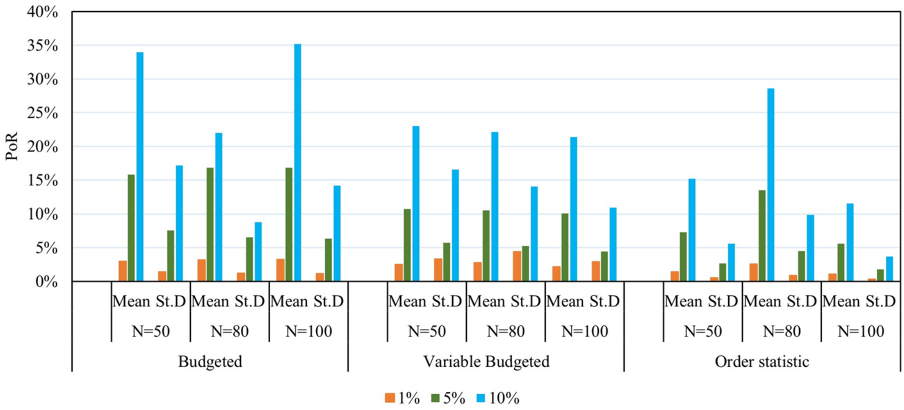

Fig. 3 The PoR for different level of data perturbations and samples

nificant challenges and may require more comprehensive interventions. These banks should prioritise reducing inefficiencies by significantly improving their resource utilisation. This could involve re-evaluating and re-allocating resources, enhancing operational processes, and addressing major inefficiencies to boost performance. The goal is to reduce inputs without compromising outputs, thereby improving efficiency and robustness under uncertainty. Overall, this clustering approach provides a framework for understanding performance variability under data uncertainty. By targeting strategies to optimise resource usage and manage uncertainties, banks can improve their efficiency, better handle data perturbations, and achieve more stable and robust performance outcomes.

4.3 Managerial implications

In the banking sector, where the conversion of multiple financial inputs into outputs is complex, benchmarking methods such as DEA are frequently employed to assess performance (Fukuyama et al. 2023; Tzeremes 2015). However, traditional DEA models often overlook data uncertainty—an inherent challenge in banking due to market volatility, regulatory changes, and other unpredictable factors (Zervopoulos et al. 2023). As a result, these deterministic DEA models may be less effective in capturing the true performance of banks under uncertain conditions.

Our study develops robust DEA models and compares them with existing robust DEA approaches, using empirical data from European banks. The findings highlight that incorporating uncertainty into efficiency measurements can indeed be costly for banks. In particular, our results show that the costs associated with robustness vary based on uncertainty levels (1%, 5%, and 10%) and sample sizes (50, 80, and 100). This variation offers managers the flexibility to choose the appropriate robust DEA model based on their risk tolerance and expertise, allowing for more informed decision-making in an unpredictable environment. The study's findings highlight that the proposed robust DEA models represent a substantial enhancement over traditional and existing robust DEA approaches. By accounting for various types of uncertainty and different sample sizes, we demonstrate that these models provide a more reliable

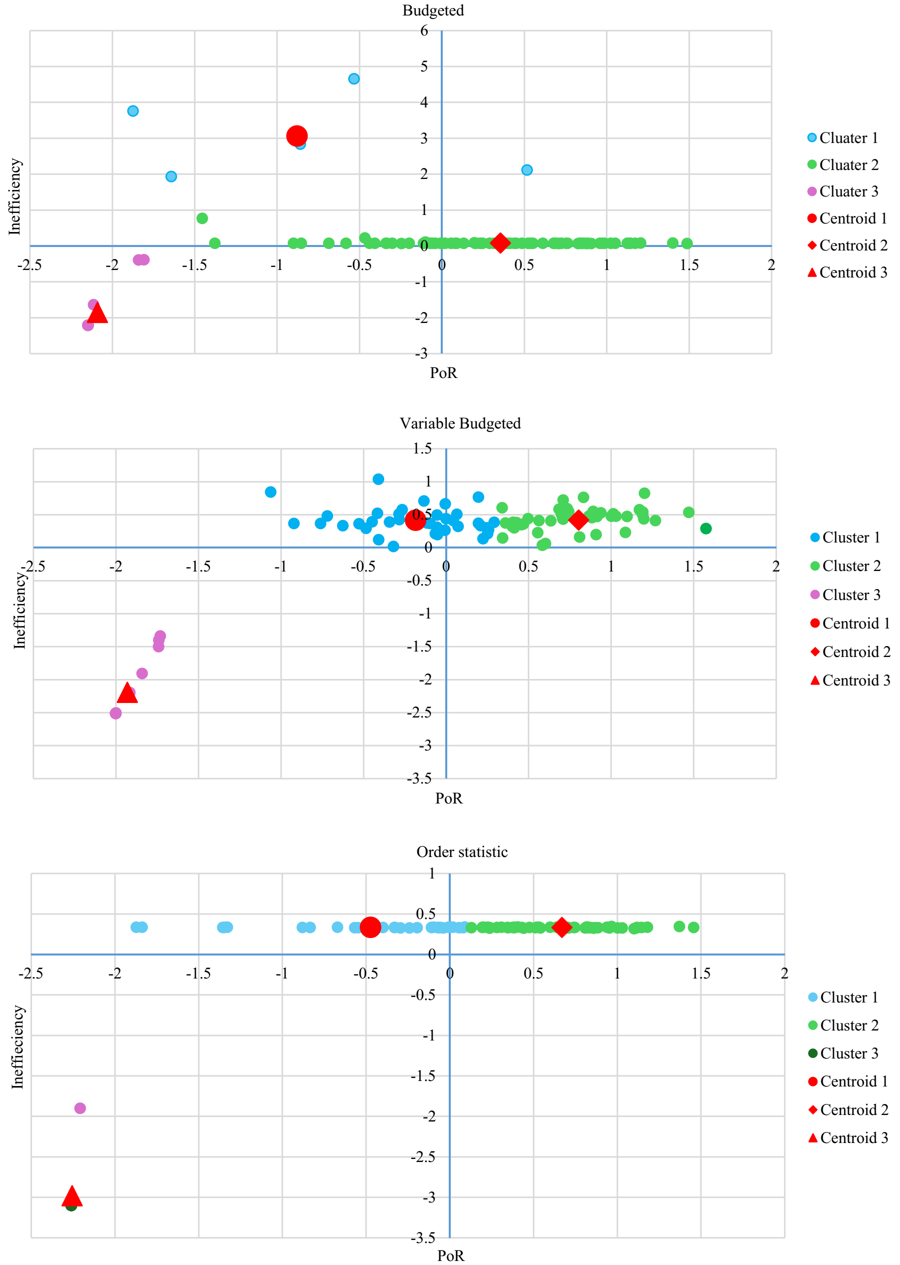

Fig. 4 Clustering distribution positions of the banks

and resilient assessment of efficiency. While our empirical study focuses on a specific financial institution, the robust DEA approach is adaptable to various banking contexts, including online banking and branch networks. Given the complexities of the banking industry—such as regulatory changes, market volatility, and economic disruptions—handling data uncertainty and developing robust DEA models to align with real-world conditions is crucial. The proposed model enables bank managers to evaluate performance under varying levels of conservatism, revealing how different banking regions or branches respond to varying uncertainties. These insights enable managers to develop effective strategies that enhance decision-making and improve bank performance in the face of unpredictable events, such as financial crises, interest rate fluctuations, and economic downturns.

The findings indicate that only a few banks—such as Abbey National Treasury Services Plc, ABC International Bank Plc, Airbus Group Bank GmbH, Aletti & C. Banca di Investimento Mobiliare SpA-Banca Aletti & C. SpA, Alior Bank Spółka Akcyjna, American Express Austria Bank GmbH, AXA Bank Europe SA/NV, Banca Aletti & C. Spa, and Banca Mediolanum SpA—maintain robust efficiency under both deterministic and uncertain conditions. These institutions exemplify best practices in managing operational and environmental uncertainties, demonstrating resilience and adaptability in a volatile sector. Policymakers should take note of these examples and prioritise addressing operational uncertainties in their strategies. By improving service quality, banks can attract a larger customer base and increase transaction volumes. Furthermore, effective management practices that tackle unforeseen events and operational challenges, such as fluctuations in transaction volumes and customer wait times, are crucial for sustaining high performance.

In conclusion, the study highlights the importance of incorporating uncertainty considerations into efficiency assessments. The robust DEA models developed in this paper provide a valuable framework for evaluating bank performance in a more realistic and comprehensive manner. For banks and policymakers alike, understanding and applying these models can lead to more informed decisions, improved service quality, and enhanced overall performance in the face of an unpredictable financial landscape.

5 Conclusion

DEA models are widely used to evaluate performance but often neglect data uncertainties that are prevalent in real-world environments, such as financial institutions and hospitals. This oversight can lead to unreliable outcomes when minor data fluctuations occur. To address this challenge, robust optimisation has been incorporated into DEA models to improve their reliability under uncertain conditions. This paper introduces two new robust DEA models—using order statistic and variable budgeted uncertainty sets—to extend existing robust DEA models and reduce inefficiencies in the presence of uncertainties in both inputs and outputs. It discusses how existing robust DEA models under budgeted uncertainty sets represent a special case of the proposed models when the robust parameter is appropriately selected. The budgeted uncertainty set is superior for stable environments, offering simplicity and computational efficiency, while the variable budgeted uncertainty set is suited for dynamic contexts with fluctuating uncertainties. The order statistic uncertainty set is particularly valuable for managing extreme risk factors. The decision to use one uncertainty set over another depends on the nature of the uncertainties involved and the specific priorities in risk management. This choice allows for tailoring performance assessments and risk management strategies according to specific needs and risk profiles.

Future research could explore several avenues to further improve robust DEA models. Key areas include investigating performance factors that significantly impact robustness in both deterministic and uncertain situations, extending empirical analysis to larger datasets to understand the role of data uncertainty in big-data analytics (Khezrimotlagh et al. 2019), and applying various uncertainty sets to other DEA methods, such as the Malmquist and Luenberger productivity indices, to analyse productivity changes over time. A promising direction for future research is to extend the proposed model using a directional distance function (DDF), as demonstrated by Arabmaldar et al. (2023) for budgeted uncertainty sets, to further enhance its modelling flexibility. In addition, studies could focus on developing robust network and dynamic network DEA models using diverse uncertainty sets, improving their applicability in complex and uncertain environments. We also acknowledge the importance of data-driven methods for defining uncertainty sets in robust optimisation, as highlighted in Bertsimas et al. (2018). The application of data-driven techniques to define uncertainty sets directly from empirical distributions could further enhance the practical relevance of robust DEA models, particularly in data-rich environments where leveraging historical patterns can support more informed and adaptive decision-making.

Appendix: Proofs

Proof of Proposition 1 According to the defined uncertainty sets in (4) and the inner optimisation problem in model (3) based on $\mathcal{U}^B(\boldsymbol{\Gamma}^x)$ and $\mathcal{U}^B(\boldsymbol{\Gamma}^y)$, we have the following problems for the i^{th} input and r^{th} output constraints:

$$\begin{aligned}
&\underline{i^{\text{th}}\text{ input constraint}}\\
&\max \sum_{j\in J_i^x, j\neq o} \widehat{x}_{ij}\,|\lambda_j|\,z_{ij}^x + \widehat{x}_o\,|\lambda_o-\theta_o|\,z_{io}^x\\
&\text{s.t.}\\
&\sum_{j\in j_i^x} z_{ij}^x \le \Gamma_i^x,\\
&0\le z_{ij}^x\le 1, \quad \forall j\in J_i^x.
\end{aligned}$$

$$\begin{aligned}
&\underline{r^{\text{th}}\text{ Output constraint}}\\
&\max \sum_{j\in J_r^y, j\neq o} \widehat{y}_{rj}\,|\lambda_j|\,z_{rj}^y + \widehat{y}_o\,|\lambda_o-1|\,z_{ro}^y\\
&\text{s.t.}\\
&\sum_{j\in j_r^y} z_{rj}^y \le \Gamma_r^y,\\
&0\le z_{rj}^y\le 1, \quad \forall j\in J_r^y.
\end{aligned} \tag{A1}$$

Without loss of generality, let us focus on the input constraint. Examining (8) closely reveals that the optimal solution of model (A1) for the input constraint is $Z^* = \left(z_{ij}^{x^*}, z_{io}^{x^*}\right), \forall j \in J_i^x$. At optimality, the variables $z_{ij}^{x^*}$ and $z_{io}^{x^*}$ are equal to 1 to satisfy the integer part of Γ_i^x, denoted by $[\Gamma_i^x]$, while the fractional part $\Gamma_i^x - [\Gamma_i^x]$ is distributed between $z_{ij}^{x^*}$ and $z_{io}^{x^*}$ to maximise the objective function of (A1). This is equivalent to selecting a subset of $\{S_i^x \cup \{\alpha_i^x\} | S_i^x \subseteq J_i^x, |S_i^x| = [\Gamma_i^x], \alpha_i^x \in J_i^x \backslash S_i^x\}$ with the corresponding objective function (6). □

Proof of Theorem 1 The protection functions $\beta_i(\boldsymbol{\lambda}^*, \mathcal{U}_i^{VB}(\boldsymbol{\lambda}))$ and $\beta_r\left(\boldsymbol{\lambda}^*, \mathcal{U}_r^{VB}(\boldsymbol{\lambda})\right)$ in model (9), which correspond to their input and output constraints, can be represented by the following optimisation problems, respectively:

$$
\begin{aligned}
&\max \sum_{j \in J_i^x (j \neq o)} |\lambda_j| \, z_{ij}^x \hat{x}_{ij} + |\lambda_o - \theta_o| \, z_{io}^x \hat{x}_{io} \\
&\text{s.t.} \\
&\sum_{j \in J_i^x} z_{ij}^x \leq \gamma_i^x(\boldsymbol{\lambda}), \\
&z_{ij}^x \leq 1, \quad \forall j \in J_i^x, \\
&z_{ij}^x \geq 0, \quad \forall j \in J_i^x,
\end{aligned} \tag{A2}
$$

$$
\begin{aligned}
&\max \sum_{j \in J_r^y (j \neq o)} |\lambda_j| \, y_{rj}^y \hat{y}_{rj} + |\lambda_o - 1| \, z_{ro}^y \hat{y}_{ro} \\
&\text{s.t.} \\
&\sum_{j \in J_r^y} z_{rj}^y \leq \gamma_r^y(\boldsymbol{\lambda}), \\
&z_{rj}^y \leq 1, \quad \forall j \in J_r^y, \\
&z_{rj}^y \geq 0, \quad \forall j \in J_r^y.
\end{aligned} \tag{A3}
$$

Consider the dual models of (A2) and (A3) as follows:

$$
\begin{aligned}
&\min p_i^{x'} \gamma_i^x(\boldsymbol{\lambda}) + \sum_{j \in J_i^x, j \neq o} q_{ij}^{x'} \\
&\text{s.t.} \\
&p_i^{x'} + q_{ij}^{x'} \geq |\lambda_j| \, \hat{x}_{ij}, \quad \forall j \in J_i^x, j \neq o, \\
&p_i^{x'} + q_{io}^{x'} \geq |\lambda_o - \theta_o| \, \hat{x}_{io}, \quad o \in J_i^x, \\
&p_i^{x'}, q_{ij}^{x'} \geq 0, \quad \forall j \in J_i^x,
\end{aligned} \tag{A4}
$$

$$\min p_r^{y'} \gamma_r^y(\boldsymbol{\lambda}) + \sum_{j \in J_i^x, j \neq o} q_{rj}^{y'}$$
$$\text{s.t.}$$
$$p_r^{y'} + q_{rj}^{y'} \geq |\lambda_j| \hat{y}_{rj}, \quad \forall j \in J_r^y, j \neq o,$$
$$p_r^{y'} + q_{ro}^{y'} \geq |\lambda_o - 1| \hat{y}_{ro}, \quad o \in J_r^y, \tag{A5}$$
$$p_r^{y'}, q_{rj}^{y'} \geq 0, \quad \forall j \in J_r^y,$$

where $p_i^{x'}(p_r^{y'})$ and $q_{ij}^{x'}$ $(q_{rj}^{y'})$ are the dual variables associated with the first and second sets of constraints in models (A2) and (A3), respectively. Since $\lambda_{j \neq 0} \geq 0, (\theta_o - \lambda_o) \geq 0$, and $(1 - \lambda_o) \geq 0$, the absolute signs in models (A4) and (A5) can be removed. As a result, the input constraint sets in model (9) can be expressed as follows:

$$\sum_{j \in J} \lambda_j x_{ij} + p_i^{x'} \gamma_i^x(\boldsymbol{\lambda}) + \sum_{j \in J_i^x, j \neq o} q_{ij}^{x'} \leq \theta_o x_{io}, \quad \forall i \in I,$$
$$p_i^{x'} + q_{ij}^{x'} \geq \lambda_j \hat{x}_{ij}, \quad \forall i \in I, \forall j \in J_i^x, j \neq o,$$
$$p_i^{x'} + q_{io}^{x'} \geq (\theta_o - \lambda_o) \hat{x}_{io}, \quad \forall i \in I, o \in J_i^x, \tag{A6}$$
$$p_i^{x'}, q_{ij}^{x'} \geq 0, \quad \forall i \in I, \forall j \in J_i^x,$$
$$\lambda_j \geq 0, \quad \forall j \in J.$$

Likewise, the output constraint sets in model (9) can be reformulated with the following set of constraints:

$$\sum_{j \in J} \lambda_j y_{rj} - p_r^{y'} \gamma_r^y(\boldsymbol{\lambda}) - \sum_{j \in J_i^x, j \neq o} q_{rj}^{y'} \geq y_{ro}, \quad \forall r \in R$$
$$p_r^{y'} + q_{rj}^{y'} \geq \lambda_j \hat{y}_{rj}, \quad \forall r \in R, \forall j \in J_r^y, j \neq o,$$
$$p_r^{y'} + q_{ro}^{y'} \geq (1 - \lambda_o) \hat{y}_{ro}, \quad \forall r \in R, o \in J_r^y, \tag{A7}$$
$$p_r^{y'}, q_{rj}^{y'} \geq 0, \quad \forall r \in R, \forall j \in J_r^y,$$
$$\lambda_j \geq 0, \quad \forall j \in J.$$

The bilinear terms $p_i^{x'} \gamma_i^x(\boldsymbol{\lambda}) = p_i^{x'}(\gamma_0^x + \sum_{j \in J_i^x} \gamma_j^x \lambda_j)$ and $p_r^{y'} \gamma_r^y(\boldsymbol{\lambda}) = p_r^{y'}(\gamma_0^y + \sum_{j \in J_r^y} \gamma_j^y \lambda_j)$ in models (A6) and (A7), respectively, should be linearised. In doing so, we introduce the variable alteration $w_{ij}^x = p_i^{x'} \lambda_j$ and $w_{rj}^y = p_r^{y'} \lambda_j$, and then apply the approach proposed by Poss (2013, Proposition 1, page 86), as formulated below:

$$\begin{aligned}
&\sum_{j\in J}\lambda_j x_{ij}+p_i^{x'}\gamma_0^x+\sum_{j\in J_i^x}\gamma_j^x w_{ij}^x+\sum_{j\in J_i^x}q_{ij}^{x'}\leq\theta_0 x_{io},\quad \forall i\in I,\\
&p_i^{x'}+q_{ij}^{x'}\geq\lambda_j\hat{x}_{ij},\quad \forall i\in I,\forall j\in J_i^x, j\neq o,\\
&p_i^{x'}+q_{io}^{x'}\geq(\theta_o-\lambda_o)\,\hat{x}_{io},\quad \forall i\in I, o\in J_i^x,\\
&w_{ij}^x-p_i^{x'}\geq -M_x\,(1-\lambda_j)\,,\quad \forall i\in I,\forall j\in J_i^x,\\
&\lambda_j\leq M_x' h_j^x,\quad \forall j\in J,\\
&p_i^{x'}, q_{ij}^{x'}, w_{ij}^x\geq 0,\quad \forall i\in I,\forall j\in J_i^x,\\
&h_j^x\in\{0,1\}^n,\quad \forall j\in J,\\
&\lambda_j\geq 0,\quad \forall j\in J,
\end{aligned}\tag{A8}$$

and

$$\begin{aligned}
&\sum_{j\in J}\lambda_j y_{rj}-p_r^{y'}\gamma_0^y-\sum_{j\in J_r^y}\gamma_j^y w_{rj}^y-\sum_{j\in J_r^y}q_{rj}^{y'}\geq y_{ro},\quad \forall r\in R,\\
&p_r^{y'}+q_{rj}^{y'}\geq\lambda_j\hat{y}_{rj},\quad \forall r\in R,\forall j\in J_r^y, j\neq o,\\
&p_r^{y'}+q_{ro}^{y'}\geq(1-\lambda_o)\,\hat{y}_{ro},\quad \forall r\in R, o\in J_r^y,\\
&w_{rj}^y-p_r^{y'}\geq -M_y\,(1-\lambda_j)\,,\quad \forall r\in R,\forall j\in J_r^y,\\
&\lambda_j\leq M_y' h_j^y,\quad \forall j\in J,\\
&p_r^{y'}, q_{rj}^{y'}, w_{rj}^y,\geq 0,\quad \forall r\in R, j\in J_r^y,\\
&h_j^y\in\{0,1\}^n,\quad \forall j\in J,\\
&\lambda_j\geq 0,\quad \forall j\in J,
\end{aligned}\tag{A9}$$

where M_x and M_y are sufficiently large constants. Since each $p_i^{x'}$ and $p_r^{y'}$ must satisfy $p_i^{x'}+q_{ij}^{x'}\geq\lambda_j\widehat{x}_{ij}$ and $p_r^{y'}+q_{rj}^{y'}\geq\lambda_j\widehat{y}_{rj}$, it is sufficient for M_x and M_y to be as large as $\max_j(\widehat{x}_{ij})$ $(\forall i\in I)$ and $\max_j(\widehat{y}_{rj})$ $(\forall r\in R)$, respectively. Moreover, since constraints (A8) and (A9) do not impose any additional restrictions on $p_i^{x'}$ and $p_r^{y'}$, we can choose M_x and M_y equal to $\max_j(\widehat{x}_{ij})$ $(\forall i\in I)$ and $\max_j(\widehat{y}_{rj})$ $(\forall r\in R)$, respectively. The auxiliary binary variables, $h_j^x, h_j^y\in\{0,1\}^n$, are introduced in models (A8) and (A9) to ensure that $\lambda_j\leq M_x' h_j^x, q_i^x\in B_{M_x'}(0)$ and $\lambda_j\leq M_y' h_j^y, q_r^y\in B_{M_y'}(0)$, where $B_{M_x'}(0)$ and $B_{M_y'}(0)$ represent the balls centred at the origin with radius M_x' and M_y', respectively, that are sufficiently large to maintain the feasibility of model (A8) and (A9)[12]. In other words, the optimal values of λ_j^* for models (A8) and (A9) fall within the boundaries defined by $B_{M_x'}(0)$ and $B_{M_y'}(0)$. Simply put, any optimal values of $\lambda_j^*(\forall j\in J)$ that lie within $B_{M_x'}(0)$ and $B_{M_y'}(0)$ are considered feasible

[12] It should be noted that the concept of the budgeted uncertainty set was initially proposed for binary decision variables and later extended to bounded real or integer variables (Poss 2013, Sect. 5, Theorem 2).

(see Poss 2013 for further details). Given that $\lambda^*_{(o\in\{1,2,\ldots,n\})} \leq 1$, it is assumed that $M'_x = M'_y = 1$. Rewriting constraints (A6)–(A9) completes the proof. □

Proof of Theorem 2 Assume $\lambda_o = 0\,(\forall j \in J_i^x, j \neq o)\,, \theta_o^{VB} = 1,\ \ p_i^{x'} = p_r^{y'} = w_{ij}^x = w_{rj}^y = q_{ij}^{x'} = q_{rj}^{y'} = 0\,(\forall i \in I, \forall r \in R, \forall j \in J_i^x)$, and $h_j^x = h_j^y = 1(\forall j \in J)$. This feasible solution completes the proof of the first part of the theorem.

To prove that $\theta_o^{VB*} \leq 1$, we first consider the feasible solution described in (i). For model (10), which is a minimisation model, the objective function value θ_o^{VB*} is at most 1. Next, we show that $\theta_o^{VB*} > 0$ by contradiction. Assume $\theta_o^{VBU*} = 0$ and the input constraints are converted to $\sum\limits_{j\in J} \lambda_j x_{ij} + p_i^{x'} \gamma_0^x + \sum\limits_{j\in J_i^x} \gamma_j^x w_{ij}^x + \sum\limits_{j\in J_i^x} q_{ij}^{x'} \leq \theta_0 x_{io}, \forall i \in I$. Given the non-negative assumption for the data, this implies $\lambda_j = p_i^{x'} = w_{ij}^x = q_{ij}^{x'} = 0, (\forall j \in J_i^x, \forall i \in I)$. This leads to a contradiction because the second constraint, $\sum\limits_{j\in J} \lambda_j y_{rj} - p_r^{y'} \gamma_0^y - \sum\limits_{j\in J_r^y} \gamma_j^y w_{rj}^y - \sum\limits_{j\in J_r^y} q_{rj}^{y'} \geq y_{ro}, \forall r \in R$, requires λ_j to be non-zero. Thus, combining these results, we have $0 < \theta_o^{VBU*} \leq 1$. □

Proof of Theorem 3 Consider the following equations (a) and (b) that corresponds to the input constraints in models (5) and (10), respectively:

$$\sum_{j\in J} \lambda_j x_{ij} + \sum_{j\in J_i^x} q_{ij}^x + \Gamma_i^x p_i^x \leq \theta_o x_{io} \Rightarrow \theta_o \geq \frac{\sum_{j\in J} \lambda_j x_{ij} + \sum_{j\in J_i^x} q_{ij}^x + \Gamma_i^x p_i^x}{x_{io}}, \forall i \in I, (a)$$

$$\sum_{j\in J} \lambda_j x_{ij} + p_i^{x'} \gamma_0^x + \sum_{j\in J_i^x} \gamma_j^x w_{ij}^x + \sum_{j\in J_i^x} q_{ij}^{x'} \leq \theta_0 x_{io}, \forall i \in I \Rightarrow$$

$$\theta_o \geq \frac{\sum_{j\in J} \lambda_j y_{rj} + p_i^{x'} \gamma_0^x + \sum_{j\in J_i^x} \gamma_j^x w_{ij}^x + \sum_{j\in J_i^x} q_{ij}^{x'}}{x_{io}}, \forall i \in I, \qquad (b)$$

where θ_o is a decision variable. We consider the following two cases to complete the proof:

Case (i): if $\gamma_i^x(\boldsymbol{\lambda}) = \Gamma_i^x$ and $\gamma_r^y(\boldsymbol{\lambda}) = \Gamma_r^y$, then the numerators in inequalities (a) and (b) coincide, resulting in $\theta_0^{B*} = \theta_0^{VB*}$.

Case(ii): Since $\gamma_i^x(\boldsymbol{\lambda})$ and $\gamma_r^y(\boldsymbol{\lambda})$ are the best over-approximating affine functions of Γ_i^x and Γ_r^y, respectively, we have $\theta_0^{VB*} > \theta_0^{B*}$. □

Proof of Theorem 4 By considering the protection functions in (15), Proposition 2, and the dual formulation of models (13) and (14), we obtain the following models:

$$\begin{aligned}&\min \textstyle\sum_{j\in J_i^x}\left(\psi_{ij}^x+\Phi_{ij}^x\right)\\&\text{s.t.}\\&\psi_{ij}^x+\Phi_{ik}^x\geq \widehat{x}_{ij}\left|\lambda_j\right|\rho_{ijk}^x, \quad \forall j,k\in J_i^x, j\neq o,\\&\psi_{io}^x+\Phi_{io}^x\geq \widehat{x}_{io}\left|\lambda_o-\theta_o\right|\rho_{io}^x, \quad o\in J_i^x.\end{aligned} \tag{A10}$$

$$\begin{aligned}&\min \textstyle\sum_{j\in J_r^y}\left(\psi_{rj}^y+\Phi_{rj}^y\right)\\&\text{s.t.}\\&\psi_{rj}^y+\Phi_{rk}^y\geq \widehat{y}_{rj}\left|\lambda_j\right|\rho_{rjk}^y, \quad \forall j,k\in J_r^y, j\neq o,\\&\psi_{ro}^y+\Phi_{ro}^y\geq \widehat{x}_{ro}\left|\lambda_o-1\right|\rho_{ro}^y, \quad o\in J_r^y.\end{aligned} \tag{A11}$$

It is straightforward to verify that the new robust DEA model (16) can be derived by applying strong duality and substituting models (A10) and (A11) into the input and output constraints (15). □

Proof of Theorem 5 Assume $\lambda_o=1, \lambda_j=0\,(\forall j\in J, j\neq o), \theta_o=1, \psi_{ij}^x=\Phi_{ik}^x=0\,(\forall i\in I, \forall j,k\in J_i^x), \psi_{rj}^y=\Phi_{rk}^y=0(\forall r\in R, \forall j,k\in J_r^y)$. This feasible solution completes the proof of the first part of the theorem.

(ii) First, we prove that $\theta_o^{OS*}\leq 1$. Given the feasible solution presented in (i) and considering model (16) as a minimisation problem, the objective function value θ_o^{OS*} is less than or equal to 1. Then, to show that $\theta_o^{OS*}>0$, we proceed by contradiction. Assume $\theta_o^{OS*}=0$, which changes the input constraints of model (16) to $\sum_{j\in J}x_{ij}\lambda_j+\sum_{j\in J_i^x}\left(\psi_{ij}^x+\Phi_{ij}^x\right)\leq 0, i\in I$, or $\sum_{j\in J_i^x}\left(\psi_{ij}^x+\Phi_{ij}^x\right)\leq -\sum_{j\in J}x_{ij}\lambda_j$, $\forall i\in I$. By summing up the third and fourth constraints in model (16), we arrive at $\sum_{j\in J_i^x}\left(\psi_{ij}^x+\Phi_{ij}^x\right)\geq -\sum_{j\in J}\widehat{x}_{ij}\lambda_j\rho_{ijk}$ leading to $\sum_{j\in J_i^x}\left(\psi_{ij}^x+\Phi_{ij}^x\right)=0$, which, as per the first constraint, is not feasible. □

Proof of Theorem 6 Without loss of generality, we consider the input constraint and assume that there is only one constraint for both the budgeted uncertainty set and the order statistic uncertainty set, thereby removing the i index.

Considering the robust counterpart DEA models (3) with the protection functions $\beta_i(\boldsymbol{\lambda}^*,\mathcal{U}^B(\boldsymbol{\Gamma}^x))$ and $\beta_r(\boldsymbol{\lambda}^*,\mathcal{U}^B(\boldsymbol{\Gamma}^y))$ and (15), we need to prove that the optimal objective values $\beta(\boldsymbol{\lambda}^*,\mathcal{U}^{BS}(\boldsymbol{\Gamma}^x))$ and $\beta(\boldsymbol{\lambda}^*,\mathcal{U}^{OS}(\boldsymbol{\varepsilon}))$ are equal. Since model (13) is the linear relaxation of the maximum weight assignment problem, which is known to have an integer optimal solution, for each $j\in J^x$, there exists a unique $k\in J^x$ such that $\eta_{jk}^x=1$. If $\eta_{jk}^x=1$, then $\widehat{x}_{ij}\left|\lambda_j\right|$ and $\left|\lambda_o-\theta_o\right|$ are paired with ρ_{jk}. As a result, based on $\beta(\lambda^*,\mathcal{U}^{BS}(\boldsymbol{\Gamma}^x))$ introduced in Proposition 1, and $\beta_r(\lambda^*,\mathcal{U}_r^{OS}(\boldsymbol{\rho}^{\boldsymbol{x}}))$ in Proposition 2, we have the following three cases:

If $1\leq k\leq |J^x|-\lfloor\boldsymbol{\Gamma}^x\rfloor-1$, then $\hat{x}_{ij}\left|\lambda_j\right|$ and $\left|\lambda_o-\theta_o\right|$ are paired with 0,
If $k=|J^x|-\lfloor\boldsymbol{\Gamma}^x\rfloor$, then $\hat{x}_{ij}\left|\lambda_j\right|$ and $\left|\lambda_o-\theta_o\right|$ are paired with $\boldsymbol{\Gamma}^x-\lfloor\boldsymbol{\Gamma}^x\rfloor$,
If $|J^x|-\lfloor\boldsymbol{\Gamma}^x\rfloor+1\leq k\leq |J^x|$, then $\widehat{x}_{ij}\left|\lambda_j\right|$ and $\left|\lambda_o-\theta_o\right|$ are paired with 1,

Hence, for all $\hat{x}_{i1}\left|\lambda_1\right|, \hat{x}_{i2}\left|\lambda_2\right|, \ldots, \hat{x}_{i|J^x|}\left|\lambda_{|J^x|}\right|$, we know that $\lfloor \mathbf{\Gamma}^x \rfloor$ of these terms are paired with 1, one term is paired with $\mathbf{\Gamma}^x - \lfloor \mathbf{\Gamma}^x \rfloor$, and the remaining terms are paired with 0. Therefore, model is equivalent to model (6), and their optimal objective values are the same, i.e., $\beta(\lambda^*, \mathcal{U}^{BS}(\mathbf{\Gamma}^{\boldsymbol{x}}))=\beta(\lambda^*, \mathcal{U}^{OS}(\boldsymbol{\rho}^{\boldsymbol{x}}))$. Thus, the proof is complete. □

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00291-025-00832-z.

References

Akbarian D (2020) Overall profit Malmquist productivity index under data uncertainty. Financ Innov 6(1):6

Arabmaldar A, Hatami-Marbini A, Loske D, Hammerschmidt M, Klumpp M (2024) Robust data envelopment analysis with variable budgeted uncertainty. Eur J Oper Res 315(2):626–641

Arabmaldar A, Jablonsky J, Hosseinzadeh Saljooghi F (2017) A new robust DEA model and super-efficiency measure. Optim Lett 66(5):723–736

Arabmaldar A, Sahoo BK, Ghiyasi M (2023) A generalized robust data envelopment analysis model based on directional distance function. Eur J Oper Res 311(2):617–632

Ayadi R, Naceur SB, Casu B, Quinn B (2016) Does Basel compliance matter for bank performance? J Financ Stab 23:15–32

Bandi C, Bertsimas D, Youssef N (2015) Robust queueing theory. Oper Res 63(3):676–700

Bandi C, Gupta D (2020) Operating room staffing and scheduling. Manuf Serv Oper Manage 22(5):958–974

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30(9):1078–1092

Ben-Tal A, Bertsimas D, Brown DB (2010) A soft robust model for optimization under ambiguity. Oper Res 58(4):1220–1234

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton

Ben-Tal A, Nemirovski A (1998) Robust convex optimization. Math Oper Res 23(4):769–805

Ben-Tal A, Nemirovski A (2000) Robust optimization of linear programming problems contamined with uncertain data. Math Program 88(3):411–424

Bertsimas D, Brown DB (2009) Constructing uncertainty sets for robust linear optimization. Oper Res 57(6):1483–1495

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. SIAM Rev 53(3):464–501

Bertsimas D, Gupta V, Kallus N (2018) Data-driven robust optimization. Math Program 167:235–292

Bertsimas D, Sim M (2004) The price of robustness. Oper Res 52(1):35–53

Chambers RG, Quiggin J (2000) Uncertainty, production, choice, and agency: the state-contingent approach. p 373

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision-making units. Eur J Oper Res 2:429–444

Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and Dea-Solver software, 2nd edn. Springer, Berlin

Daraio C, Simar L (2007) Advanced robust and nonparametric methods in efficiency analysis methodology and applications. Springer, Berlin

Dehghani Filabadi M, Mahmoudzadeh H (2022) Effective budget of uncertainty for classes of robust optimization. INFORMS J Optim 4(3):249–277

Despotis DK, Smirlis YG (2002) Data envelopment analysis with imprecise data. Eur J Oper Res 140:24–36

Emrouznejad A, Tavana M, Hatami-Marbini A (2014) The state of the art in fuzzy data envelopment analysis. In: Emrouznejad A, Tavana M (eds) Performance measurement with fuzzy data envelopment analysis. Springer, Berlin

Fukuyama H, Tsionas M, Tan Y (2023) Dynamic network data envelopment analysis with a sequential structure and behavioural-causal analysis: application to the Chinese banking industry. Eur J Oper Res 307(3):1360–1373

Goerigk M, Khosravi M (2023) Optimal scenario reduction for one- and two-stage robust optimization with discrete uncertainty in the objective. Eur J Oper Res 310(2):529–551

Goerigk M, Lendl S, Wulf L (2022) Recoverable robust representatives selection problems with discrete budgeted uncertainty. Eur J Oper Res 303(2):567–580

Gregory C, Darby-Dowman K, Mitra G (2011) Robust optimization and portfolio selection: the cost of robustness. Eur J Oper Res 212(2):417–428

Hatami-Marbini A (2019) Benchmarking with network DEA in a fuzzy environment. RAIRO - Oper Res 53(2):687–703

Hatami-Marbini A, Arabmaldar A (2021) Robustness of Farrell cost efficiency measurement under data perturbations: evidence from a US manufacturing application. Eur J Oper Res 295(2):604–620

Hatami-Marbini A, Arabmaldar A, Asu JO (2022a) Robust productivity growth and efficiency measurement with undesirable outputs: evidence from the oil industry. Or Spectr 44(4):1213–1254

Hatami-Marbini A, Arabmaldar A, Toloo M, Mahmoodi Nehrani A (2022b) Robust non-radial data envelopment analysis models under data uncertainty. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2022.118023

Hatami-Marbini A, Emrouznejad A, Tavana M (2011) A taxonomy and review of the fuzzy data envelopment analysis literature: two decades in the making. Eur J Oper Res 214(3):457–472

Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31(8):651–666

Kerstens K, Sadeghi J, Toloo M, Van de Woestyne I (2022) Procedures for ranking technical and cost-efficient units: with a focus on nonconvexity. Eur J Oper Res 300(1):269–281

Khezrimotlagh D, Zhu J, Cook WD, Toloo M (2019) Data envelopment analysis and big data. Eur J Oper Res 274(3):1047–1054

Li X, Xu G, Wu J, Xu C, Zhu Q (2024) Evaluation of bank efficiency by considering the uncertainty of nonperforming loans. Omega. https://doi.org/10.1016/j.omega.2024.103069

Liu P, Yang W, Guo T (2016) A discussion on the conservatism of robust linear optimization problems. Optim 65(8):1641–1650

Mostafaee A, Saljooghi FH (2010) Cost efficiency measures in data envelopment analysis with data uncertainty. Eur J Oper Res 202(2):595–603

Olesen OB, Petersen NC (1995) Chance constrained efficiency evaluation. Manage Sci 41(3):442–457

Olesen OB, Petersen NC (2016) Stochastic data envelopment analysis - a review. Eur J Oper Res 251(1):2–21

Pham M, Simar L, Zelenyuk V (2023) Statistical Inference for Aggregation of Malmquist Productivity Indices. Oper Res 174:1076–1086

Poss M (2013) Robust combinatorial optimization with variable budgeted uncertainty. 4OR 11(1):75–92

Poss M (2014) Robust combinatorial optimization with variable cost uncertainty. Eur J Oper Res 237(3):836–845

Sadjadi SJ, Omrani H (2008) Data envelopment analysis with uncertain data: an application for Iranian electricity distribution companies. Energy Policy 36(11):4247–4254

Salahi M, Toloo M, Torabi N (2021) A new robust optimization approach to common weights formulation in DEA. J Oper Res Soc 72(6):1390–1402

Sealey CW, Lindley JT (1977) Inputs, outputs, and a theory of production and cost at depository financial institutions. J Finance 32(4):1251–1266

Sengupta JK (1992) A fuzzy systems approach in data envelopment analysis. Comput Math Appl 24(8–9):259–266

Shokouhi AH, Hatami-Marbini A, Tavana M, Saati S (2010) A robust optimization approach for imprecise data envelopment analysis. Comput Ind Eng 59(3):387–397

Shokouhi AH, Shahriari H, Agrell PJ, Hatami-Marbini A (2014) Consistent and robust ranking in imprecise data envelopment analysis under perturbations of random subsets of data. Or Spectr 36(1):133–160
Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Manage Sci 44(1):49–61
Soyster AL (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper Res 21(5):1154–1157
Thiele A (2010) A note on issues of over-conservatism in robust optimization with cost uncertainty. Optim 59(7):1033–1040
Toloo M, Keshavarz E, Hatami-Marbini A (2021) An interval efficiency analysis with dual-role factors. Or Spectr 43:187–255
Toloo M, Mensah EK (2019) Robust optimization with nonnegative decision variables: a DEA approach. Comput Ind Eng 127:313–325
Toloo M, Mensah EK, Salahi M (2022) Robust optimization and its duality in data envelopment analysis. Omega 108:102583
Tzeremes NG (2015) Efficiency dynamics in Indian banking: a conditional directional distance approach. Eur J Oper Res 240(3):807–818
Zervopoulos PD, Triantis K, Sklavos S, Kanas A (2023) An alternative Bayesian data envelopment analysis approach for correcting bias of efficiency estimators. J Oper Res Soc 74(4):1021–1041
Zhang P, Gupta D (2023) Robust optimization with order statistic uncertainty set. Eur J Oper Res 311(3):1023–1039
Zhu J (2003) Imprecise data envelopment analysis (IDEA): a review and improvement with an application. Eur J Oper Res 144(3):513–529

Authors and Affiliations

Aliasghar Arabmaldar[1] · **Adel Hatami-Marbini**[2]

✉ Adel Hatami-Marbini
a.hatamimarbini@hud.ac.uk

Aliasghar Arabmaldar
a.arabmaldar@herts.ac.uk

[1] Hertfordshire Business School, University of Hertfordshire, Hatfield, UK

[2] School of Business, Education and Law, University of Huddersfield, Huddersfield, UK