Check for updates

# Robust data envelopment analysis via ellipsoidal uncertainty sets with application to the Italian banking industry

**Emmanuel Kwasi Mensah[1]** 📧

## Abstract

This paper extends the conventional DEA models to a robust DEA (RDEA) framework by proposing new models for evaluating the efficiency of a set of homogeneous decision-making units (DMUs) under ellipsoidal uncertainty sets. Four main contributions are made: (1) we propose new RDEA models based on two uncertainty sets: an ellipsoidal set that models unbounded and correlated uncertainties and an interval-based ellipsoidal uncertainty set that models bounded and correlated uncertainties, and study the relationship between the RDEA models of these two sets, (2) we provide a robust classification scheme where DMUs can be classified into fully robust efficient, partially robust efficient and robust inefficient, (3) the proposed models are extended to the additive DEA model and its efficacy is analyzed with two imprecise additive DEA models in the literature, and finally, (4) we apply the proposed models to study the performance of banks in the Italian banking industry. We show that few banks which were resilient in their performance can be robustly classified as partially efficient or fully efficient in an uncertain environment.

---

📧 Emmanuel Kwasi Mensah
ekmensah@uninsubria.it; kwasimensah87@gmail.com

1 Department of Economics, University of Insubria, Via Monte Generoso, 71, 21100 Varèse, VA, Italy

🙋 Springer

## 1 Introduction

Data envelopment analysis (DEA) is a nonparametric optimization model for assessing the relative performance of a set of peer decision-making units (DMU) with multiple inputs and multiple outputs. Specifically, the DEA model is based on linear programming to measure the efficiency of the $i$-th DMU under evaluation relative to the other DMUs of the set. The model was initially proposed under the constant returns to scale assumption of a firm's production activity by Charnes et al. (1978) and later extended to the variable returns to scale by Banker et al. (1984). The multiple advantages of the DEA make it an important analysis tool for benchmarking in various scientific areas such as operational research, decision analysis, management science, social science, etc. However, the learning procedure through which DMUs are benchmarked against each other implicitly assumes precision in data and ignores the uncertainties and noise inherent in the inputs and outputs. Hence, the traditional models as proposed in Charnes et al. (1978) and Banker et al. (1984) can be practically unsuitable in application. For instance, in many applications, some of the data are only known within specified bounds or in ordered relations while others are described vaguely such that the real values are unknown or uncertain. These uncertainties when neglected can affect the reliability of the efficiency scores and the stability of management decisions.

The classical approach to deal with uncertainties in management decisions is the stochastic programming and sensitivity analysis (see Land et al. 1993; Olesen and Petersen 1995; Cooper et al. 1998). To address the issue of imprecision and uncertainty which conceivably have potential feasibility and optimality concerns in DEA, many researchers adopt advanced deterministic models such as imprecise DEA, fuzzy DEA and the robust DEA (RDEA). We refer the reader to Zhu (2003) and Hatami-Marbini et al. (2011) for extensive reviews of these approaches.

The RDEA approach was proposed by Sadjadi and Omrani (2008) to deal with uncertainties in input and output data. The approach offers to immunize the uncertain inputs and outputs data of DMUs in a user-defined uncertainty set and provides a probability guarantee for reliable efficiency scores, robust discrimination and ranking of DMUs. The RDEA is based on the robust optimization (RO) technique which was initially introduced by Soyster (1973) and extended by the likes of Ben-Tal and Nemirovski (1998, 1999, 2000) and El Ghaoui et al. (1998). The theory and applications of the RO have been reviewed in Bertsimas et al. (2011) for a wider reading. A good background and review of the RO to DEA have also been provided in Mensah (2019). When applying the RO to DEA, two main techniques are enhanced to characterize the uncertainties in the input and output data: scenario-based and uncertainty set-based techniques. The former is proposed by Mulvey et al. (1995) and researched further by Laguna (1998). Its application to DEA includes the works of Zahedi-Seresht et al. (2017) and Esfandiari et al. (2017). The scenario-based RO, however, has a drawback of changing the structure of the deterministic program.

On the other hand, uncertainty set-based techniques have been developed in Ben-Tal and Nemirovski (1998, 1999, 2000), Bertsimas and Sim (2004). The general concept of this approach thrives on formulating an alternative optimization model called the robust counterpart which seeks all or most possible realization of the

uncertain parameters in decision-maker-defined uncertainty set. Broadly speaking, a major modeling concern of the RO in this direction is the design of a tractable[1] robust formulation for the nominal problem that guarantees constraints feasibility with high probability. To deal with model tractability and conservativess of robust solutions, different robust concepts: reliability of the robust solution (Ben-Tal and Nemirovski 2000), price of robustness (Bertsimas and Sim 2004), adjustable robustness (Ben-Tal et al. 2004) and light robustness (Fischetti and Monaci 2009; Mensah and Rocca 2019) among others have been proposed in the past few years. These concepts are largely based on two main sets: the ellipsoidal uncertainty sets of Ben-Tal and Nemirovski (1999, 2000) and a family of polyhedral uncertainty sets otherwise called the budget of uncertainty of Bertsimas and Sim (2004). These sets ensure model tractability and offer the decision maker the ability to control the conservativeness of the robust solution. The main disadvantage of the former, however, is that the nominal linear optimization model is transformed into a nonlinear problem which can be computational demanding.

Bertsimas and Sim (2004) model approach preserves the linearity of the nominal problem. As a result, the concept has been applied to different classes of DEA problems. For instance, Shokouhi et al. (2010) proposed a general RDEA model in which inputs and outputs are constrained in an uncertainty set with data uncertainties covering the interval DEA approach. The authors applied the robust approach of Bertsimas and Sim (2004) and Monte Carlo simulation to compute for the range of Gamma values for the conformity of the ranking of the DMUs. Omrani (2013) introduced an RDEA to find the common set of weights (CSW) in DEA with uncertain data under a similar uncertainty set. In a related paper, Salahi et al. (2016) developed an optimistic RO approach for the CSW in DEA. Arabmaldar et al. (2017) proposed a robust super-efficiency DEA model. Toloo and Mensah (2019) studied the computational complexity RDEA models in their reduced form using the concept of the budget of uncertainty. It is noteworthy that, RDEA models with ellipsoidal uncertainty seem to be relatively unexplored. To the best of our knowledge, Sadjadi and Omrani (2008), Lu (2015), Wu et al. (2017), Salahi et al. (2018) and Lu et al. (2019) are the only few researchers who have made advances to RDEA considering uncertainty in an ellipsoid. Most of these studies are, however, limited to output data uncertainty due to the larger concern of considering input data uncertainty in the equality normalization constraint. Uncertainty in the constraints of the DEA models must be strictly satisfied to obtain a feasible solution for the RDEA counterpart. The issue equality constraint in RDEA is well addressed in Toloo and Mensah (2019) and considered in this paper.

This paper will focus on the ellipsoidal uncertainty sets introduced in Ben-Tal and Nemirovski (1999, 2000) to identify inefficiencies of DMUs using the risk preference of the decision maker (DM). From the mathematical point of view, the ellipsoidal uncertainty set provides a convenient entity and offers the decision maker the ability to control the conservativeness of the efficiency solution to

---

[1] By tractability, we mean the existence of an explicit polynomial time algorithm to an equivalent formulation of the nominal optimization problem.

different data perturbations via the semi-axis of the ellipsoid. These sets are also practically useful for modeling correlation (if they exist) among the inputs (output) data which is relevant to prevent the effect of correlation on the efficiency mean (Farzipoor Sean et al. 2005). To be more specific, we adopt an ellipsoidal uncertainty set to model unbounded distribution of uncertainties while input and output data with bounded random distribution are modeled with an interval-based ellipsoidal uncertainty set. Another contribution of this paper provides a classification scheme based on the proposed models. The scheme allows DMUs to be classified into fully robust efficient, partially robust efficient and robust inefficient. We further extend our robust approach to the non-radial additive model where a newly proposed robust additive model is compared with peer imprecise additive models proposed in Lee et al. (2002) and Matin et al. (2007).

The structure of the paper is as follows. Section 2 will provide the background of the DEA and RDEA models. This is followed by RDEA models developed from the two ellipsoidal uncertainty sets in Sect. 3. A robust classification scheme and a numerical example are also given in this section. Section 4 will extend the RDEA approach to a robust additive DEA model and will compare it to some imprecise DEA models in the literature. The penultimate section illustrates the applicability of the RDEA models with banking studies in Italy. Finally, Sect. 6 will provide conclusions and further research.

## 2 Background

### 2.1 The DEA models

Consider $n$ DMUs indexed as $j = 1, \ldots, n$ where each $DMU_j$ consumes $m$ inputs $x_j = (x_{1j}, \ldots, x_{mj})$ to produce $s$ outputs $y_j = (y_{1j}, \ldots, y_{sj})$ which are denoted by $(x_j, y_j) \in \mathbb{R}^{m+s}$. Charnes et al. (1978) proposed the following fractional DEA programming to maximize the ratio of the weighted sum of outputs to the weighted sum of inputs of a unit subject to the condition that the same ratio of all other units are less than or equal to unity:

$$
\begin{aligned}
&\max \ \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\
&\text{s.t.} \\
&\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \quad j = 1, \ldots, n \\
&v_i \geq 0 \qquad\qquad i = 1, \ldots, m \\
&u_r \geq 0 \qquad\qquad r = 1, \ldots, s
\end{aligned}
\tag{1}
$$

where $u_r$ and $v_i$ are the respective $r$th output and $i$th input weights. Charnes et al. (1978) further reduced the above nonlinear CCR model to a linear form in the following:

$$\max\theta = \sum_{r=1}^{s} u_r y_{ro}$$

s.t.

$$\sum_{i=1}^{m} v_i x_{io} \leq 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \quad j = 1, \ldots, n \qquad (2)$$

$$v_i \geq 0 \qquad\qquad\qquad i = 1, \ldots, m$$

$$u_r \geq 0 \qquad\qquad\qquad r = 1, \ldots, s$$

Model (2) involves $n+1$ and $m+s$ decision variables (weights) with an objective function that estimates the efficiency of DMU$_o$ at $n$ solution instance. It must be noted here that the "$\leq$" sign is used for the normalization constraint rather than the usual "$=$" sign used in the case of the standard CCR model. The consideration of the inequality sign addresses the issue of equality constraint in RDEA. In other words, model (2) overcomes the situation where input uncertainty and robust analysis in the standard normalization constraint $\sum_{i=1}^{m} v_i x_{io} = 1$ could lead to a restriction on the constraint and probable model infeasibility (Ben-Tal et al. 2009). For details on equality constraint in RDEA, see Toloo and Mensah (2019). It is worth mentioning also that the CCR model with either "$\leq$" or "$=$" sign in constraint 1 yields an equivalent efficiency for the DMUs (see Toloo 2014). Thus, suppose the optimal solution in model (2) is $(\boldsymbol{u}^*, \boldsymbol{v}^*) = \left(u_1^*, \ldots u_s^*; v_1^*, \ldots v_m^*\right)$. The efficiency of DMUs is given by the following definition.

**Definition 1** DMU$_o$ is CCR efficient if $\theta^* = 1$, and there exists at least one strictly positive optimal solution (i.e., $\forall i, v_i^* > 0, \forall r, u_r^* > 0$), otherwise it is CCR inefficient.

## 2.2 Robust counterpart DEA

Model (2) shows the case where the inputs and outputs are deterministic, i.e., nominal data are used. As aforementioned, the model is not useful when uncertain factors in the inputs and outputs prevail. In this section, we employ the RO technique and develop RDEA model to address this nondeterministic data problem.

The common technique of the RDEA is to consider the *worst-case scenario* for the uncertain input and output data and trade-off between performance and robustness as different scenarios occur. Let the input and output variables with uncertainty be expressed as $\tilde{x}_{ij} = x_{ij} + \xi_{ij}^x \hat{x}_{ij}$; $\tilde{y}_{rj} = y_{rj} + \xi_{rj}^y \hat{y}_{rj}$ where $\hat{x}_{ij} = \delta^x x_{ij}, \hat{y}_{rj} = \delta^y y_{rj}$ are deviations from the nominal values, $x_{ij}, y_{rj}$ and $\delta^x$ and $\delta^y$ are a given uncertainty level or percentage of perturbation, respectively. By definition, DMU$_k$ with $(\tilde{\boldsymbol{x}}_k, \tilde{\boldsymbol{y}}_k)$ is uncertain if there exist $i \in I_k$ or $r \in R_k$ where $I_j$ and $R_j$ represent the set of inputs and outputs of DMU$_j$ that are subject to uncertainty. That is $I_j = \emptyset$ and $R_j = \emptyset$ present the case where there is no uncertainty. All the uncertainties are subject to a known set $\mathcal{U}$ called the uncertainty set. Therefore, considering the uncertain inputs and outputs set, we could impose constraints on the uncertain DEA model such as $\tilde{x}_{ij}, \tilde{y}_{rj} \in \mathcal{U}$. The robust counterpart to the uncertain DEA will be the following:

$$\max \theta$$
$$\text{s.t.}$$
$$\theta - \sum_{r=1}^{s} u_r \tilde{y}_{ro} \leq 0$$
$$\sum_{i=1}^{m} v_i \tilde{x}_{io} \leq 1$$
$$\sum_{r=1}^{s} u_r \tilde{y}_{rj} - \sum_{i=1}^{m} v_i \tilde{x}_{ij} \leq 0 \; j = 1, \ldots, n \qquad (3)$$
$$v_i \geq 0 \qquad\qquad i = 1, \ldots, m$$
$$u_r \geq 0 \qquad\qquad r = 1, \ldots, s$$
$$\forall (\tilde{x}, \tilde{y}) \in \mathcal{U}$$

and

$$\mathcal{U} = \left\{ \Phi(\xi) | \parallel \mathbf{P}\xi \parallel_2 \leq \Omega \right\}$$

where $\mathcal{U}$ is 'ellipsoidal'; $\mathbf{P}$ is a non-singular matrix of perturbations and $\Omega$ is the safety or robust parameter defined by the DM. In this formulation $\theta > 0$ holds since the weight variables are assumed to be nonnegativity. Thus, the objective in model (2) is converted to a constraint in model (3) which is related to the uncertain data for the DMU$_o$ to be evaluated. This additional constraint is free in index since it is determined at the outset and is in contact with the objective in the model.

The ellipsoid uncertainty set is considered for two main reasons: first, to adjust the risk tolerance of the DM by controlling the size of the ellipsoid via the parameter $\Omega$, i.e., as the size of the ellipsoid increases, the risk aversion of the DM increases and vice versa. Second, to overcome the aggressive conservatism of the robust solution (c.f Soyster (1973). It should be mentioned here that the robust counterpart for the ellipsoidal sets is nonlinear, however, its formulation is tractable. In other words, the robust counterpart leads to second-order quadratic programming which can be solved once in polynomial time with many solution algorithms including solvers such as GUROBI in GAMS.

## 3 RDEA models under ellipsoidal uncertainty sets

We distinguish between two kinds of uncertainty including randomness inherent in the inputs and outputs data: unbounded and bounded correlated uncertainties. We consider the following uncertainty sets:

1. the usual ellipsoidal uncertainty set
2. box (interval)-based ellipsoidal uncertainty set.

The first set models uncertainties that have unbounded distribution while the second set models uncertainties with bounded random distribution. Below, we discuss in detail these uncertainty sets and their RDEA formulation.
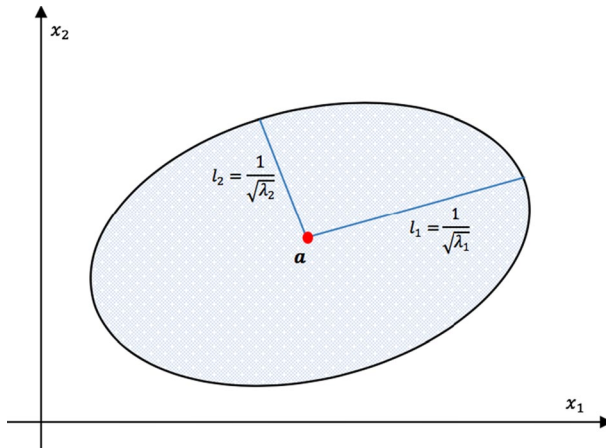
**Fig. 1** The geometry of an ellipsoid in $\mathbb{R}^2$

## 3.1 The usual ellipsoid case

Let's begin with the simplest case where $\mathcal{U}_e$ is a usual ellipsoid or a constraint-wise uncertainty with every constraint uncertainty set $\mathcal{U}_e$ being an ellipsoid. We describe the following ellipsoid

$$\mathcal{U}_e = \left\{ \boldsymbol{a}_o + \mathbf{P}\boldsymbol{u} \mid \| \boldsymbol{u} \|_2 \leq 1 \right\} \tag{4}$$

where the vector $\boldsymbol{a}_o \in \mathbb{R}^n$ is the center of the ellipsoid, Rank $(\mathbf{P}) = m \leq n$ is the *shape matrix* of the ellipsoid and the random variable $\boldsymbol{u} \in \mathbb{R}^n$. The representation above can handle different cases of the ellipsoid including "ellipsoidal cylinders" and "flat" ellipsoids such as points and intervals (Calafiore and El Ghaoui 2004). An alternative description involves the squared shape matrix $\mathbf{A} = \mathbf{P}\mathbf{P}^{\mathrm{T}}$ for which, when $\mathbf{A} > 0$ we obtain an equivalent representation of (4) as $\mathcal{U}(\boldsymbol{a}_o, \mathbf{A}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \boldsymbol{a}_o)^{\mathrm{T}} \mathbf{A}^{-1} (\boldsymbol{x} - \boldsymbol{a}_o) \leq 1 \right\}$. Figure 1 shows such an ellipsoid in $\mathbb{R}^2$ (shaded) with center $\boldsymbol{a}_o$ and axis—length $l_i$; $\lambda_i = \left( \frac{1}{l_i} \right)^2$ in the direction $\boldsymbol{\pi}_i$ where $\lambda_i$ and $\boldsymbol{\pi}_i$ are, respectively, the eigenvalues and eigenvectors corresponding to the symmetric positive definite matrix $\mathbf{A}^2$

Let $\quad P^x = \left[ \bar{\rho}^x_{ij} \right] \in \mathbb{R}^{m \times n}, \quad P^y = \left[ \bar{\rho}^y_{rj} \right] \in \mathbb{R}^{s \times n}, \quad \Xi^x = \left[ \bar{\xi}^x_{ij} \right] \in \mathbb{R}^{m \times n}, \quad$ and $\Xi^x = \left[ \bar{\xi}^y_{rj} \right] \in \mathbb{R}^{s \times n} \quad$ where $\quad \bar{\rho}^x_{ij} = \begin{cases} \rho^x_{ij} & \text{if } i \in I_j \\ 0 & \text{otherwise} \end{cases}, \quad \bar{\rho}^y_{rj} = \begin{cases} \rho^y_{rj} & \text{if } r \in R_j \\ 0 & \text{otherwise} \end{cases},$

---

[2] Notice that, $\mathbf{A}$ is symmetric positive definite, and we can obtain the real eigenvalues of the matrix using the Cholesky decomposition. The eigenvector $\boldsymbol{\pi}_i$ with eigenvalue $\lambda_i$ represent the orientations of the principal axes of the ellipsoid. That is geometrically, $\boldsymbol{\pi}_i$ is the axis-vectors of the ellipsoid since it shows the direction of the $i$th axis of the ellipsoid.

$\bar{\xi}_{ij}^{x} = \begin{cases} \xi_{ij}^{x} & \text{if } i \in I_j \\ 0 & \text{otherwise} \end{cases}$, $\bar{\xi}_{rj}^{y} = \begin{cases} \xi_{rj}^{y} & \text{if } r \in R_j \\ 0 & \text{otherwise} \end{cases}$. Then, for all DMUs, the simple ellipsoid where $(x, y) \in \mathbb{R}^{m+s}$ is described as follows:

$$
\begin{aligned}
\mathcal{U}_j^e &= \left\{ (\tilde{X}, \tilde{Y}) \left| \begin{array}{l} \tilde{X} = (X + P^x \Xi^x), \| \xi_j^x \|_2 \leq 1 \\ \tilde{Y} = (Y + P^y \Xi^y), \| \xi_j^y \|_2 \leq 1 \end{array} ; \quad j = 1, \ldots, n \right. \right\} \subset \mathbb{R}^{n \times (m+s)} \\
&= \left\{ (\tilde{x}, \tilde{y}) \left| \begin{array}{l} \tilde{x}_{ij} = x_{ij} + \sum_{i \in I_j} \rho_{ij}^x \xi_{ij}^x, \quad \| \xi_j^x \|_2 \leq 1 \\ \tilde{y}_{rj} = y_{rj} + \sum_{r \in R_j} \rho_{rj}^y \xi_{rj}^y, \quad \| \xi_j^y \|_2 \leq 1 \end{array} ; j = 1, \ldots, n \right. \right\}
\end{aligned}
\tag{5}
$$

where $\rho_j^x (= \hat{x}_{ij})$ and $\rho_j^y (= \hat{y}_{rj})$ are deviation vectors defining the deviation of inputs (outputs) from their nominal values. Following Wu et al. (2017), the weight vectors $u$ or $v$ is, respectively, mapped by the following relationships:

$$
\xi_j^x = \frac{(\rho_j^x)^T v}{\| (\rho_j^x)^T v \|_2}, \quad \xi_j^y = \frac{(\rho_j^y)^T u}{\| (\rho_j^y)^T u \|_2}
$$

To formulate the robust counterpart of the DEA under the uncertainty (4), the following lemma on the worst-case robust counterpart is important.

**Lemma 1** *Consider the linear inequality $a_o^T x \leq b_i$ where the vector a is uncertain and belongs to the ellipsoid $\mathcal{U}_e = \{ a_o + Pu | \| u_2 \|_2 \leq 1 \}$. Then its robust counterpart is $a_o^T x + \| Px \|_2 \leq b_i$*

**Proof** See for e.g., Bertsimas et al. (2011)

**Theorem 1** *The robust counterpart CCR described under the ellipsoidal set (5) lead to the following nonlinear model:*

$$
\begin{aligned}
& \max w \\
& \text{s.t.} \\
& w - \sum_{r=1}^{s} u_r y_{ro} + \sqrt{\sum_{r \in R_o} u_r^2 \hat{y}_{ro}^2} \leq 0 \\
& \sum_{i=1}^{m} v_i x_{io} + \sqrt{\sum_{i \in I_o} v_i^2 \hat{x}_{io}^2} \leq 1 \\
& \sum_{r=1}^{s} u_r y_{ro} - \sum_{i=1}^{m} v_i x_{io} - \sqrt{\sum_{r \in R_o} u_r^2 \hat{y}_{ro}^2} - \sqrt{\sum_{i \in I_o} v_i^2 \hat{x}_{io}^2} \leq 0 \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \sqrt{\sum_{r \in R_j} u_r^2 \hat{y}_{rj}^2} + \sqrt{\sum_{i \in I_j} v_i^2 \hat{x}_{ij}^2} \leq 0 \quad \forall j \neq o \\
& v_i \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad \forall i \\
& u_r \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad \forall r
\end{aligned}
\tag{6}
$$

**Proof** Using Lemma 1 and the CCR model 1, the robust counterpart DEA with the ellipsoid uncertainty set (5) is formulated first as the following model:

$$
\max w
$$
$$
\text{s.t.}
$$
$$
w - \frac{\sum_{r=1}^{s} u_r y_{ro} + \inf_{\|\xi_o^y\|_2 \leq 1} \left( \sum_{r \in R_o} u_r \xi_{ro}^y \rho_{ro}^y \right)}{\sum_{i=1}^{m} v_i x_{io} + \sup_{\|\xi_o^x\|_2 \leq 1} \left( \sum_{i \in I_o} v_i \xi_{io}^x \rho_{io}^x \right)} \leq 0
$$
$$
\frac{\sum_{r=1}^{s} u_r y_{ro} + \inf_{\|\xi_o^y\|_2 \leq 1} \left\{ \sum_{r \in R_o} u_r \xi_{ro}^y \rho_{ro}^y \right\}}{\sum_{i=1}^{m} v_i x_{io} + \sup_{\|\xi_o^x\|_2 \leq 1} \left\{ \sum_{i \in I_o} v_i \xi_{io}^x \rho_{io}^x \right\}} \leq 1 \qquad (7)
$$
$$
\frac{\sum_{r=1}^{s} u_r y_{rj} + \sup_{\|\xi_j^y\|_2 \leq 1} \left\{ \sum_{r \in R_j} u_r \xi_{rj}^y \rho_{rj}^y \right\}}{\sum_{i=1}^{m} v_i x_{ij} + \inf_{\|\xi_j^x\|_2 \leq 1} \left\{ \sum_{i \in I_j} v_i \xi_{ij}^x \rho_{ij}^x \right\}} \leq 1 \qquad \forall j \neq o
$$
$$
v_i \geq 0 \qquad\qquad\qquad \forall i
$$
$$
u_r \geq 0 \qquad\qquad\qquad \forall r
$$

In this application, the $n$ constraints in model (1) are transformed into one and $n-1$ constraints. This is to ensure that the maximal efficiency of DM $U_o$ does not exceed unity but attains it at optimum, given also that the set of data used in the first constraint of model (7) are reused in the second constraint. Model (7) is further modified from the fractional program to linear form by employing Charnes and Cooper (1962) transformation: Denote $t = \dfrac{1}{\sum_{i=1}^{m} v_i x_{io} + \sup_{\|\xi_o^x\|_2 \leq 1} \left( \sum_{i \in I_o} v_i \xi_{io}^x \rho_{io}^x \right)} > 0$ and $u_r = tu_r, v_i = tv_i$ in the first constraint. Consequently, the following DEA model is achieved:

$$
\max w
$$
$$
\text{s.t.}
$$
$$
w - \sum_{r=1}^{s} u_r y_{ro} + \inf_{\|\xi_o^y\|_2 \leq 1} \left( \sum_{r \in R_o} u_r \xi_{ro}^y \rho_{ro}^y \right) \leq 0
$$
$$
\sum_{i=1}^{m} v_i x_{io} + \sup_{\|\xi_o^x\|_2 \leq 1} \left( \sum_{i \in I_o} v_i \xi_{io}^x \rho_{io}^x \right) \leq 1
$$
$$
\left( \sum_{r=1}^{s} u_r y_{ro} + \inf_{\|\xi_o^y\|_2 \leq 1} \left\{ \sum_{r \in R_o} u_r \xi_{ro}^y \rho_{ro}^y \right\} \right) - \left( \sum_{i=1}^{m} v_i x_{io} + \sup_{\|\xi_o^x\|_2 \leq 1} \left\{ \sum_{i \in I_o} v_i \xi_{io}^x \rho_{io}^x \right\} \right) \leq 0
$$
$$
\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \sup_{\|\xi_j^y\|_2 \leq 1} \sup_{\|\xi_j^x\|_2 \leq 1} \left\{ \sum_{r \in R_j} u_r \xi_{rj}^y \rho_{rj}^y + \sum_{i \in I_j} v_i \xi_{ij}^x \rho_{ij}^x \right\} \leq 0 \qquad \forall j \neq o
$$
$$
v_i \geq 0 \qquad\qquad\qquad \forall i
$$
$$
u_r \geq 0 \qquad\qquad\qquad \forall r
$$
$$
(8)
$$

Next, we solve the inner problems in model (8). The last term of the first constraint in a re-casted objective function form arrives at the following robust counterpart:

$$\max_{\boldsymbol{u}\in\mathbb{R}^s} \inf_{\tilde{\boldsymbol{y}}_o\in\mathcal{U}_e} \left\{ \sum_{r=1}^{s} u_r y_{ro} + \sum_{r\in R_o} u_r \xi_{ro}^y \rho_{ro}^y \right\} = \max_{\boldsymbol{u}\in\mathbb{R}^s} \left\{ \sum_{r=1}^{s} u_r y_{ro} + \inf_{|\xi_o^y|_2\leq 1} (\boldsymbol{\xi}_o^y)^{\mathrm{T}} (\boldsymbol{\rho}_o^y)^{\mathrm{T}} \boldsymbol{u} \right\}$$

$$= \max_{\boldsymbol{u}\in\mathbb{R}^s} \left\{ \sum_{r=1}^{s} u_r y_{ro} - \frac{\boldsymbol{u}^{\mathrm{T}} \boldsymbol{\rho}_o^y (\boldsymbol{\rho}_o^y)^{\mathrm{T}} \boldsymbol{u}}{\| (\boldsymbol{\rho}_o^y)^T \boldsymbol{u} \|_2} \right\}$$

$$= \max_{\boldsymbol{u}\in\mathbb{R}^s} \left\{ \sum_{r=1}^{s} u_r y_{ro} - \frac{\| (\boldsymbol{\rho}_o^y)^T \boldsymbol{u} \|_2^2}{\| (\boldsymbol{\rho}_o^y)^T \boldsymbol{u} \|_2} \right\}$$

$$= \max_{\boldsymbol{u}\in\mathbb{R}^s} \left\{ \sum_{r=1}^{s} u_r y_{ro} - \| (\boldsymbol{\rho}_o^y)^{\mathrm{T}} \boldsymbol{u} \|_2 \right\}$$

(9)

The robust counterpart for the second term of the second constraint is:

$$\sup_{\tilde{\boldsymbol{x}}_o\in\mathcal{U}_e} \left\{ \sum_{i=1}^{m} v_i x_{io} + \sum_{i\in I_o} v_i \xi_{io}^x \rho_{io}^x \right\} = \sup_{\xi_{o2}^x\leq 1} \left\{ \sum_{i=1}^{m} v_i x_{io} + (\boldsymbol{\xi}_o^x)^{\mathrm{T}} (\boldsymbol{\rho}_o^x)^{\mathrm{T}} \boldsymbol{v} \right\}$$

$$= \sum_{i=1}^{m} v_i x_{io} + \| (\boldsymbol{\rho}_o^x)^{\mathrm{T}} \boldsymbol{v} \|_2$$

(10)

The robust counterpart for the third term of the fourth constraint is:

$$\sup_{\tilde{\boldsymbol{y}}_j\in\mathcal{U}_e} \sup_{\tilde{\boldsymbol{x}}_j\in\mathcal{U}_e} \left\{ \sum_{r=1}^{s} u_r y_{rj} + \sum_{r\in R_j} u_r \xi_{rj}^y \rho_{rj}^y - \sum_{i=1}^{m} v_i x_{ij} + \sum_{i\in I_j} v_i \xi_{ij}^x \rho_{ij}^x \right\}$$

$$= \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \sup_{|\xi_j^y|_2\leq 1} \sup_{|\xi_j^x|_2\leq 1} \left\{ (\boldsymbol{\xi}_j^y)^{\mathrm{T}} (\boldsymbol{\rho}_j^y)^{\mathrm{T}} \boldsymbol{u} + (\boldsymbol{\xi}_j^x)^{\mathrm{T}} (\boldsymbol{\rho}_j^x)^{\mathrm{T}} \boldsymbol{v} \right\}$$

(11)

$$= \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \| (\boldsymbol{\rho}_j^y)^{\mathrm{T}} \boldsymbol{u} \|_2 + \| (\boldsymbol{\rho}_j^x)^{\mathrm{T}} \boldsymbol{v} \|_2$$

Finally, we write $\| \boldsymbol{\rho}_j^x \boldsymbol{v} \|_2 = \sqrt{\sum_{i\in I_j} v_i^2 \hat{x}_{ij}^2}$ for $i \in I_j$ and $\| \boldsymbol{\rho}_j^y \boldsymbol{u} \|_2 = \sqrt{\sum_{r\in R_j} u_r^2 \hat{y}_{rj}^2}$ for $r \in R_j$ and substitute the results from (9) to (11) into model (8) and the proof is complete. □

## 3.2 The combined interval and ellipsoid case

We now look at the uncertainty set designed with an ellipsoid and interval uncertainties. We assume that the uncertain input and outputs are obtained from the nominal values by the random perturbation: $\tilde{x}_{ij} = \left(1 + \delta^x \zeta_{ij}^x\right) x_{ij}$ and $\tilde{y}_{rj} = \left(1 + \delta^y \zeta_{rj}^y\right) y_{rj}$ where $\left\{\zeta_{ij}^x\right\}_{i\in I_j}$ and $\left\{\zeta_{rj}^y\right\}_{r\in R_j}$ ($\zeta_{ij}^x = \zeta_{rj}^y = 0$ for $i \notin I_j, r \notin R_j$) are the independent random variables symmetrically distributed in the interval bound $[-1, 1]$ and $\delta^x$ and
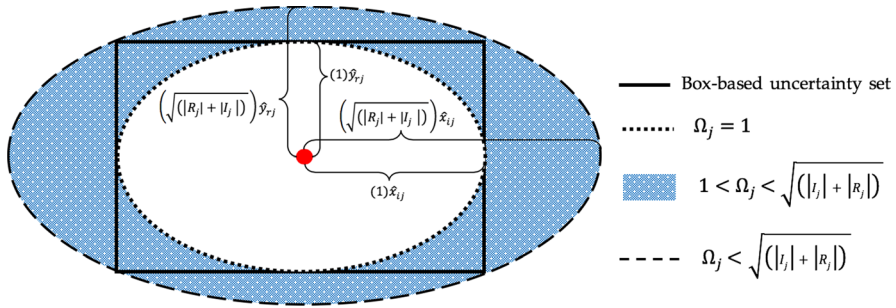
**Fig. 2** Illustration of feasible region for varying $\Omega_j$ values

$\delta^y$ are given uncertainty levels of the inputs and outputs (Ben-Tal and Nemirovski 2000). In this situation, we speak of an RDEA which pass from deterministic to probabilistic in the sense that the underlying inputs and output variable are purely random and the probability of the constraint $\sum_{i=1}^{m} v_i \tilde{x}_{io} \leq 1$, for instance, is $\hbar$ where $\hbar \geq 0$ is a given reliability level. For the uncertainty set here, we proceed with the following useful definitions.

**Definition 2** Consider a unitary interval denoted by $\mathbf{B} = [-1, 1]$. An interval uncertainty set for the random variables $\boldsymbol{\zeta}$ is equipped with the infinity norm given as $\mathcal{U}_{\text{int}} = \left\{ \boldsymbol{\zeta} \in \mathbf{R}^{m \times s} : \| \boldsymbol{\zeta} \|_{\infty} \leq 1 \right\}$.

**Definition 3** Given the random variables $\boldsymbol{\zeta}$, the ellipsoid normalized to a ball of radius $\Omega$ centered at the origin is the set with $l_2$-norm given as $\mathcal{U}_{\text{ell}} = \left\{ \boldsymbol{\zeta} \in \mathbf{R}^{m \times s} : \| \boldsymbol{\zeta} \|_2 \leq \Omega \right\}$.

The uncertainty set for the uncertain input and output dynamics using both $l_2$-norm and the infinity norm from the above definitions is stated as the following:

$$
\mathcal{U}_j^{\Omega} = \left\{ (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \middle| \begin{array}{l} \tilde{x}_{ij} = \left(1 + \delta^x \zeta_{ij}^x\right) x_{ij}, \ \left\| \zeta_j^x \right\|_{\infty} \leq 1, \left\| \zeta_j^x \right\|_2 \leq \Omega_j^x \\ \tilde{y}_{rj} = \left(1 + \delta^y \zeta_{rj}^y\right) y_{rj} \ \left\| \zeta_j^y \right\|_{\infty} \leq 1, \left\| \zeta_j^y \right\|_2 \leq \Omega_j^y \end{array} ; \quad j = 1, \dots, n \right\}
$$

(12)

where $\Omega_j^x$ and $\Omega_j^y$ are the lengths of semi-axes of the ellipsoid for the uncertain input and output data, respectively. Let $\Omega_j = \Omega_j^x + \Omega_j^y$, where $\Omega_j \leq \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5}$ depicts the allowable conservative preference of the DM and $\left| I_j \right|$ and $\left| R_j \right|$ are the cardinalities of the uncertain inputs and outputs, respectively. It is clear that $\mathcal{U}_j^{\Omega}$ is an intersection of unit boxes and balls centered at the origin with radii $\Omega_j^x$ and $\Omega_j^y$. The largest volume ellipsoid contained in the box occurs when $\Omega_j = 1$ and the smallest volume ellipsoid containing the box occurs when $\Omega_j = \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5}$. Figure 2 illustrates the different scenarios of the feasible region for the ellipsoid intersection with the

box which is adapted similarly from Hanks et al. (2017). Here, while it is possible to set $\Omega_j > \left(\left|R_j\right| + \left|I_j\right|\right)^{0.5}$ for the uncertain input and output data, wlog, we consider only the case where $0 \leq \Omega_j \leq \left(\left|R_j\right| + \left|I_j\right|\right)^{0.5}$. The following lemma is crucial for the robust counterpart DEA model:

**Lemma 2** *The constraint in the uncertain optimization problem* $\min\left\{cx : a^{\mathrm{T}}x + \max_{\zeta \in \mathbb{U}} \zeta^{\mathrm{T}}x \leq b_i\right\}, \mathbb{U} = \left\{\zeta : \left\|\zeta_j\right\|_{\infty} \leq 1, \left\|\zeta_j\right\|_2 \leq \Omega, \forall j \in J_i\right\}$ *has a second-order cone constraint* $\sum_{j=1}^{n} a_j x_j + \sum_{j \in J_i} \hat{a}_{ij}\left|x_j - z_{ij}\right| + \Omega\sqrt{\sum_{j \in J_i} \hat{a}_{ij}^2 z_{ij}^2} \leq b_i$

**Proof** See for e.g., Ben-Tal and Nemirovski (2000).

Consequently, the RDEA under the uncertainty set (10) leads to the following second-order cone programming CCR model.

**Theorem 2** *The robust counterpart CCR model described under the ellipsoidal set (13) leads to the following second-order cone programming DEA model:*

$$
\begin{aligned}
&\max z \\
&\text{s.t.} \\
&z - \sum_{r=1}^{s} u_r y_{ro} + \sum_{r \in R_o} \mu_r \hat{y}_{ro} + \Omega_o^y \sqrt{\sum_{r \in R_o} v_r^2 \hat{y}_{ro}^2} \leq 0 \\
&\sum_{i=1}^{m} v_i x_{io} + \sum_{i \in I_o} \lambda_{io} \hat{x}_{io} + \Omega_o^x \sqrt{\sum_{i \in I_o} v_{io}^2 \hat{x}_{io}^2} \leq 1 \\
&\sum_{r=1}^{s} u_r y_{ro} - \sum_{i=1}^{m} v_i x_{io} - \sum_{r \in R_o} \mu_r \hat{y}_{ro} - \sum_{i \in I_o} \lambda_{io} \hat{x}_{io} \\
&\quad -\Omega_o^y \sqrt{\sum_{r \in R_o} v_{ro}^2 \hat{y}_{ro}^2} - \Omega_o^x \sqrt{\sum_{i \in I_o} v_{io}^2 \hat{x}_{io}^2} \leq 0 \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \sum_{r \in R_j} \mu_r \hat{y}_{rj} + \sum_{i \in I_j} \lambda_{ij} \hat{x}_{ij} \\
&\quad +\Omega_j^y \sqrt{\sum_{r \in R_j} v_{rj}^2 \hat{y}_{rj}^2} + \Omega_j^x \sqrt{\sum_{i \in I_j} v_{ij}^2 \hat{x}_{ij}^2} \leq 0 \qquad j \neq o \\
&-\mu_{rj} \leq u_r - v_{rj} \leq \mu_{rj} && \forall r \in R_j \\
&-\lambda_{ij} \leq v_i - v_{ij} \leq \lambda_{ij} && \forall i \in I_j \\
&v_i \geq 0 && \forall i \\
&u_r \geq 0 && \forall r \\
&\lambda_{ij}, \mu_{rj} \geq 0 && \forall i \in I_j, \forall r \in R_j
\end{aligned}
\tag{13}
$$

where $\mu_{rj}$ and $\lambda_{ij}$ are auxiliary output and input variables; $\upsilon_{rj}$ and $\nu_{ij}$ are interval uncertainty parameters.[3] The robust counterpart model is feasible with the same probability as the original problem if all the constraints are satisfied with probability guarantee $\hbar = \exp{-(\Omega/2)}$.

**Proof** The proof follows similarly from Theorem 1.

### 3.3 The efficiency of the RDEA models

The design of uncertainty set for the two models developed in this paper is related to the distribution of the uncertainty. Thus, if the uncertainty of the inputs and outputs is subject to unbounded distribution, i.e., the size of the uncertainty is not restricted, the simple ellipsoid is considered appropriate and model (6) is constructed. On the other hand, to subject the uncertainty to a bounded distribution, the interval set is required to limit the uncertainty in their bounds to avoid an unnecessarily large uncertainty set. The combined interval and ellipsoid uncertainty set is used for model (13). The formulation of these RDEA models are tractable, feasible and their robust efficiency is obtained according to the following theorems:

**Theorem 3** *The optimal objective values of model* (6) *is less than or equal to 1.*

**Proof** Let $(w^*, \boldsymbol{v}^*, \boldsymbol{u}^*)$ be the optimal solution of model (6). We have $w^* \leq \sum_{r=1}^{s} u_r^* y_{ro} - \sqrt{\sum_{r \in R_o} u_r^{*2} \hat{y}_{ro}^2}$ according to the first constraint of model (6). For every $(x_{ij}, y_{rj}), i \in I_o, r \in R_o,$ we have $\sum_{r=1}^{s} u_r^* y_{ro} - \sqrt{\sum_{r \in R_o} u_r^{*2} \hat{y}_{ro}^2} \leq \sum_{r=1}^{s} u_r^* y_{ro} + \sqrt{\sum_{r \in R_o} u_r^{*2} \hat{y}_{ro}^2}$ and since $\sqrt{\sum_{r \in R_o} v_r^{*2} \hat{y}_{ro}^2}$ is nonnegative, taking the second and third sets of constraints arrive at $\sum_{r=1}^{s} u_r y_{ro} + \sqrt{\sum_{r \in R_o} v_r^{*2} \hat{y}_{ro}^2} \leq \sum_{i=1}^{m} v_i^* x_{io} + \sqrt{\sum_{i \in I_o} v_i^{*2} \hat{x}_{io}^2} \leq 1$ for each $j = o$. Consequently, $w^* \leq 1$ which completes the proof. □

**Theorem 4** *The optimal objective values of model* (13) *is less than or equal to 1.*

**Proof** The proof is similar to Theorem 3.

Let $w^*$ and $z^*$ be the optimal objective value of models (6) and (13), respectively. The robust efficiency for $\text{DMU}_o$ is given by the following definition.

**Definition 4** $\text{DMU}_o$ is $R$-efficient, if and only if it satisfies the following two conditions:

---

[3] Note that, the RDEA to the ellipsoidal uncertainty set is practically tractable and convenient to handle with nonlinear solvers (such as Gurobi, MOSEK, BARON) and any efficient optimization software (see Sadjadi and Omrani 2008; Wu et al. 2017; Shakouri et al. 2019).

(i)  it is CCR efficient and

(ii)  $w^* = 1$ or $z^* = 1$.

The efficiency here is also referred to strong efficiency or fully robust efficiency in the sense of Pareto–Koopmans efficiency as opposed to Farrell's measure of efficiency, which ignores the slacks as sources of inefficiency (see Cooper et al. 1999; Park 2007). We shall incorporate other existing concepts of the efficiency classifications later in this section. From Theorems 3 and 4, it is clear that $w^*, z^* \in (0, 1]$. Therefore, a CCR—non efficient DMU will be robust—inefficient irrespective of the uncertainty level.

Note that the solution $\left(z^*, \boldsymbol{v}^*, \boldsymbol{u}^*, \lambda^*, \boldsymbol{\mu}^*\right)$ in model (13) can be 'less conservative' than the solution $(w^*, \boldsymbol{v}^*, \boldsymbol{u}^*)$ in model (6): if $\Omega_j$ increases in the former. In fact, in the case of large uncertainty set indicating high assurance for robustness, the efficiency of DMUs in model (13) can be less efficient than in model (6). Indeed, the second term in the constraint model (13): $z - \sum_{r=1}^{s} u_r y_{ro} + \sum_{r \in R_o} \mu_r \hat{y}_{ro} + \Omega_o^y \sqrt{\sum_{r \in R_o} v_r^2 \hat{y}_{ro}^2} \leq 0$ enforces protection of the random perturbation of the inputs and outputs in the interval which makes the model much restrictive for higher efficiency. In the special case of $\Omega_j = 1$ (see Fig. 2), the ellipsoid is exactly inscribed by the box/interval, and so we obtain an equivalency in the optimal objective values of model (6) and model (13). For higher values of $\Omega_j$, the performance of DMUs worsen indicating the price paid for robustness (Bertsimas and Sim 2004). As a result, DMUs characterized as inefficient by the latter are equally characterized as inefficient by the former. However, the reverse case is not entirely true.

As observed so far, the efficiency of DMUs under the robust model (13) depends on the risk preference of the DM determined by the uncertainty level. That is, the DM is at will to vary $\Omega_j$ according to the following:

- $\Omega_j = 0 \Rightarrow$ the robust model shrinks to the nominal DEA problem.
- $\Omega_j = 1 \Rightarrow$ the uncertainty denotes the largest volume of ellipsoid contained in the interval and
- $\Omega_j = \left(\left|R_j\right| + \left|I_j\right|\right)^{0.5} \Rightarrow$ the highest robust solution is sought for the uncertain inputs and outputs in the model since all the uncertain inputs and outputs are immunized.

The specific value of $\Omega_j$ to the model is carefully chosen to avoid an overly conservative solution. Here, we provide a suggestion for the classification of DMUs using the conservativeness of the DM, i.e., $\Omega_j \leq \left(\left|R_j\right| + \left|I_j\right|\right)^{0.5}$.

**Definition 5** The robust efficiency for DMUs under the robust model (13) can be classified into three mutually exclusive subsets:

(i)  (*Full R-efficiency*). DMU$_o$ is fully $R$-efficient if and only if $z^* = 1$ when $\Omega_j = \left(\left|R_j\right| + \left|I_j\right|\right)^{0.5}, \forall j.$

**Table 1** Data for uncertainty analysis

| DMU | Input | | | | Output | | | |
|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $y_1$ | $y_2$ | $y_3$ | $y_4$ |
| 1 | 32 | 50 | 82 | 46 | 47 | 93 | 54 | 65 |
| 2 | 61 | 56 | 68 | 37 | 88 | 56 | 92 | 80 |
| 3 | 42 | 58 | 45 | 34 | 94 | 65 | 80 | 80 |
| 4 | 73 | 39 | 88 | 81 | 50 | 53 | 93 | 97 |
| 5 | 45 | 38 | 68 | 41 | 47 | 42 | 70 | 52 |
| 6 | 86 | 62 | 44 | 32 | 86 | 45 | 100 | 47 |
| 7 | 38 | 74 | 71 | 74 | 83 | 91 | 62 | 74 |
| 8 | 61 | 54 | 70 | 62 | 79 | 60 | 72 | 98 |
| 9 | 84 | 52 | 38 | 47 | 85 | 68 | 51 | 41 |
| 10 | 87 | 47 | 31 | 52 | 78 | 95 | 70 | 92 |

(ii) (*Partial R-efficiency*). $DMU_o$ is partially *R*-efficient if $z^* < 1$ when $\Omega_j = \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5}$ and there exist $\Omega_j > 0$ such that $z^* = 1$

(iii) (*R-inefficiency*). A $DMU_o$ is said to be *R*-inefficient if $z^* < 1$ for all $\Omega_j \in \left( 0, \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5} \right]$.

The above classification can be denoted as RE$^{++}$ ~ full R-efficiency, RE$^+$ ~ partial robust efficiency or PR-efficiency and RE$^-$ ~ *R*-inefficiency. The set RE$^{++}$ consist of DMUs that are robust R-efficient in any combination of uncertain inputs and outputs at all robust levels defined by the DM. This category of efficient DMUs is obtained under the most conservative evaluation of the uncertain data. It is evident that any DMU in the fully robust efficient group is always efficient uncertain data and can hence be regarded as the best performer. So, logically, a DMU is robust efficient if and only if it is fully robust efficiency. The set RE$^+$ consists of DMUs that are *R*-efficient at maximal sense but cannot maintain R-efficiency at certain conservative levels for inputs and outputs. The PR-efficiency is therefore obtained in a less stringent manner than the full robust efficiency and as a result, it's efficiency values are higher than the *R*-efficient units. Finally, the set RE$^-$ consists of *R*-inefficient DMUs which are always inefficient in the least consideration of uncertainty for any input and output combinations. It is therefore clear that $DMU_o$ cannot be efficient in the robust sense unless this DMU is partially robust efficient and that $DMU_o$ will be partially robust efficient if it is first DEA efficient.

To further demonstrate the classification scheme provided in Definition 5 and also compare the efficiency of DMUs under the robust models (6) and (13), we consider a numerical example with data from Hatami-Marbini and Toloo (2017). See Table 1. The input and output data are taken 5% perturbation from their nominal values. The results for the CCR model and robust CCR models are shown in Table 2. Omega values, $\Omega_j = 0, 0.5, 1, 2$ and 2.8 in model (13) are arbitrary chosen bearing in mind $\Omega_j = 0$ is equivalent to the CCR efficiency in column two of

**Table 2** Results of CCR and RCCR models

| DMU | $\theta^*$ | $w^*$ | $z^*$ | | | | Classification |
|-----|-----------|-------|-------------------|-------------------|------------------|-------------------|----------------|
| | | | $(\Omega_j = 0.5)$ | $(\Omega_j = 1)$ | $(\Omega_j = 2)$ | $(\Omega_j = 2.8)$ | |
| 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| 2 | 1.000 | 0.944 | 1.000 | 0.944 | 0.899 | 0.899 | $RE^{+}$ |
| 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| 5 | 1.000 | 0.884 | 0.959 | 0.884 | 0.842 | 0.853 | $RE^{-}$ |
| 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| 7 | 1.000 | 0.956 | 1.000 | 0.956 | 0.931 | 0.931 | $RE^{+}$ |
| 8 | 1.000 | 0.875 | 0.947 | 0.875 | 0.839 | 0.839 | $RE^{-}$ |
| 9 | 0.994 | 0.833 | 0.905 | 0.833 | 0.812 | 0.798 | $RE^{-}$ |
| 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |

Table 2 and $\Omega_j = \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5} \cong 2.8$ is the highest conservatism the DM can tolerate for the uncertain data. From column 2, all other DMUs are CCR efficient except $DMU_9$. However, the efficiency scores in the robust model (13) decrease when uncertainty is considered in the data and $\Omega_j$ increases. Figure 3 shows the efficiency of DMUs as $\Omega_j$ increases from $\Omega_j = 0$ to $\Omega_j = 2.8$ at an interval of 0.4. Here, the robust efficiency of $DMU_1$, $DMU_3$, $DMU_4$, $DMU_6$ and $DMU_{10}$ remain the same at 1 for all values of $\Omega_j$. These DMUs are called $R$-efficient.

Considering columns three and five of Table 2, it is evident the equivalency of the robust models (6) and (13) when the ellipsoid is inscribed by the box/interval. Also, the robust efficiency scores of model (13) include that of model (6) at $\Omega_j = 1.0$. Here, the DMUs which are $R$-inefficient in the later model are also $R$-inefficient in the former model. However, as mentioned earlier, it is possible that the maximum realization of the uncertain data may occur at the corners of the interval (see Fig. 2) which implies that model (13) can be more conservative and with higher complexity than model (6) at full protection of the uncertain data. The efficiency classification according to the DM conservativeness is shown in the last column of Table 2. It is observed that the R-efficient DMUs are $RE^{++} = \left\{ DMU_1, DMU_3, DMU_4, DMU_6, DMU_{10} \right\}$, the PR-efficient DMUs are $RE^{+} = \left\{ DMU_2, DMU_7 \right\}$ while finally, the $R$-inefficient DMUs are $RE^{-} = \left\{ DMU_5, DMU_8, DMU_9 \right\}$.

## 4 Extension to the additive DEA model and imprecise data

In this section, we extend the robust approach to the additive (ADD) model with imprecise data. Consider the additive (ADD) model proposed in Charnes et al. (1985) to evaluate the efficiency of DMUs:
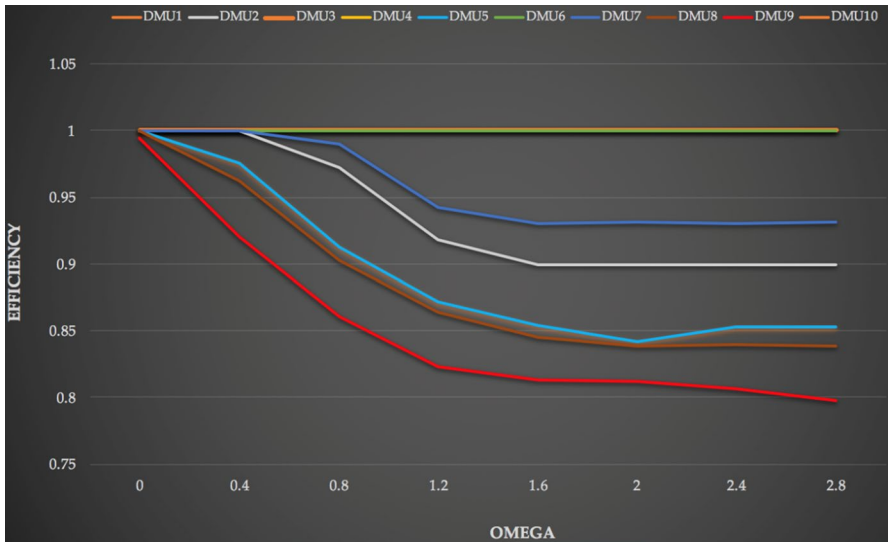
**Fig. 3** Efficiency scores of DMUs under different values of omega

$$s_o^* = \max \sum_{j=1}^{m} s_i^- + \sum_{j=1}^{s} s_r^+$$

s.t

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0} \qquad i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{r0} \qquad r = 1, \ldots, s \qquad (14)$$

$$\lambda_j \geq 0 \qquad j = 1, \ldots, n$$

$$s_i^- \geq 0 \qquad i = 1, \ldots, m$$

$$s_r^+ \geq 0 \qquad r = 1, \ldots, s$$

where $s_i^-$ and $s_r^+$ are the slacks for the input and output, respectively. To extend the RO to the additive model above, first, consider the dual formulation of model (14). Again, this is to avoid any possible infeasibility resulting from uncertainty analysis in the equality constraints. The dual of model (14) is the following:

$$\min \omega$$

s.t.

$$\omega - \sum_{i=1}^{m} v_i x_{io} + \sum_{r=1}^{s} u_r y_{ro} \leq 0$$

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0 \qquad j = 1, \ldots, n \qquad (15)$$

$$v_i \geq 0 \qquad i = 1, \ldots, m$$

$$u_r \geq 0 \qquad r = 1, \ldots, s$$

where $\omega$ is the efficiency of DMU$_o$. It is easily verifiable that $\omega^* \geq 0$; thus, an efficient point $(x_{ij}, y_{rj})$ will lie on the facet-defining hyperplane with equation $\sum_{i=1}^m v_i^* x_{rj} - \sum_{r=1}^s u_r^* y_{ro} = 0$. Then, a DMU $j$ is efficient if $\omega^* = 0$ and inefficient if $\omega^* > 0$ or alternatively, $\omega^* > 0$ and $(v^*, u^*) \geq 1_{m+s}$ measures the inefficiencies of the DMUs. In particular, to obtain an efficiency preserving unit to data perturbation, we consider the ellipsoidal-interval uncertainty defined in (12), and similarly to model (13), we propose the following robust additive model (RADD):

$$\min \varpi$$
s.t.
$$\varpi - \sum_{i=1}^m v_i x_{io} + \sum_{r=1}^s u_r y_{ro} + \sum_{i \in I_o} \lambda_{io} \hat{x}_{io}$$
$$+ \Omega_o^x \sqrt{\sum_{i \in I_o} v_{io}^2 \hat{x}_{io}^2} + \sum_{r \in R_o} \mu_{ro} \hat{y}_{ro} + \Omega_o^y \sqrt{\sum_{r \in R_o} v_{ro}^2 \hat{y}_{ro}^2} \leq 0$$
$$\sum_{i=1}^m v_i x_{io} - \sum_{r=1}^s u_r y_{ro} + \sum_{r \in R_o} \mu_{ro} \hat{y}_{ro} + \sum_{i \in I_o} \lambda_{io} \hat{x}_{io}$$
$$\Omega_o^y \sqrt{\sum_{r \in R_o} v_{ro}^2 \hat{y}_{ro}^2} + \Omega_o^x \sqrt{\sum_{i \in I_o} v_{io}^2 \hat{x}_{io}^2} \geq 0$$
$$\sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} + \sum_{r \in R_j} \mu_{rj} \hat{y}_{rj} + \sum_{i \in I_j} \lambda_{ij} \hat{x}_{ij}$$
$$\Omega_j^y \sqrt{\sum_{r \in R_j} v_{rj}^2 \hat{y}_{rj}^2} + \Omega_j^x \sqrt{\sum_{i \in I_j} v_{ij}^2 \hat{x}_{ij}^2} \geq 0 \qquad\qquad j \neq o$$
$$-\mu_{rj} \leq u_r - v_{rj} \leq \mu_{rj} \qquad\qquad \forall r \in R_j$$
$$-\lambda_{ij} \leq v_i - v_{ij} \leq \lambda_{ij} \qquad\qquad \forall i \in I_j$$
$$v_i \geq 1 \qquad\qquad \forall i$$
$$u_r \geq 1 \qquad\qquad \forall r$$
$$\lambda_{ij}, \mu_{rj} \geq 0 \qquad\qquad \forall i \in I_j, \forall r \in R_j$$

(16)

where $\varpi$ is the robust additive efficiency of DMU$_o$

The RADD model (16) can be compared to the imprecise additive models developed in Lee et al. (2002) and Matin et al. (2007). Consider the numerical example given in Cooper et al. (1999) and presented in Table 3. The column headings indicate the data to be dealt with in ordinal and bounded forms as well as in the customary exact forms represented by the conditions $y_r \in D_r^+, x_i \in D_i^-$ where $D_r^+$ and $D_i^-$. DEA models described by these data are nonlinear and usually converted to linear standard DEA with exact data by using the transformation approach suggested in Zhu (2003). It must be noted that the robust model is not able to deal with ordinal and bounded data. The approach adopted in this paper follows the transformation of bound and ordinal data in Table 3 to exact data in Lee et al. (2002). The result of the retrieved exact data is given in Table 4. From this data, we compare the result of the RADD model with the two-stage imprecise additive model of Lee et al. (2002) and the one-stage imprecise additive model of Matin et al. (2007). Table 5 presents the inefficiency of DMUs proposed by the different methods. The efficiency of DMUs provided by the proposed robust model ($\Omega_j = 0$) indicated in Table 5 is the same as the former two methods where the RADD model yields larger scores for the inefficient

**Table 3** Exact and imprecise data adapted from Cooper et al. (1999)

| DMU | Inputs | | Outputs | |
|---|---|---|---|---|
| | Exact | Bound | Exact | Ordinal |
| | Cost | Judgment | Revenue | Satisfaction |
| $j$ | $x_{1j}$ | $x_{2j}$[a] | $y_{1j}$ | $y_{2j}$[b] |
| 1 | 100 | [0.6, 0.7] | 2000 | 4 |
| 2 | 150 | [0.8, 0.9] | 1000 | 2 |
| 3 | 150 | 1 | 1200 | 5 |
| 4 | 200 | [0.7, 0.8] | 900 | 1 |
| 5 | 200 | 1 | 600 | 3 |

[a]Ordinal ranking such that 5 = highest rank, …, 1 = lowest rank (i.e., $y_{23} \geq \cdots \geq y_{24}$)

[b]Ratio bound based on the reference DMUs 3 or 5 (e.g., $0.6 \leq x_{21} \leq 0.7$ with $x_{23} = 1$)

DMUs and with higher discriminating power. The performance of DMUs on the three models are indifferent and their efficiency score according to Table 5 is ranked as follows:

$$\text{DMU}_1 \sim \text{DMU}_3 \succ \text{DMU}_2 \succ \text{DMU}_4 \succ \text{DMU}_5$$

where the symbol ''~'' denotes ''indifferent to'' and the symbol ''≻'' denotes ''superior to''. It should be noted that the RADD model lightens the computational burden compared to the imprecise DEA models and provides the flexibility for controlling the conservativeness of solution to data perturbations. Thus, for some imprecise data, the proposed model in this paper is more computationally effective and flexible in robustly ranking the efficiency of DMUs.

## 5 Application to banking efficiency in Italy

We demonstrate the real-world application of the proposed robust CCR models by analyzing the performance of banks operating in Italy. The Italian banking industry is emerging from a prolonged period of distress following the global financial crises in 2008 and the slowdown of the Italian economy.[4] Although the banking system has shown enough resilience and recovering over the years, competition in the global uncertain environment, particularly in Europe has required that the banks operate efficiently and robustly. Indeed, banking idiosyncratic uncertainties translate into

---

[4] The banking crises that engulfed Italy and ongoing mildly can be attributed to two main sources. First is the financial market crises in 2008 that was caused by mortgage crises and largely the failure of the Lehman Brothers. The second one stems from the sovereign debt crises that affected Greece and some peripheral countries of the European monetary union: Italy, Spain, Portugal, Ireland. The Italian government through the bank of Italy in its supervisory capacity instituted measures such as the provision of liquidity, strengthening and supporting of banks, recapitalization of distress banks and including the so-called "Tremonti bond". The measures were to revitalize the banking industry, protect depositors and also finance the economy.

**Table 4** Retrieved exact data adapted from Lee et al. (2002)

| DMUs | $x_{1j}$ | $x_{2j}$ | $y_{1j}$ | $y_{2j}$ |
|---|---|---|---|---|
| 1 | 100 | 0.7 | 2000 | 50 |
| 2 | 150 | 0.8 | 1000 | 20 |
| 3 | 150 | 1 | 1200 | 100 |
| 4 | 200 | 0.8 | 900 | 10 |
| 5 | 200 | 1 | 600 | 20 |

Exact data when $\text{DMU}_j (j = 2, 3, 4, 5)$ is under evaluation. $\text{DMU}_{o \equiv 1}$ for $x_{21} = 0.6$

**Table 5** Computed inefficiency with different additive models

| DMUs | Lee et al. (2002)[a] | Martin et al. (2007)[b] | RADD model |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 2 | 1321.429 | 1050.1 | 1358.57 |
| 3 | 0 | 0 | 0 |
| 4 | 1200 | 1200 | 1518.57 |
| 5 | 2314.286 | 1500.3 | 2365.71 |

[a]The result here is taken from Lee et al. (2002) two-stage approach

[b]The result here is taken from Matin et al. (2007) one-stage approach

uncertainties in banking data. Therefore, to obtain robust performance evaluation of the banks, we consider the RDEA models developed in this paper and assume banks as decision-making units that consume inputs, e.g., assets and equity to generate an amount of output level, e.g., loans and revenue in an uncertain environment.

## 5.1 Bank data and variable selection

Data comprising 29 main banks in Italy for the accounting year 2015 were collected from the Bureau van Dick—Bankscope database (Bank scope 2015). See also Alfiero et al. (2019). The selected banks operate under a common set of rules and regulations set up by the Bank of Italy and by extension the Central Bank of Europe which implies that they have a common current denominator for which comparison of performance can be smoothly made. Next is the selection of inputs and outputs which is crucial in the banking efficiency measurement. In the banking sector, similar to other sectors, a consensus is reached on the classification of some factors as inputs and outputs. However, the classification of others particularly deposit is unclear and controversial. The debate on bank deposit which in the DEA literature is termed as a flexible measure or dual-role factors (see Toloo 2012; Toloo et al. 2018) is that, depending on the operational activities of the bank, in one hand, deposit

could be regarded as an input (*intermediation approach*) and on the other hand as an output (*production approach*) or as a major component involved in the creation of added value (*value-added approach*). Different researchers select different measures. Casu and Girardone (2002) examined the cost efficiency of the Italian bank conglomerates by assessing the cost characteristics of bank parent companies and their subsidiaries. Favoring the intermediation approach, they considered as inputs labor cost, deposits and physical capital whiles total loans and other earning assets were used as outputs. Aiello and Bonanno (2016) considered the role of banks in Italy as an intermediary and used deposits, capital and labor as input factors whiles they used loans, securities and commission income as output factors.

According to popular studies on banking efficiency, the intermediation approach is used since banks are essentially seen as financial intermediaries, whose main activities are to borrow funds from depositors and lend to others (Fethi and Pasiouras 2010). Kao and Liu (2014) among other studies consider demand deposits as outputs in the intermediation approach. Within this context and following the survey of Mostafa (2009) in which deposit is mostly used as outputs, we select as input factors; employees, assets and equity and as output factors; deposits from banks, loans and revenue. Table 6 shows the input and output factors and statistics for the Italian banks used for this study (see Table 9 in "Appendix A" for details of the banks). All the inputs and outputs are expressed in monetary values. It is assumed that the actual values of some of the input and output factors are uncertain. A bank has uncertainty characterization if any of its input or output data for the performance measurement is uncertain. Here, we perceive uncertainty in banking data to be the result of errors from measurement and statistical computations and other errors such as from forecast values of loans, non-performing loans, deposit, etc. Following this development, we then apply the proposed models (6) and (13) to assess the robust performance of the banks.

## 5.2 Efficiency results

In the proposed robust models, we seek to obtain an acceptable performance level of the banks by optimizing the worst-case values of the uncertain inputs and outputs values in the ellipsoids. For each bank, uncertainty is considered in some or all the inputs and outputs where the realization of their values are restricted to the uncertainty sets. We suppose that the inputs and outputs deviate from their nominal values by a percentage of perturbation, $\delta = 0.05$. The result of the model implementation is reported in Table 7. The third column shows the efficiency ranking by the DEA model (2), and the fifth and last column show the efficiency ranking by the robust models (6) and (13). A comparative view of the efficiency of these three models is given in Fig 4. Note that, the result obtained in model (13) for $\Omega_j = \left( \left| R_j \right| + \left| I_j \right| \right)^{0.5} \cong 2.5$ indicates the highest conservativeness of decision makers which occurs at the full protection of the inputs and outputs against all uncertainties. Column 2 of Table 7 shows 8 banks with an efficiency score equal to 1 which are efficient under the DEA model (2) and two banks (B13 and B22) which are *R*-efficient in the two robust models. The robust efficiency decreases relative to the

**Table 6** Data for 29 major banks in Italy in 2015

| Bank | Inputs | | | Outputs | | |
|------|--------|--------|--------|----------|--------|--------|
| | Employees | Assets | Equity | Deposits | Loans | Revenue |
| B01 | 125,510.00 | 860,433.40 | 53,485.50 | 163,050.90 | 445,293.90 | 12,764.70 |
| B02 | 90,807.00 | 676,496.00 | 48,593.00 | 79,743.00 | 350,010.00 | 10,259.00 |
| B03 | 25,731.00 | 169,012.00 | 9622.70 | 28,068.10 | 106,680.40 | 2905.50 |
| B04 | 16,972.00 | 120,509.60 | 8546.80 | 24,078.00 | 71,902.80 | 1843.70 |
| B05 | 17,718.00 | 117,200.80 | 10,517.80 | 16,626.80 | 83,815.70 | 1724.50 |
| B06 | 13,371.00 | 77,494.50 | 5649.00 | 22,222.70 | 60,523.20 | 1750.30 |
| B07 | 11,447.00 | 61,261.20 | 5651.80 | 7385.60 | 43,702.60 | 1362.30 |
| B08 | 8197.00 | 51,373.20 | 5138.10 | 4290.60 | 36,462.50 | 1017.50 |
| B09 | 7743.00 | 50,203.30 | 4647.40 | 9000.70 | 33,953.90 | 870.50 |
| B10 | 2651.00 | 44,710.20 | 2070.10 | 8803.90 | 7430.60 | 322.20 |
| B11 | 5273.00 | 39,783.40 | 2552.10 | 9973.50 | 25,068.30 | 584.70 |
| B12 | 6019.00 | 37,455.30 | 2479.70 | 6190.70 | 22,649.40 | 509.30 |
| B13 | 3195.00 | 35,537.60 | 2649.40 | 3029.20 | 23,290.40 | 547.10 |
| B14 | 6263.00 | 33,349.30 | 2153.40 | 6913.00 | 22,012.20 | 556.70 |
| B15 | 5034.00 | 30,298.90 | 2489.10 | 3352.70 | 20,395.20 | 383.80 |
| B16 | 5868.00 | 27,916.70 | 1687.20 | 449.90 | 18,004.80 | 697.50 |
| B17 | 4123.00 | 26,901.70 | 2187.70 | 4194.10 | 18,263.50 | 544.90 |
| B18 | 3927.00 | 24,186.20 | 1663.40 | 5811.20 | 19,070.70 | 543.90 |
| B19 | 3588.00 | 21,861.10 | 2323.70 | 2323.10 | 18,736.10 | 426.20 |
| B20 | 3064.00 | 14,968.20 | 1297.20 | 1661.60 | 13,121.70 | 352.90 |
| B21 | 3194.00 | 14,809.50 | 1084.70 | 2680.60 | 9414.80 | 266.30 |
| B22 | 916.00 | 13,852.60 | 226.30 | 3731.30 | 4942.40 | 141.70 |
| B23 | 2208.00 | 13,545.30 | 1387.60 | 1467.40 | 12,295.50 | 261.00 |
| B24 | 2570.00 | 13,205.90 | 1258.50 | 2137.90 | 7945.80 | 277.20 |
| B25 | 1863.00 | 12,276.90 | 1006.20 | 2375.20 | 6795.10 | 175.50 |
| B26 | 2371.00 | 12,248.10 | 922.10 | 905.20 | 9386.30 | 285.70 |
| B27 | 1207.00 | 11,769.20 | 750.10 | 1612.40 | 6394.80 | 175.40 |
| B28 | 2443.00 | 11,615.50 | 724.70 | 471.10 | 9328.80 | 292.30 |
| B29 | 2989.00 | 10,765.90 | 771.20 | 381.50 | 6734.00 | 184.50 |
| Mean | 13,319.38 | 90,863.50 | 6328.84 | 14,583.86 | 52,193.98 | 1449.20 |
| SD | 27,280.19 | 192,987.03 | 12,690.11 | 32,430.64 | 99,685.54 | 2879.44 |
| Max | 916.00 | 10,765.90 | 226.30 | 381.50 | 4942.40 | 141.70 |
| Min | 125,510.00 | 860,433.40 | 53,485.50 | 163,050.90 | 445,293.90 | 12,764.70 |

DEA efficiency which indicates the worst-case and reliable performance of the banks in uncertain conditions. The DEA result indicates an average overall technical efficiency (0.898) for the banks under study which further indicates that although the banks are performing averagely well, the number of banks which are efficient with or without uncertainty analysis is quite small. The least performing bank includes UniCredit SpA (B01) with an efficiency score of 0.738.

**Table 7** Efficiency scores and ranking—CCR and RCCR models

| Banks | CCR (2) | Rank | RCCR (6) | Rank | RCCR (13)[a] | Rank |
|-------|---------|------|----------|------|-------------|------|
| B01 | 0.738 | 29 | 0.649 | 27 | 0.602 | 29 |
| B02 | 0.769 | 26 | 0.648 | 28 | 0.629 | 26 |
| B03 | 0.897 | 16 | 0.767 | 16 | 0.733 | 13 |
| B04 | 0.847 | 21 | 0.729 | 21 | 0.692 | 20 |
| B05 | 0.876 | 18 | 0.748 | 17 | 0.717 | 16 |
| B06 | 1.000 | 1 | 0.963 | 4 | 0.947 | 3 |
| B07 | 0.932 | 13 | 0.792 | 13 | 0.728 | 15 |
| B08 | 0.899 | 15 | 0.780 | 15 | 0.713 | 18 |
| B09 | 0.860 | 20 | 0.737 | 20 | 0.702 | 19 |
| B10 | 0.815 | 22 | 0.676 | 24 | 0.667 | 21 |
| B11 | 0.975 | 9 | 0.840 | 10 | 0.791 | 9 |
| B12 | 0.785 | 25 | 0.667 | 25 | 0.642 | 25 |
| B13 | 1.000 | 1 | 1.000 | 1 | 1.000 | 1 |
| B14 | 0.874 | 19 | 0.748 | 18 | 0.714 | 17 |
| B15 | 0.803 | 23 | 0.683 | 22 | 0.656 | 22 |
| B16 | 1.000 | 1 | 0.859 | 8 | 0.833 | 8 |
| B17 | 0.938 | 12 | 0.793 | 12 | 0.745 | 12 |
| B18 | 1.000 | 1 | 0.921 | 5 | 0.850 | 6 |
| B19 | 0.972 | 10 | 0.842 | 9 | 0.791 | 9 |
| B20 | 1.000 | 1 | 0.882 | 7 | 0.840 | 7 |
| B21 | 0.792 | 24 | 0.680 | 23 | 0.648 | 23 |
| B22 | 1.000 | 1 | 1.000 | 1 | 1.000 | 1 |
| B23 | 1.000 | 1 | 0.912 | 6 | 0.868 | 5 |
| B24 | 0.887 | 17 | 0.738 | 19 | 0.643 | 24 |
| B25 | 0.759 | 27 | 0.653 | 26 | 0.617 | 27 |
| B26 | 0.963 | 11 | 0.829 | 11 | 0.765 | 11 |
| B27 | 0.922 | 14 | 0.784 | 14 | 0.731 | 14 |
| B28 | 1.000 | 1 | 0.969 | 3 | 0.900 | 4 |
| B29 | 0.753 | 28 | 0.629 | 29 | 0.614 | 28 |

[a]Note that this result is obtained for $\Omega_j = 2.5$

For the robust classification of banks, the robust parameter $\Omega_j$ is set to a range from 0 when no uncertainty in data is anticipated to $\Omega_j = 2.5$ when full protection for uncertainty is anticipated. The choice of appropriate $\Omega_j$ within this range is selected arbitrarily. Table 8 shows the result of the robust classification of the banks. In exchange for higher guaranteed robustness, higher values of $\Omega_j$ is selected. The efficiency of banks decrease as $\Omega_j$ increases and the DM can express preferences with different values of $\Omega_j$ and robust efficiency which is similar to the approach proposed in Ben-Tal and Nemirovski (2000) and Sadjadi and Omrani (2008). Full protection of the inputs and outputs, only B13 and B22 have $R$-efficiency. Banks B06, B16, B18, B20, B23 and B28 are PR-efficient at
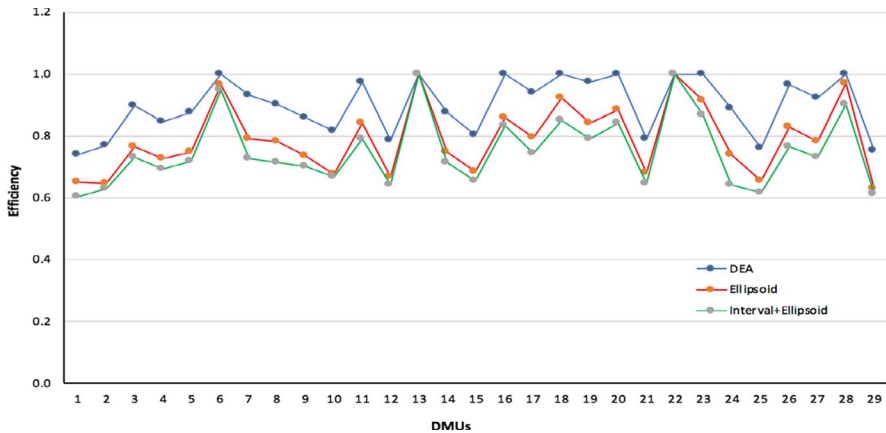
**Fig. 4** The result from ellipsoid and interval-based ellipsoid $\left(\Omega_j = 2.5\right)$ sets

different conservativeness level. The rest of the DMUs are R-inefficient. The last column of Table 8 shows the classification of the banks as given in Definition 5. Because many banks were inefficient in the traditional DEA evaluation, it is unsurprising the number of efficient banks which are partially or fully robust efficient. As observed, 2 banks and 6 banks are fully or partially robust efficient at different levels from the 8 DEA efficient banks.

## 6 Concluding remarks

In this paper, we proposed new robust DEA models based on the ellipsoidal uncertainty and interval-based ellipsoidal uncertainty sets designed in Ben-Tal and Nemirovski (1999, 2000). This has been done in a manner that immunizes arbitrary bounded or unbounded uncertainties partly or in all inputs and outputs data simultaneously. By constraining the uncertain data in an ellipsoidal uncertainty sets, the models developed in this paper become less pessimistic and in contrast offer the advantage over the interval DEA models which mostly evaluate the performance of DMUs based on their extreme lower and upper bounds of efficiency. The developed RDEA models provides the DM the flexibility of controlling the level of robustness. Another important contribution that is made in this paper is the design of a classification scheme which enables the DM to classify DMUs into fully robust efficient, partially robust efficient and robust inefficient. We provide numerical examples to illustrate the proposed models particlularly, for our proposed robust additive model which is compared with some IDEA models to show its efficacy, potential and applicability. Furthermore, the proposed robust models are applied for the evaluation and classification of banks in Italy. The proposed model enables bank managers to classify banks into fully, partially and robust (in)efficient units. Employing the RDEA models with different uncertainties to classify DMUs in applications can be considered for future research.

**Table 8** Classification of banks based on the robust model (13)

| Banks | $\delta = 0.05$ | | | | | | | Classification |
|---|---|---|---|---|---|---|---|---|
| | $\Omega_j = 0$ | $\Omega_j = 0.1$ | $\Omega_j = 0.5$ | $\Omega_j = 1.0$ | $\Omega_j = 1.5$ | $\Omega_j = 2.0$ | $\Omega_j = 2.5$ | |
| B01 | 0.738 | 0.728 | 0.690 | 0.649 | 0.611 | 0.603 | 0.602 | $RE^-$ |
| B02 | 0.769 | 0.755 | 0.704 | 0.648 | 0.629 | 0.629 | 0.629 | $RE^-$ |
| B03 | 0.897 | 0.882 | 0.827 | 0.767 | 0.741 | 0.733 | 0.733 | $RE^-$ |
| B04 | 0.847 | 0.835 | 0.786 | 0.729 | 0.693 | 0.693 | 0.692 | $RE^-$ |
| B05 | 0.876 | 0.863 | 0.810 | 0.748 | 0.717 | 0.717 | 0.717 | $RE^-$ |
| B06 | 1.000 | 1.000 | 1.000 | 0.963 | 0.951 | 0.947 | 0.947 | $RE^+$ |
| B07 | 0.932 | 0.916 | 0.856 | 0.792 | 0.749 | 0.735 | 0.728 | $RE^-$ |
| B08 | 0.899 | 0.886 | 0.837 | 0.780 | 0.732 | 0.713 | 0.713 | $RE^-$ |
| B09 | 0.860 | 0.847 | 0.796 | 0.737 | 0.703 | 0.703 | 0.702 | $RE^-$ |
| B10 | 0.815 | 0.799 | 0.739 | 0.676 | 0.667 | 0.667 | 0.667 | $RE^-$ |
| B11 | 0.975 | 0.960 | 0.905 | 0.840 | 0.798 | 0.794 | 0.791 | $RE^-$ |
| B12 | 0.785 | 0.772 | 0.723 | 0.667 | 0.649 | 0.642 | 0.642 | $RE^-$ |
| B13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| B14 | 0.874 | 0.859 | 0.807 | 0.748 | 0.715 | 0.715 | 0.714 | $RE^-$ |
| B15 | 0.803 | 0.790 | 0.740 | 0.683 | 0.664 | 0.657 | 0.656 | $RE^-$ |
| B16 | 1.000 | 1.000 | 0.931 | 0.859 | 0.836 | 0.834 | 0.833 | $RE^+$ |
| B17 | 0.938 | 0.922 | 0.859 | 0.793 | 0.766 | 0.759 | 0.745 | $RE^-$ |
| B18 | 1.000 | 1.000 | 0.985 | 0.921 | 0.873 | 0.858 | 0.850 | $RE^+$ |
| B19 | 0.972 | 0.957 | 0.904 | 0.842 | 0.791 | 0.791 | 0.791 | $RE^-$ |
| B20 | 1.000 | 1.000 | 0.960 | 0.882 | 0.854 | 0.848 | 0.840 | $RE^+$ |
| B21 | 0.792 | 0.780 | 0.733 | 0.680 | 0.651 | 0.648 | 0.648 | $RE^-$ |
| B22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | $RE^{++}$ |
| B23 | 1.000 | 1.000 | 0.988 | 0.912 | 0.885 | 0.875 | 0.868 | $RE^+$ |
| B24 | 0.887 | 0.870 | 0.806 | 0.738 | 0.698 | 0.668 | 0.643 | $RE^-$ |
| B25 | 0.759 | 0.748 | 0.704 | 0.653 | 0.621 | 0.617 | 0.617 | $RE^-$ |
| B26 | 0.963 | 0.948 | 0.893 | 0.829 | 0.784 | 0.767 | 0.765 | $RE^-$ |
| B27 | 0.922 | 0.906 | 0.847 | 0.784 | 0.758 | 0.748 | 0.731 | $RE^-$ |
| B28 | 1.000 | 1.000 | 1.000 | 0.969 | 0.916 | 0.900 | 0.900 | $RE^+$ |
| B29 | 0.753 | 0.739 | 0.688 | 0.629 | 0.614 | 0.614 | 0.614 | $RE^-$ |

# Appendix A

See Table 9.

**Table 9** Major banks in Italy used for the analysis

| Banks | Bank name |
|-------|-----------|
| B01 | UniCredit SpA |
| B02 | Intesa Sanpaolo |
| B03 | Banca Monte dei Paschi di Siena SpA-Gruppo Monte dei Paschi di Siena |
| B04 | Banco Popolare - Società Cooperativa-Banco Popolare |
| B05 | Unione di Banche Italiane Scpa-UBI Banca |
| B06 | Banca Nazionale del Lavoro SpA |
| B07 | Banca popolare dell'Emilia Romagna |
| B08 | Cassa di Risparmio di Parma e Piacenza SpA |
| B09 | Banca Popolare di Milano SCaRL |
| B10 | Banca Mediolanum SpA |
| B11 | Banca Popolare di Vicenza Societa cooperativa per azioni |
| B12 | Credito Emiliano SpA-CREDEM |
| B13 | Banca Popolare di Sondrio Societa Cooperativa per Azioni |
| B14 | Veneto Banca scpa |
| B15 | Banca Carige SpA |
| B16 | Banco di Napoli SpA |
| B17 | Banca Piccolo Credito Valtellinese-Credito Valtellinese Soc Coop |
| B18 | Deutsche Bank SpA |
| B19 | Banca Popolare di Bergamo SpA |
| B20 | Cassa di Risparmio del Veneto SpA |
| B21 | Banca Popolare di Bari Soc. Coop.P.A |
| B22 | CheBanca SpA |
| B23 | Banco di Brescia San Paolo Cab SpA |
| B24 | Banco di Sardegna SpA |
| B25 | Cassa di risparmio di Asti SpA |
| B26 | Banco di Desio e della Brianza SpA-Banco Desio |
| B27 | Banca di Credito Cooperativo di Roma |
| B28 | Unipol Banca Spa |
| B29 | Banca Sella SpA |

# References

Aiello, F., Bonanno, G.: Bank efficiency and local market conditions: evidence from Italy. J. Econ. Bus. **83**, 70–90 (2016)

Alfiero, S., Esposito, A., Mensah, E.K., Toloo, M.: A dataset of European banks in performance evaluation under uncertainty. Data Brief **22**, 214–217 (2019)

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

Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag. Sci. **30**(9), 1078–1092 (1984)

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

Ben-tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Oper. Res. Lett. **25**(1), 1–13 (1999)

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

Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A.: Adjustable robust solutions of uncertain linear programs. Math. Program. **99**(2), 351–376 (2004)

Ben-Tal, A., El Ghaoui, L., Nemirovskiĭ, A.S.: Robust Optimization. Princeton University Press, Princeton (2009)

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

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

Calafiore, G., El Ghaoui, L.: Ellipsoidal bounds for uncertain linear equations and dynamical systems. Automatica **40**(5), 773–787 (2004)

Casu, B., Girardone, C.: A comparative study of the cost efficiency of Italian bank conglomerates. Manag. Finance **28**(6), 34–45 (2002)

Charnes, A., Cooper, W.W.: Programming with linear fractional functionals. Nav. Res. Logist. Q. **9**(3–4), 181–186 (1962)

Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)

Charnes, A., Cooper, W., Golany, B., Seiford, L., Stutz, J.: Foundations of DEA fo Pareto-koopmans efficient empirical production functions. J. Econom. **30**, 91–107 (1985)

Cooper, W.W., Huang, Z., Lelas, V., Li, S.X., Olesen, O.B.: Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. J. Prod. Analy. **9**(1), 53–79 (1998)

Cooper, W.W., Park, K.S., Yu, G.: IDEA and AR-IDEA: models for dealing with imprecise data in DEA. Manag. Sci. **45**(4), 597–607 (1999)

El Ghaoui, L., Oustry, F., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM J. Optim. **9**(1), 33–52 (1998)

Esfandiari, M., Hafezalkotob, A., Khalili-Damghani, K., Amirkhan, M.: Robust two-stage DEA models under discrete uncertain data. Int. J. Manag. Sci. Eng. Manag. **12**(3), 216–224 (2017)

Farzipoor Sean, R., Memariani, A., Lotfi, F.H.: The effect of correlation coefficient among multiple input vectors on the efficiency mean in data envelopment analysis. Appl. Math. Comput. **162**(2), 503–521 (2005)

Fethi, M.D., Pasiouras, F.: Assessing bank efficiency and performance with operational research and artificial intelligence techniques: a survey. Eur. J. Oper. Res. **204**(2), 189–198 (2010)

Fischetti, M., Monaci, M.: Light robustness. In: Ahuja, R.K., Möhring, R.H., Zaroliagis, C.D. (eds.) Robust and Online Large-Scale Optimization. Vol. 5868 LNCS, pp. 61–84. Springer, Berlin (2009)

Hanks, R.W., Weir, J.D., Lunday, B.J.: Robust goal programming using different robustness echelons via norm-based and ellipsoidal uncertainty sets. Eur. J. Oper. Res. **262**(2), 636–646 (2017)

Hatami-Marbini, A., Toloo, M.: An extended multiple criteria data envelopment analysis model. Exp. Syst. Appl. **73**, 201–219 (2017)

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

Kao, C., Liu, S.T.: Measuring performance improvement of Taiwanese commercial banks under uncertainty. Eur. J. Oper. Res. **235**(3), 755–764 (2014)

Laguna, M.: Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. Manag. Sci. **44**(11), S101–S110 (1998)

Land, K.C., Lovell, C.A.K., Thore, S.: Chance-constrained data envelopment analysis. Manag. Decis. Econ. **14**(6), 541–554 (1993)

Lee, Y.K., Sam Park, K., Kim, S.H.: Identification of inefficiencies in an additive model based IDEA (imprecise data envelopment analysis). Comput. Oper. Res. **29**(12), 1661–1676 (2002)

Lu, C.-C.: Robust data envelopment analysis approaches for evaluating algorithmic performance. Comput. Ind. Eng. **81**, 78–89 (2015)

Lu, C., Tao, J., An, Q., et al.: A second-order cone programming based robust data envelopment analysis model for the new-energy vehicle industry. Ann Oper Res **292**, 321–339 (2020)

Matin, R.K., Jahanshahloo, G.R., Vencheh, A.H.: Inefficiency evaluation with an additive DEA model under imprecise data, an application on IAUK departments. J. Oper. Res. Soc. Jpn. **50**(3), 163–177 (2007)

Mensah, E.K.: Robust Optimization in Data Envelopment Analysis: Extended theory and applications. University of Insubria, PhD Thesis (2019)

Mensah, Ek, Rocca, M.: Light robust goal programming. Math. Comput. Appl. **24**(4), 85 (2019)

Mostafa, M.M.: Modeling the efficiency of top Arab banks: a DEA-neural network approach. Exp. Syst. Appl. **36**(1), 309–320 (2009)

Mulvey, J.M., Vanderbei, R.J., Zenios, S.A.: Robust optimization of large-scale systems. Oper. Res. **43**(2), 264–281 (1995)

Olesen, O.B., Petersen, N.C.: Chance constrained efficiency evaluation. Manag. Sci. **41**(3), 442–457 (1995)

Omrani, H.: Common weights data envelopment analysis with uncertain data: a robust optimization approach. Comput. Ind. Eng. **66**(4), 1163–1170 (2013)

Park, K.S.: Efficiency bounds and efficiency classifications in DEA with imprecise data. J. Oper. Res. Soc. **58**(4), 533–540 (2007)

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

Salahi, M., Torabi, N., Amiri, A.: An optimistic robust optimization approach to common set of weights in an optimistic robust optimization approach to common set of weights in DEA. Measurement **93**, 67–73 (2016)

Salahi, M., Toloo, M., Hesabirad, Z.: Robust Russell and enhanced Russell measures in DEA. J. Oper. Res. Soc. **70**, 1275–1283 (2018). **(in press)**

Shakouri, R., Salahi, M., Kordrostami, S.: Stochastic p-robust approach to two-stage network DEA model. Quant. Finance Econ. **3**(2), 315–346 (2019)

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

Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper. Res. **21**(3), 1154–1157 (1973)

Toloo, M.: Alternative solutions for classifying inputs and outputs in data envelopment analysis. Comput. Math Appl. **63**(6), 1104–1110 (2012)

Toloo, M.: An epsilon-free approach for finding the most efficient unit in DEA. Appl. Math. Model. **38**(13), 3182–3192 (2014)

Toloo, M., Mensah, E.K.: Robust optimization with nonnegative decision variables: a DEA approach. Comput. Ind. Eng. **127**, 313–325 (2019)

Toloo, M., Allahyar, M., Hančlová, J.: A non-radial directional distance method on classifying inputs and outputs in DEA: application to banking industry. Exp. Syst. Appl. **92**, 495–506 (2018)

Wu, D., Ding, W.D., Koubaa, A., Chaala, A., Luo, C.C.: Robust DEA to assess the reliability of methyl methacrylate-hardened hybrid poplar wood. Ann. Oper. Res. **248**(1–2), 515–529 (2017)

Zahedi-Seresht, M., Jahanshahloo, G.R., Jablonsky, J.: A robust data envelopment analysis model with different scenarios. Appl. Math. Model. **52**, 306–319 (2017)

Zhu, J.: Imprecise data envelopment analysis (IDEA): a review and improvement with an application. Eur. J. Oper. Res. **144**(3), 513–529 (2003)