Uncertain Negative Data in DEA: An Application of Banking in Malaysia

Rokhsaneh Yousef Zehi(✉) and Noor Saifurina Nana Khurizan

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
yousefzehi.rokhsaneh@gmail.com

Abstract. DEA models and their applicability is heavily depended on the type of data that has been used for efficiency assessment. Conventional DEA models assume the all the involved data in the efficiency evaluation are non-negative, which in many cases seems unrealistic specially when the profit or the rate of growth are involved in the evaluation of organizations. Moreover, the perturbation in data is unavoidable in real-world applications and negative data also might be affected by error. In this paper we propose a robust DEA model to handle uncertain negative data that guarantees the robustness of solution against the uncertainty in data. The proposed robust DEA model is constructed under a box-ellipsoidal uncertainty set and an application of banking in Malaysia is presented to validate the applicability of proposed model and evaluate the effect of uncertainty in efficiency assessment and ranking of 30 banks in Malaysia. The result shows that our proposed model provides a better and more discriminative ranking of banks.

Keywords: Data envelopment analysis (DEA) · Mathematical programming · Robust optimization · Uncertainty · Negative data

1 Literature Review

Data envelopment analysis (DEA) is a non-parametric approach for assessing the relative efficiency of decision-making units (DMUs) based on linear programming, which construct a frontier based on the observed input-output ratios. Charnes et al. [1] proposed the first DEA model called CCR model with the assumption of constant return to scale (CRS). Banker et al. [2] extended the CCR model and proposed BBC model which allows variable return to scale (VRS). Since introducing first DEA models, this method has been extensively applied in management sciences and various real-world applications such as education, agriculture, health care, banking, etc. (see [3] and [4] for a Survey of DEA applications).

Conventional DEA models assume non-negative values for input and output observations which is unrealistic and in many real-life applications of DEA models some variables can take both positive and negative values such as rate of growth, profit, rate of return, operational cost savings etc. One of the firstly presented approaches to cope with negative data is applying data transformation whereas an arbitrary lager number is added to into all the values of a variable which can turn negative values into positive values [5]. However, it should be mentioned that based on the applied model the results are

N. Wahi et al. (Eds.): ICMSS 2022, ACSR 98, pp. 182–194, 2023.
https://doi.org/10.2991/978-94-6463-014-5_18

different and the solutions may not be invariant to data transformations. Portela et al. [6] introduced a range directional measure (RDM) model based on the directional distance function to handle negative data. Sharp et al. [7] developed a modified slack base measure that considers both positive and negative data. One of the popular approaches for modelling negative data is based on partitioning positive and negative variables which was applied by Emrouznejad et al. [8] and based on that they proposed a semi oriented radial measure (SORM) to evaluate the efficiency of DMUs. Later several models and approaches were proposed to cope with negative data in DEA models which some of these studies can be seen in [9, 10]. It should be noted that in the presence of negative data in a technology, the CRS assumption is not possible, as in the CRS technology it as assumed that any activity can be radially expanded or contracted to create other feasible activities or in another words any proportion of an efficient unit is efficient as well, which is not consistent if some of the values of a variable are negative. Therefore, the VRS technologies are required to be assumed in the presence of negative data.

In conventional DEA models it is assumed that all data are accurate and crisp values and the uncertainty and perturbation in data is ignored. However, uncertainty in data is inevitable in many real-world applications and ignoring the perturbation in data may lead to unreliable efficiency scores and ranking of DMUs and also unattainable management decisions. There are several approaches in DEA to cope with uncertainty in data such as fuzzy DEA models [11], imprecise DEA [12] and robust DEA (RDEA) [13]. Robust DEA was firstly introduced by Sadjadi and Omrani [13] and since then it has gain lots of attentions by researchers in both theory and application perspective. RDEA has been proposed based on the robust optimization approach which was originally introduced by Soyster [14] and extended in the work of Ben-Tal and Nemirovski [15] and Bertsimas and Sim [16]. In robust optimization, the optimal solutions will be determined in a robust counterpart of the nominal problem which is constructed based on a predefined uncertainty set that ensure the optimal solution remain feasible when the data changes in the predefined uncertainty set. RDEA is one of the most popular approaches for handling uncertainty in DEA models. Sadjadi et al. [17] proposed a robust counterpart based on the Bental and Nemirovski's approach [15] for super efficiency DEA model to evaluate the gas companies. Shirazi and Mohammdi [18] evaluated Iranian airlines by developing a robust slack base measure (SBM) with undesirable outputs. Dehokhalaji et al. [19] presented the robust counterpart of the envelopment form of the CCR model based on the Ben-Tal and Nemirovski's robust approach [15] in a situation where inputs and outputs are assumed to take value from a symmetric box. RDEA is one of the most popular approaches for handling uncertainty in DEA and many researchers applied the robust optimization approaches to cope with data uncertainty in basic and advanced DEA models. A survey of RDEA studies can be seen in [20].

The focus of this paper is to propose a robust DEA model to handle uncertain negative data. As the equality constraints are challenging for constructing a robust counterpart for SORM model, and the multiplier form of SORM model contains equality constraint, we modify an equivalent model to be applied to propose a robust SORM model. The robust counterpart of the SORM model is constructed based on a box-ellipsoidal uncertainty set proposed by Ben-al and Nemirovski [15]. The level of conservativeness in this approach is controllable and it depends on the decision makers preferences. The rest of this paper is organized as follows: Sect. 2, a background on the SORM model and an equivalent model is proposed to be applied to propose a robust SORM model. Section 3 presents a robust SORM (RSORM) model under a box-ellipsoidal uncertainty set. An application on 30 banks in Malaysia is given in Sect. 4 to validate the proposed model and show the applicability of the model. Finally, the conclusion is presented in Sect. 5.

2 SORM Model

The SORM model for handling negative data has been developed by Emrouznejad et al. [8] based on partitioning approach to divide positive and negative values of a specific variable and replace negative values by a difference of two non-negative values. The first element includes positive variable, and the second element includes the absolute value of the negative part. Assume there are n DMUs, which each of them produces s outputs $\{y_{rj}; r = 1, \ldots, s\}$ using m inputs $\{x_{ij}; i = 1, \ldots, s\}$. Let us assume the input variables take only positive values and output variables can take positive values for some DMUs and negative values for the other. The output variables are divided into two subsets: $O = \{Y_r;\text{ take paositive values for all DMUs}\}$ and $O^N = \{Y_k;\text{ take paositive values some DMUS and negative values for others}\}$, where as $O \cup O^N = \{1, \ldots, s\}$ and $O \cap O^N = \emptyset$. The output variables $Y_k \in O^N$ are defined as $Y_k = Y_k^1 - Y_k^2$ where $Y_1^k, Y_1^k \geq 0$ and take value for DMU_j as follows:

$$Y_{kj}^1 = \begin{cases} Y_{kj}, & \textit{if } Y_{kj} \geq 0 \\ 0, & \textit{if } Y_{kj} < 0 \end{cases} \qquad \& \qquad Y_{kj}^2 = \begin{cases} 0, & \textit{if } Y_{kj} \geq 0 \\ -Y_{kj}, & \textit{if } Y_{kj} < 0 \end{cases}$$

Therefore, the production possibility set (PPS) in the presence of negative data for the output oriented SORM model, denoted by P_{SORM} is defined as follows:

$$P_{SORM} = \left\{ (x, y) \middle| \begin{array}{l} x \geq \sum_{j=1}^{n} \lambda_j x_j, \\ y \leq \sum_{j=1}^{n} \begin{pmatrix} Y_r \\ Y_k \end{pmatrix} \lambda_j, \\ \sum_{j=1}^{n} \lambda_j = 1, \\ \lambda_j \geq 0 \quad \forall j. \end{array} \right\}$$

Based on the above partitions and defined variables and assumptions the output oriented SORM model is defined as the following mathematical programming:

$$\begin{array}{lll} \text{Max } \varphi_o & & \forall i \in I \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}, & \forall i \in I \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geq \varphi_o y_{ro}, & \forall r \in O \\ & \sum_{j=1}^{n} \lambda_j y_{kj}^1 \geq \varphi_o y_{kj}^1, & \forall k \in O^N \\ & \sum_{j=1}^{n} \lambda_j y_{kj}^2 \leq \varphi_o y_{kj}^2, & \forall r \in O^N \\ & \sum_{j=1}^{n} \lambda_j = 1, & \forall j \\ & \lambda_j \geq 0. & \end{array} \tag{1}$$

In model (1) it can be seen that y_{kj}^2 is treated as inputs since the absolute value of the negative variables should be decreased to improve the performance of the under evaluation DMU. Let v_i represent the weights factors related to the i-th input and u_r the weights assigned to outputs belong to O and u_r^1 and u_r^2 are the wrights factors related to y_{kj}^1 and y_{kj}^2 respectively, so the dual form of model (1) is formulated as follow:

$$
\begin{aligned}
&\text{Min} \quad \sum\nolimits_{i\in I} v_i x_{io} + v_o \\
&\text{s.t.} \ \sum\nolimits_{r\in O} u_r y_{ro} + \sum\nolimits_{k\in O^N} u_k^1 y_{ko}^1 - \sum\nolimits_{k\in O^N} u_k^2 y_{ko}^2 = 1, \\
&\sum\nolimits_{i\in I} v_i x_{ij} - \sum\nolimits_{r\in O} u_r y_{rj} - \sum\nolimits_{k\in O^N} u_k^1 y_{kj}^1 + \sum\nolimits_{k\in O^N} u_k^2 y_{ko}^2 + v_o \geq 0, \\
&\forall j \in J \\
&u_r \geq 0, \qquad\qquad \forall r \in O \\
&u_k^1, u_k^2 \geq 0 \qquad\quad \forall k \in O^N \\
&v_i \geq 0 \qquad\qquad\quad \forall i \in I \\
&v_o \ \textit{is free in sign.}
\end{aligned}
\tag{2}
$$

The optimal value of model (2) represents the efficiency score of DMU_O. DMU_O is called efficient if the optimal value of model (2) is equal to 1 and it is called inefficient if the optimal value of model (2) is greater than 1.

2.1 Equality Constraints in the Multiplier SORM Model and an Equivalent Model

One of the challenges in constructing a tractable robust counterpart for some of the DEA models is the existence of equality constraints in some of the models. In the multiplier form of the output oriented SORM model the constraint (2.1) is in equality form, hence if the outputs variables are assumed to be uncertain, this constraint may restrict the feasible region or results in a non-feasible solution for the robust counterpart of model (2) [21]. In order to avoid such complications one can assume that only input variables are under uncertainty to avoid the problems caused by equality constraint related to output variables which is not an appropriate way. Therefore, we develop an equivalent model to model (2), without any equality constraint which is more applicable in cases where both inputs and outputs variables are subjected to uncertainty or in our case to handle the uncertainty in output variables in an output-oriented model. In this section we modify an alternative formulation to convert the equality constraint to inequality. Towards this end, suppose the equality constraint (2.1) is fixed at an arbitrary positive parameter, thus model (2) can be reformulated as follows:

$$
\begin{aligned}
&\text{Min} \ \sum\nolimits_{i\in I} v_i x_{io} + v_o \ \text{s.t.} \sum\nolimits_{r\in O} u_r y_{ro} + \sum\nolimits_{k\in O^N} u_k^1 y_{ko}^1 - \sum\nolimits_{k\in O^N} u_k^2 y_{ko}^2 = t, \\
&\sum\nolimits_{i\in I} v_i x_{ij} - \sum\nolimits_{r\in O} u_r y_{rj} - \sum\nolimits_{k\in O^N} u_k^1 y_{kj}^1 + \sum\nolimits_{k\in O^N} u_k^2 y_{ko}^2 + v_o \geq 0, \\
&\forall j \in J \\
&u_r \geq 0, \qquad\qquad \forall r \in O
\end{aligned}
$$

$$
\begin{aligned}
& u_k^1, u_k^2 \geq 0 && \forall k \in O^N \\
& v_i \geq 0 && \forall i \in I \\
& v_o \textit{ is free in sign.}
\end{aligned} \tag{3}
$$

Proposition 1. Model (3) is equivalent to SORM model.

Proof. It is clear that $\left(u_r^*, u_k^{1*}, u_k^{2*}, v_i^*, v_0^*\right)$ is an optimal solution for the SORM model if and only if $\left(tu_r^*, tu_k^{1*}, tu_k^{2*}, tv_i^*, tv_0^*\right)$ is an optimal solution of model (3). Therefore, it can be concluded that model (3) is equivalent to SORM model.

Theorem 1. SORM model is equivalent to the following model:

$$
\begin{aligned}
& \text{Min} \quad \sum\nolimits_{i \in I} v_i x_{io} + v_o \\
& \text{s.t.} \quad \sum\nolimits_{r \in O} u_r y_{ro} + \sum\nolimits_{k \in O^N} u_k^1 y_{ko}^1 - \sum\nolimits_{k \in O^N} u_k^2 y_{ko}^2 \geq t, \\
& \sum\nolimits_{i \in I} v_i x_{ij} - \sum\nolimits_{r \in O} u_r y_{rj} - \sum\nolimits_{k \in O^N} u_k^1 y_{kj}^1 + \sum\nolimits_{k \in O^N} u_k^2 y_{ko}^2 + v_o \geq 0, \\
& \forall j \in J \\
& u_r \geq 0, \qquad \forall r \in O \\
& u_k^1, u_k^2 \geq 0 \qquad \forall k \in O^N \\
& v_i \geq 0 \qquad \forall i \in I \\
& v_o \textit{ is free in sign.}
\end{aligned} \tag{4}
$$

Proof. By proposition 1, model (3) is equivalent to SORM model, hence it is sufficient to prove that the dual of model (3) is equivalent to the dual of model (4). The dual form of model (3) is the following mathematical programming:

$$
\begin{array}{lll}
\text{Max } t\varphi_o & & \\
\text{s.t. } \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}, & \forall i \in I & (5.1) \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq \varphi_o y_{ro}, & \forall r \in O & (5.2) \\
\sum_{j=1}^{n} \lambda_j y_{kj}^1 \geq \varphi_o y_{kj}^1, & \forall k \in O^N & (5.3) \\
\sum_{j=1}^{n} \lambda_j y_{kj}^2 \leq \varphi_o y_{kj}^2, & \forall k \in O^N & (5.4) \\
\sum_{j=1}^{n} \lambda_j = 1, & \forall j \in J & (5.5) \\
\lambda_j \geq 0 & & (5.6) \\
\varphi_o \textit{ is free in sign} & & (5.7)
\end{array} \tag{5}
$$

and the dual of model (4) is as follows:

$$
\begin{array}{ll}
\text{Max } \ t\varphi_o & \\
\text{s.t. } \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}, & \forall i \in I \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq \varphi_o y_{ro}, & \forall r \in O \\
\sum_{j=1}^{n} \lambda_j y_{kj}^1 \geq \varphi_o y_{kj}^1, & \forall k \in O^N \\
\sum_{j=1}^{n} \lambda_j y_{kj}^2 \leq \varphi_o y_{kj}^2, & \forall k \in O^N \\
\sum_{j=1}^{n} \lambda_j = 1, & \forall j \in J \\
\lambda_j \geq 0 & \\
\varphi_o \geq 0 &
\end{array} \tag{6}
$$

Let φ_o^* be the optimal solution of model (5) and (6). The only difference between the dual models is the sign of φ_o^* which is free in sign in (5) and non-negative in (6). It is straightforward to show that $\varphi_o^* > 0$ in (5) since if $\varphi_o^* \le 0$, then $\lambda^* = 0_n$ in constraints (5.2), (5.3) and (5.4) which contradicts with the convexity constraint and violate it, that is impossible. Therefore, it can be concluded that the dual models (5) and (6) are equivalent and subsequently their primal models (3) and (4) are equivalent.

As mentioned previously in order to avoid the challenges regarding the normalization constraint, various models in the RDEA literature adopted the input-oriented models when output variables are assumed to be uncertain and similarly the output-oriented model are adopted when input variable are uncertain. One of the advantages of model (4) is that the constraints in this model are all in the form of inequality which make it suitable to be applied in cases when uncertain inputs and outputs appear simultaneously. Since in model (4), the parameter t is ab arbitrary positive number, therefore it can be considered to set $t = 1$ and the following model (7) is a suitable model to handle uncertainty in output variables in an output-oriented model in the presence of negative data.

$$
\begin{aligned}
&(SORM_E)\quad \text{Min} \sum\nolimits_{i \in I} v_i x_{io} + v_o \\
&\text{s.t.} \sum\nolimits_{r \in O} u_r y_{ro} + \sum\nolimits_{k \in O^N} u_k^1 y_{ko}^1 - \sum\nolimits_{k \in O^N} u_k^2 y_{ko}^2 \ge t, \\
&\sum\nolimits_{i \in I} v_i x_{ij} - \sum\nolimits_{r \in O} u_r y_{rj} - \sum\nolimits_{k \in O^N} u_k^1 y_{kj}^1 + \sum\nolimits_{k \in O^N} u_k^2 y_{ko}^2 + v_o \ge 0, \\
&\forall j \in J \\
&u_r \ge 0, \qquad\qquad \forall r \in O \\
&u_k^1, u_k^2 \ge 0 \qquad\quad \forall k \in O^N \\
&v_i \ge 0 \qquad\qquad\; \forall i \in I \\
&v_o \textit{ is free in sign}.
\end{aligned}
\tag{7}
$$

3 Robust Counterpart of the $SORM_E$ Model

In this section we modify a robust counterpart for the $SORM_E$ model based on Ben-Tal and Nemirovski's approach [15] when uncertain data change in a box-ellipsoidal uncertainty set. Without loss of generality, we assume input variables are certain and accurate and only output variables are subjected to uncertainty. Let J_j^y be the set of uncertain non-negative outputs $\left(y_{rj}\right)$ and $J_j^{y^1}$ and $J_j^{y^2}$ be the set of uncertain outputs that are non-negative valued for some DMUs (y_{kj}^1) and negative valued for the other DMUs (y_{kj}^2), in j-th constraint. The uncertain outputs are expressed as $\tilde{y}_{rj} = y_{rj} + \eta_{rj}\varepsilon^y y_{rj}$, $\tilde{y}_{kj}^1 = y_{kj}^1 + \eta_{kj}^1 \varepsilon^{y^1} y_{kj}^1$ and $\tilde{y}_{kj}^2 = y_{kj}^2 + \eta_{kj}^2 \varepsilon^{y^2} y_{kj}^2$ where ε^y, ε^{y^1} and ε^{y^2} are the level of perturbations in outputs and η_{rj}, η^1_{kj} and η^2_{kj} are the scale deviation from the nominal value. The uncertain outputs change in a box-ellipsoidal uncertainty set defined as follows:

$$\mathcal{U}(1,\varphi) = \left\{ \begin{array}{ll} \left|\eta_{rj}\right| \le 1, & \sqrt{\sum_{j\in J_i} \eta^2_{rj}} \le \varphi^y_j \quad \forall r \in J^y_j \\ \left|\eta^1_{kj}\right| \le 1, & \sqrt{\sum_{j\in J_i} \left(\eta^1_{kj}\right)^2} \le \varphi^{y^1}_j \ \ \forall k \in J^{y^1}_j \\ \left|\eta^2_{kj}\right| \le 1, & \sqrt{\sum_{j\in J_i} \left(\eta^2_{kj}\right)^2} \le \varphi^{y^2}_j \ \ \forall k \in J^{y^2}_j \\ & \end{array} \right\}$$

where, parameters φ^y_j, $\varphi^{y^1}_j$ and $\varphi^{y^2}_j$ are the lengths of the semi-axes of the ellipsoid for the uncertain outputs that control the size of ellipsoid and the level of reliability. The level of $\varphi_j \le \left(\left|J^y_j\right| + \left|J^{y^1}_j\right| + \left|J^{y^2}_j\right|\right)^{0.5}$ adjust the level of protection against uncertainty and $\varphi_j = 0$ means the model is not protected against uncertainty and the robust counterpart is reduced to the $SORM_E$ model. Note $\varphi_j = \left(\left|J^y_j\right| + \left|J^{y^1}_j\right| + \left|J^{y^2}_j\right|\right)^{0.5}$ implies the smallest volume of ellipsoid contained in the box and it is the highest allowable level of conservatism that the decision maker can consider for the j-th constraint and $\varphi_j = 1$ implies the largest volume of ellipsoid contained in the box. Considering the uncertainty set $\mathcal{U}(1,\varphi)$, the robust counterpart of the $SORM_E$ based on Ben-Tal and Nemirovski's approach [15] can be formulated as the following mathematical programming:

$$\begin{aligned}
&\text{(RSORM) Min} \sum\nolimits_{i\in I} v_i x_{io} + v_o \\
&\sum\nolimits_{r\in O} u_r y_{ro} + \sum\nolimits_{k\in O^N} u^1_k y^1_{ko} - \sum\nolimits_{k\in O^N} u^2_k y^2_{ko} \\
&- \sum\nolimits_{r\in J^y_o} \alpha_{ro}\hat{y}_{ro}\varphi^y_o \sqrt{\sum\nolimits_{r\in J^y_o} \hat{y}^2_{ro} z^2_{ro}} - \sum\nolimits_{k\in J^{y^1}_o} \beta_{ko}\hat{y}^1_{ko} - \varphi^{y^1}_o \sqrt{\sum\nolimits_{k\in J^{y^1}_o} \hat{y}^{1^2}_{ko} z^{1^2}_{ko}} \\
&- \sum\nolimits_{k\in J^{y^2}_o} \gamma_{ko}\hat{y}^2_{ko} - \varphi^{y^2}_o \sqrt{\sum\nolimits_{k\in J^{y^2}_o} \hat{y}^{2^2}_{ko} z^{2^2}_{ko}} \ge 1, \\
&\sum\nolimits_{i\in I} v_i x_{ij} - \sum\nolimits_{r\in O} u_r y_{rj} - \sum\nolimits_{k\in O^N} u^1_k y^1_{kj} + \sum\nolimits_{k\in O^N} u^2_k y^2_{kj} + v_o - \sum\nolimits_{r\in J^y_j} \alpha_{rj}\hat{y}_{rj} \\
&- \varphi^y_j \sqrt{\sum\nolimits_{r\in J^y_j} \hat{y}^2_{rj} z^2_{rj}} - \sum\nolimits_{k\in J^{y^1}_j} \beta_{kj}\hat{y}^1_{kj} - \varphi^{y^1}_j \sqrt{\sum\nolimits_{k\in J^{y^1}_j} \hat{y}^{1^2}_{kj} z^{1^2}_{kj}} - \sum\nolimits_{k\in J^{y^2}_j} \gamma_{kj}\hat{y}^2_{kj} \\
&- \varphi^{y^2}_j \sqrt{\sum\nolimits_{k\in J^{y^1}_j} \hat{y}^{2^2}_{kj} z^{2^2}_{kj}} \ge 0, \\
&- \alpha_{rj} \le u_r - z_{rj} \le \alpha_{rj} && \forall r \in J^y_j \\
&- \beta_{kj} \le u^1_k - z^1_{kj} \le \beta_{kj} && \forall k \in J^{y^1}_j \\
&- \gamma_{kj} \le u^2_k - z^2_{kj} \le \gamma_{kj} && \forall k \in J^{y^2}_j \\
&u_r \ge 0 && \forall r \in O \\
&u^1_k, u^2_k \ge 0 && \forall k \in O^N \\
&v_i \ge 0 && \forall i \in I \\
&v_o \text{ is free in sign} &&
\end{aligned} \tag{8}$$

where α_{rj}, β_{kj}, γ_{kj}, z_{rj}, z_{kj}^1 and z_{kj}^2 are auxiliary decision variables. The probability that j-th constraint in model (8) is violated is at most $\exp\left(-\frac{\varphi_j^2}{2}\right)$. The robust counterpart of the SORM model under a box-ellipsoidal uncertainty set lead to a second order quadratic programming and it is non-linear, however its formulation is practically tractable.

4 Application

DEA is one of most popular approaches for the assessment of efficiency in banking system. As the banks play a significant role in the economic development, evaluating the efficiency of banks provides valuable information regarding the future decisions made for the banks in order to improve the efficiency scores of banks and as a result an economic growth for the banks can be expected. Generally, there are two approaches to identify the input and output variables in the assessment of efficiency in banks: the intermediation and production approaches [22]. The intermediation approach uses monetary measure such as capital and labor as inputs and loans and profit as outputs. On the other hand, the production approach considers banks as producers and uses physical inputs such as number of staff and capital to produces services as outputs such as loans.

Several research have been conducted to evaluate the efficiency of banks in Malaysia based on different approaches. Omar et al. [23] adopted the intermediary approach for input and output selection and evaluated the efficiency of 11 commercial banks in Malaysia using a CRS and VRS CCR model and investigated the change of productivity for the banks using a Malmquist index. Tahir et al. [24] employed the intermediation approach to identify inputs and outputs and evaluated the efficiency of 23 banks in Malaysia during the period of 2000–2006 using the basic DEA models. Ab Rahim et al. [25] applied DEA models to estimate the cost efficiency of 10 domestic banks in Malaysia during the period of 1995–2010. Echchabi [26] provided a review of the previous work on the efficiency evolution of banks in Malaysia and stated that there are limited studies on this area in developing countries like Malaysia in comparison with developed countries. They adopted the intermediary approach to choose the variables and estimate the efficiency of 23 Malaysian banks and examine the factors that has an impact of the efficiency score of the banks.

However, the previous studies considered the input and output as certain and non-negative variables. In fact, in the presence of profit as an output in the evaluation process it is most likely that some of DMUs experience a loss instead of profit in a specified period. Moreover, uncertainty in data is an unavoidable factor in the efficiency assessment of any organizations such as banks and it can be due to errors in computation and measurement etc. Utilizing the intermediary approach, in this study we determine the efficiency of 30 banks in Malaysia with three inputs (total assets, deposit, total equity) and two outputs (loans and profit). The SORM models is applied to cope with negative data and the proposed RSORM to handle the uncertainty in in both negative and positive variables. we assume that outputs variables are subject to uncertainty. A descriptive statistic for data sets is given in Table 1. Firstly, the efficiency scores of DMUs are evaluated applying the SORM model and the proposed RSORM model for two different level of perturbation in data (0.01 and 0.05). The result of efficiency score and ranking of DMUs are presented

Table 1. Descriptive statistics for data sets

Data set	Mean	Standard Deviation	Minimum	Maximum
Inputs				
Total assets	85938662	143794964.8	430897	602354899
Total deposits	51485240	94965298.9	101430	410839559
Total equity	11143709	17169911.37	327053	72266256
Outputs				
Loans	37667592	79757727.26	32561	365844401
Profit	630273	1299502.742	−70405	5965127

Table 2. Optimal solutions obtained from SORM Model and RSORM model ($\varphi_j = 1.4$)

Banks	SORM	Ranking	RSORM $\varepsilon = 0.01$	Ranking	RSORM $\varepsilon = 0.05$	Ranking	Banks	SORM	Ranking	RSORM $\varepsilon = 0.01$	Ranking	RSORM $\varepsilon = 0.05$	Ranking
1	1	1	1.02	1	1.11	1	16	1.05	4	1.07	4	1.16	4
2	1	1	1.02	1	1.11	1	17	1	1	1.02	1	1.58	12
3	1	1	1.02	1	1.11	1	18	1	1	1.02	1	1.11	1
4	1.56	12	1.60	14	1.73	16	19	1.42	10	1.45	12	1.63	14
5	1	1	1.02	1	1.11	1	20	1.06	5	1.08	5	1.17	5
6	1.22	9	1.24	9	1.79	17	21	1	1	1.02	1	1.11	1
7	1.54	11	2.13	17	2.43	20	22	1	1	1.56	13	1.80	18
8	1.09	6	1.14	7	1.21	6	23	1	1	1.02	1	1.47	10
9	1.02	2	1.04	2	1.13	2	24	1.15	8	1.87	15	2.18	19
10	1	1	1.41	11	1.59	13	25	1.05	4	1.07	4	1.64	15
11	1	1	1.02	1	1.11	1	26	1	1	1.02	1	1.11	1
12	1	1	1.02	1	1.11	1	27	1	1	2.10	16	2.58	21
13	1	1	1.30	10	1.45	9	28	1	1	1.15	2	1.42	8
14	1.04	3	1.06	2	1.15	3	29	1.42	10	1.45	12	1.57	11
15	1	1	1.02	1	1.11	1	30	1.11	7	1.13	6	1.23	7

in Table 2. The results in Table 2 are obtained for the highest level of conservativeness $\varphi_j = \left(\left|J_j^{y}\right| + \left|J_j^{y^1}\right| + \left|J_j^{y^2}\right|\right)^{0.5} \cong 1.4$ which means the outputs variables are fully protected against uncertainty. In Table 3, a comparison of efficiency score for different level of conservativeness and different level of perturbation in data is given.

The obtained results in Table 2 shows that by applying SORM model most of the banks are efficient and the uncertainty in output variables can significantly change the

Table 3. RSORM optimal value for different level of φ_j

Banks	$\varphi_j = 0.5\ \varepsilon = 0.01$	$\varphi_j = 0.5\ \varepsilon = 0.05$	$\varphi_j = 1\ \varepsilon = 0.01$	$\varphi_j = 1\ \varepsilon = 0.05$	$\varphi_j = 1.4\ \varepsilon = 0.01$	$\varphi_j = 1.4\ \varepsilon = 0.05$	Banks	$\varphi_j = 0.5\ \varepsilon = 0.01$	$\varphi_j = 0.5\ \varepsilon = 0.05$	$\varphi_j = 1\ \varepsilon = 0.01$	$\varphi_j = 1\ \varepsilon = 0.05$	$\varphi_j = 1.4\ \varepsilon = 0.01$	$\varphi_j = 1.4\ \varepsilon = 0.05$
1	1.01	1.05	1.02	1.02	1.02	1.02	16	1.06	1.10	1.07	1.07	1.07	1.07
1	1.01	1.05	1.02	1.02	1.02	1.02	17	1.01	1.05	1.02	1.02	1.02	1.02
3	1.01	1.05	1.02	1.02	1.02	1.02	18	1.01	1.03	1.02	1.02	1.02	1.02
4	1.58	1.64	1.60	1.60	1.60	1.60	19	1.44	1.49	1.45	1.45	1.45	1.45
5	1.01	1.03	1.01	1.02	1.02	1.02	20	1.07	1.12	1.08	1.08	1.08	1.08
6	1.23	1.28	1.24	1.24	1.24	1.24	21	1.01	1.05	1.02	1.02	1.02	1.02
7	1.56	1.62	1.58	2.13	2.13	2.13	22	1.01	1.05	1.02	1.56	1.56	1.56
8	1.10	1.14	1.11	1.14	1.14	1.14	23	1.01	1.05	1.02	1.02	1.02	1.02
9	1.03	1.07	1.04	1.04	1.04	1.04	24	1.16	1.20	1.17	1.87	1.87	1.87
10	1.01	1.05	1.02	1.41	1.41	1.41	25	1.06	1.10	1.07	1.07	1.07	1.07
11	1.01	1.05	1.02	1.02	1.02	1.02	26	1.01	1.05	1.02	1.02	1.02	1.02
12	1.01	1.05	1.02	1.02	1.02	1.02	27	1.01	1.05	1.02	2.10	2.10	2.10
13	1.01	1.05	1.02	1.30	1.30	1.30	28	1.01	1.05	1.02	1.15	1.15	1.15
14	1.05	1.09	1.06	1.06	1.06	1.06	29	1.43	1.49	1.45	1.45	1.45	1.45
15	1.01	1.05	1.02	1.02	1.02	1.02	30	1.12	1.17	1.13	1.13	1.13	1.13

efficiency score and the ranking of DMUs. For example, the ranking of Bank 27 by SORM model is reported as 1 and by RSORM ($\varepsilon = 0.01$) is 16 and by RSORM ($\varepsilon = 0.05$) is 21. The result for RSORM model in Table 2 are reported for the full protection of uncertain outputs ($\varphi_j = 1.4$). The optimal solution by RSORM model increases by increasing the level of perturbation in output variables from 0.01 to 0.05. The level of protection indicates the level of risk that decision makers are willing to allow and in this case any level of $\varphi_j \leq 1.4$ can be considered based on the preference of decision makers. Table 3 compares the obtained optimal solutions for two different level of perturbation ($\varepsilon = 0.01, 0.05$) and three different level of conservativeness ($\varphi_j = 0.5, 1, 1.4$). It shows that by increasing the level of conservativeness the optimal solution of the RSORM will be increased which indicates that the banks will become less efficient. Decision makers can apply SORM or RSORM based on the result they expect to achieve. The result shows that the SORM model is less discriminative than the RSORM model and if the decision makers are willing to accept a certain level of uncertainty in data, the RSORM model can be a better model to assess the efficiency of banks.

5 Conclusion

The input and output variables in the basic DEA models are assumed to be non-negative. However, in many real-world applications negative data are also involved in set, any type of data can be affected by uncertainty. Robust optimization is one of the popular approaches to cope with data uncertainty in DEA models. This paper focuses on the SORM model which is one of the popular DEA models to handle negative data and constructing a robust counterpart for the SORM model to ensure feasible and robust solutions. The presence of equality constraints can be problematic for constructing a robust counterpart, as such constraints may lead to a restricted feasible region or infeasible solutions. Therefore, we showed that the equality constraint can be replace by inequality and an equivalent SORM model ($SORM_E$) is presented. A robust counterpart for $SORM_E$ model is constructed under a box-ellipsoidal uncertainty set to handle the uncertainty in both positive and negative data. The proposed RSORM model can be applied in complex case studies where uncertain negative data are involved in the evolution of DMUs. In this study the proposed RSORM model is applied for assessing the efficiency of 30 banks in Malaysia in the financial year 2020 to take into consideration the uncertainty in outputs where these uncertain outputs can take both positive and negative data. Since the previous studies on assessing the efficiency of banks in Malaysia ignored the effect of uncertainty in data and also the presence of negative data, therefore the proposed RSORM model provide further opportunities for researchers to apply the proposed RSORM model and the methodological grounds of this research for future studies on different cases such as supply chain or sustainability analysis.

References

1. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision-making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978). https://doi.org/10.1016/0377-2217(78)90138-8
2. 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). https://doi.org/10.1287/mnsc.30.9.1078
3. Liu, J.S., Lu, L.Y., Lu, W.M., Lin, B.J.: A survey of DEA applications. Omega **41**(5), 893–902 (2013). https://doi.org/10.1016/j.omega.2012.11.004
4. Emrouznejad, A., Yang, G.L.: A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socioecon. Plann. Sci. **61**, 4–8 (2018). https://doi.org/10.1016/j.seps.2017.01.008
5. Seiford, L.M., Zhu, J.: Modeling undesirable factors in efficiency evaluation. Eur. J. Oper. Res. **142**, 16–20 (2002). https://doi.org/10.1016/S0377-2217(01)00293-4
6. Portela, M.C.A.S., Thanassoulis, E., Simpson, G.: A directional distance approach to deal with negative data in DEA: an application to bank branches. J. Oper. Res. Soc. **55**(10), 1111–1121 (2004). https://doi.org/10.1057/palgrave.jors.2601768
7. Sharp, J.A., Liu, W.B., Meng, W.: A modified slacks-based measure model for data envelopment analysis with natural negative outputs and inputs. J. Oper. Res. Soc. **58**(12), 1672-1677 (2007). https://doi.org/10.1057/palgrave.jors.2602318
8. Emrouznejad, A., Anouze, A.L., Thanassoulis, E.: A semi-oriented radial measure for measuring the efficiency of decision making units with negative data using DEA. Eur. J. Oper. Res. **200**(1), 297–304 (2010). https://doi.org/10.1016/j.ejor.2009.01.001

9. Kaffash, S., Kazemi Matin, R., Tajik, M.: A directional semi-oriented radial DEA measure: an application on financial stability and the efficiency of banks. Ann. Oper. Res. **264**(1–2), 213–234 (2017). https://doi.org/10.1007/s10479-017-2719-5
10. Lin, R., Liu, Y.: Super-efficiency based on the directional distance function in the presence of negative data. Omega **85**, 26–34 (2019). https://doi.org/10.1016/j.omega.2018.05.009
11. Sengupta, J.K.: A fuzzy systems approach in data envelopment analysis. Comput. Math. Appl. **24**(8–9), 259–266 (1992). https://doi.org/10.1016/0898-1221(92)90203-T
12. Despotis, D.K., Smirlis, Y.G.: Data envelopment analysis with imprecise data. Eur. J. Oper. Res. **140**(1), 24–36 (2002). https://doi.org/10.1016/S0377-2217(01)00200-4
13. 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). https://doi.org/10.1016/j.enpol.2008.08.004
14. Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper. Res. **21**(5), 1154–1157 (1973). https://doi.org/10.1287/opre.21.5.1154
15. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. Math. Program. **88**(3), 411–424 (2000). https://doi.org/10.1007/PL00011380
16. Bertsimas, D., Sim, M.: The price of robustness. Oper. Res. **52**(1), 35–53 (2004). https://doi.org/10.1287/opre.1030.0065
17. Sadjadi, S.J., Omrani, H., Abdollahzadeh, S., Alinaghian, M., Mohammadi, H.: A robust super-efficiency data envelopment analysis model for ranking of provincial gas companies in Iran. Expert Syst. Appl. **38**(9), 10875–10881 (2011). https://doi.org/10.1016/j.eswa.2011.02.120
18. Shirazi, F., Mohammadi, E.: Evaluating efficiency of airlines: a new robust DEA approach with undesirable output. Res. Transp. Bus. Manag. **33**, 100467 (2019). https://doi.org/10.1016/j.rtbm.2020.100467
19. Dehnokhalaji, A., Khezri, S., Emrouznejad, A.: A box-uncertainty in DEA: a robust performance measurement framework. Expert Syst. Appl. **187**, 115855 (2022). https://doi.org/10.1016/j.eswa.2021.115855
20. Peykani, P., Mohammadi, E., Saen, R.F., Sadjadi, S.J., Rostamy-Malkhalifeh, M.: Data envelopment analysis and robust optimization: a review. Expert Syst. **37**(4), e12534 (2020). https://doi.org/10.1111/exsy.12534
21. Gorissen, B.L., Yanıkoğlu, İ., den Hertog, D.: A practical guide to robust optimization. Omega **53**, 124–137 (2015). https://doi.org/10.1016/j.omega.2014.12.006
22. Emrouznejad, A., Anouze, A.L.: Data envelopment analysis with classification and regression tree–a case of banking efficiency. Expert. Syst. **27**(4), 231–246 (2010). https://doi.org/10.1111/j.1468-0394.2010.00516.x
23. Omar, M.A., Rahman, A.R.A., Yusof, R.M., Rasid, M.E.S.M.: Efficiency of commercial banks in Malaysia. Asian Acad. Manag. J. Account. Fina. **2**(2), 19–42 (2006)
24. Tahir, I.M., Bakar, N.A., Haron, S.: Evaluating efficiency of Malaysian banks using data envelopment analysis. Int. J. Bus. Manage. **4**(8), 96–106 (2009). https://doi.org/10.5539/ijbm.v4n8p96
25. Ab-Rahim, R., Md-Nor, N.G. Ramlee, S., Ubaidillah, N.Z.: Determinants of cost efficiency in Malaysian banking. Int. J. Bus. Soc. **13**(3), 355 (2012)
26. Echchabi, A., Olaniyi, O.N., Ayedh, A.M.: Assessing the efficiency of Malaysian banks: a data envelopment analysis approach. Afro-Asian J. Financ. Acc. **5**(1), 56–69 (2015)