Asieh GHOLAMINEZHAD, PhD Candidate
asiehgholaminezhad@gmail.com
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

Mohsen ROSTAMY-MALKHALIFEH, PhD (corresponding author)
mohsen_rostamy@yahoo.com
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

Farhad HOSSEINZADEH LOTFI, PhD
farhad@hosseinzdeh.ir
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

Tofigh ALLAHVIRANLOO, PhD
tofigh.allahviranloo@istinye.edu.tr
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran
Research Center of Performance and Productivity Analysis, Istinye University, Istanbul, Türkiye

Integrating of Data Envelopment Analysis and Discriminant Analysis to Predict the Bankruptcy of Firms: Application in Electricity Industry

Abstract. *Bankruptcy represents a pivotal aspect of financial management, signifying the ultimate phase in the economic life cycle of firms and bearing substantial ramifications for a multitude of stakeholders, including shareholders, creditors, employees, and the broader economic system. It is therefore imperative to be able to predict bankruptcy in order to mitigate risks and implement timely interventions. In recent years, researchers have developed a plethora of models designed to predict bankruptcy. These encompass a range of methodologies, from statistical techniques to machine learning approaches. Among these, discriminant analysis and data envelopment analysis have been widely employed due to their simplicity and efficiency in assessing financial health. Furthermore, the combination of these methods has been investigated with a view to enhancing prediction accuracy. In this study, we propose a novel algorithm that integrates the strengths of data envelopment analysis and a hybrid data envelopment analysis-discriminant analysis approach. This algorithm provides a robust framework for analysing and evaluating bankruptcy, offering new insights into the predictive capabilities of these techniques and their practical applications in financial decision-making. The performance of the proposed algorithm is illustrated on a real data set related to corporate bankruptcy in the electricity industry.*

Keywords: b*ankruptcy, data envelopment analysis, discriminant analysis, slack sased measure*

JEL Classification: G33, D61, J16, C44.

DOI: 10.24818/18423264/59.1.25.02

1. Introduction

Given the pivotal role of corporate success and failure in the financial sector, scholars have offered a range of definitions of financial bankruptcy. In essence, bankruptcy is intrinsically linked to the fundamental and primary objective of any business, which is to survive. The occurrence of bankruptcy can have a significant impact on a diverse range of stakeholders, including those in managerial, investor, creditor, competitive, and legal capacities. Of the aforementioned groups, investors and creditors are more likely to experience adverse effects in the event of bankruptcy, and thus exhibit a greater propensity to anticipate such an outcome. Newton (Newton, 1998) delineated the stages of an unfavourable financial situation for a firm, which he defined as follows: latency, cash deficit, lack of financial or commercial solvency, lack of full solvency and, finally, bankruptcy. While most bankruptcies adhere to this sequence of events, there are instances where a firm may become insolvent without traversing the entire spectrum of these stages. Bankruptcy represents the final stage of a firm's economic life cycle and has ramifications for all stakeholders. It is, therefore, important to be able to predict it. The intensifying competitive landscape has constrained access to resources, thereby elevating the probability of bankruptcy. A suitable model for predicting bankruptcy can be provided, allowing firms to react in a timely manner and enabling investors to assess favourable opportunities from unfavourable ones. It also enables the implementation of precautionary measures to prevent potential financial losses. The losses incurred by investors from inaccurate or delayed bankruptcy predictions may be much more significant than those resulting from actual bankruptcy.

A variety of conventional and modern techniques and instruments have been proposed for the evaluation and forecasting of bankruptcy. Two papular methods that have garnered significant interest are discriminant analysis (DA) and data envelopment analysis (DEA) (Horvathova & Mokrisova, 2023). Both approaches may be classified as linear programming methods for evaluating the efficacy of a collection of homogeneous decision-making units (DMUs). In order to distinguish between units that are successful and those that are not, a weighting scheme is considered for each collection of factors that are evaluated for all DMUs. In a similar instance of DEA-based bankruptcy assessment, the authors presented a novel sample of non-parametric DAs that computed the discriminant function weights and subsequently obtained an assessment privilege, which was utilised to identify group membership. The nonparametric DA was designated "DEA-DA" to reflect its incorporation of the nonparametric property of DEA into its computational framework, which effectively preserved its resolution capabilities.

The DA was initially developed by Fisher (Fisher, 1936) as a means of employing the methodology utilised in linear regression to solve linear equations. DA is applicable when there is one grouping qualitative variable and several independent quantitative variables, and the objective is to identify relationships that can be used to determine the membership in the grouping variable according to the independent variables. DA is a technique that can be employed for the classification of units as firms, individuals, and so forth, based on a nominal variable comprising two or more facets. In general, DA comprises two or more groups. However, as Glover (1990) and numerous other researchers have demonstrated, the two-group scenario has attracted the most attention. The application of robust discriminant analysis methods has been demonstrated to be an effective approach for the management of outliers, resulting in a higher level of classification accuracy in comparison to classical techniques (Hussain & Uraibi, 2023).

In DEA-based bankruptcy analysis, firms' financial ratios are not compared to a specific ratio or standard, such as the market average, etc. Instead, the comparison criteria are the performance of all firms relative to each other. In their research, Sueyoshi and Goto (2009a) and Premachandra et al. (2009) emphasise the significance of bankruptcy assessment, highlighting that the failure of a firm can result in significant direct and indirect costs for its stockholders. The findings of this study indicate that the DEA approach is more accurate than the logistic regression method in predicting corporate bankruptcy and is more effective than the regression method when applied to small sample sizes. Cielen et al. (2004) employed the non-radial CRR model (Charnes et al., 1978) to evaluate the efficiency of the DEA and logit regression models in predicting failure. The application of optimised thresholds in super-efficiency DEA models has been demonstrated to markedly enhance the specificity of bankruptcy predictions across a range of sectors (Staňková & Hampel, 2023). A DEA-based approach was used to evaluate the relative merits of domain knowledge features and LASSO-selected features in bankruptcy prediction models (Khalafi & Matin 2021; Yadollahi & Matin 2022; Navidi et al., 2023). A dynamic performance assessment of distress prediction models utilising DEA revealed that the incorporation of financial, market, and macroeconomic data resulted in enhanced accuracy when compared to static models, particularly with shorter training periods (Mokrišová & Horváthová, 2023). Mousavi et al. (2023), Freed and Glover (1981) conducted a comparative analysis between DEA and a programmatic DA-based mathematical method. Additionally, Cheng et al. (2006) and Ravikumar and Ravi (2007) have employed the relative model for financial analysis.

In a review of previous research on DEA-DA, Sueyoshi (1999) proposed a goal-planning (GP) approach for DA and compared it with a collective model of DEA. Sueyoshi and Kirihara (1998) provided a detailed account of how the preceding information was incorporated into the DEA-DA evaluation process. In a further development, Sueyoshi (2001) extended the DEA-DA model to accommodate negative values. The proposed approach is also based on two-step GP formulas and can be solved using any linear programming software. The initial phase of the process entails identifying the areas of overlap between the two groups of DMUs. Subsequently, the DMUs exhibiting overlap are categorised (Sueyoshi 2004, 2006).

The use of DEA-DA in urban renewal projects has been shown to facilitate improved stakeholder collaboration and reduce disagreements (Shi et al., 2024). Furthermore, the integration of DEA-DA with artificial neural networks has been demonstrated to enhance healthcare systems, particularly during pandemics, by improving patient classification and resource allocation (Yousefi et al., 2024). Other empirical studies have included the forecasting of customer group membership and the classification of customers according to the customer pyramid (Farzipoor Saen, 2013), the ranking of electricity distribution units (Tavassoli et al., 2015) and the forecasting of supplier group membership (Boudaghi & Saen, 2018).

The following section outlines the structure of the remaining study. In the initial section, an investigation is conducted on the similarities and differences between DEA, DA, and DEA-DA. Subsequently, the preceding methods of assessing bankruptcy are subjected to analysis in Section 3. In Section 4, an algorithm for determining whether a given entity is bankrupt or not is presented. An illustrative example is presented in Section 5, which demonstrates the application of the proposed algorithm. In conclusion, Section 6 presents the findings of this study.

2. The similarities and differences between DEA, DA, and DEA-DA

This section presents an investigation of the similarities and differences between DEA, DA, and DEA-DA.

2.1 A comparison of the DEA with the DA

The principal similarities and differences between DEA and DA are outlined below.

- The DA and the DEA specify a size for each DMU and then proceed to compare it to a pre-established threshold value. Moreover, both techniques construct a hyperplane with the objective of differentiating between the two groups (each unit in the DEA selects its own hyperplane).
- Both techniques incorporate a relative analogy. In both methods, a set of weights is obtained from a single problem pertaining to a specific unit. In the case of the DA, the evaluation of each new DMU is contingent upon the weights specified for previous DMUs. In DEA, a DMU is defined as an efficient DMU, that is, a DMU that is more efficient than other DMUs. Consequently, both techniques involve a relative analogy.
- Both techniques are susceptible to produce degenerate solutions. Typically, the number of constraints (decision-making units) exceeds that of the decision variables (factors).
- Both could be considered to belong to the field of goal programming.
- Both approaches categorise DMUs into two distinct groups.
- The G membership is already known in the context of DA, but not in the context of DEA.

- In DA, misclassification is permitted within each of the two groups, which may result in computational errors. In the context of DEA, there is no possibility of an erroneous classification. Consequently, any unit that is designated as inefficient is, in fact, genuinely inefficient.
- In DA, it is assumed that the DMUs are operating in the same environment, and a common weight set is selected for all DMUs, resulting in a unique hyperplane. In contrast, in DEA, not all DMUs operate in the same manner and are placed in different environments.

2.2 A comparison of the DEA with the DEA-DA

Sueyoshi and Goto (2009b) delineated the similarities and differences between the two methods as follows:

- All financial variables must be divided by the DEA into inputs and outputs. The categorisation of financial variables is not a prerequisite for DEA-DA.
- The xth weighting assessment of DEA demonstrates the degree to which the xth DMU exerts influence over the efficiency of a specific DMU. The kth weighting assessment of DEA-DA demonstrates the pivotal role of the financial variable in differentiating between the two groups.
- The selection of weights in DEA is specific to each firm, resulting in a set of weights that are unique to each firm. In contrast, the selection of weights in DEA-DA is not firm-specific, and the resulting weights are applicable to all firms (DMUs) in the industry.
- Unlike DEA, DEA-DA does not hypothesise about the positioning of DMUs below the efficient frontier. Similarly, no hypothesis is required on the distribution of groups in DEA-DA.
- In the absence of conventional statistical tests, rank-sum tests are employed for DEA. In the case of DEA-DA, no regular statistical tests are provided, but rank-sum tests are offered as an alternative.
- The DEA is unable to resolve the issue of data imbalance. The issue of data imbalance can be addressed by DEA-DA through the application of a weighting system that assigns greater significance to the data points belonging to the group under consideration.
- DEA can guarantee the optimality of the answer, but it is unable to prevent the generation of multiple responses and predictions. It is not possible for DEA-DA to guarantee the optimality of the answer. The selection of M and ε in DEA-DA affects the selected answer.
- It would be beneficial for the DEA to insert a new DMU into the original data set and recalculate the entire data set. By determining whether or not a new DMU is about to file for bankruptcy, the DEA can forecast the group membership of that DMU.
- In general, the DEA assesses the operational effectiveness of each firm. Nevertheless, the DEA may be employed to evaluate performance in relation to bankruptcy, whereby the latter may be regarded as a financial

indicator. The DEA-DA calculates the Altman Z-score for both bankrupt and non-bankrupt firms, subsequently obtaining a financial performance indicator based on the aforementioned Z-scores.

3. Description of SBM and DEA-DA Models for Bankruptcy Assessment

The SBM model (Tone, 2001) is a model that simultaneously considers both an increase in outputs and a decrease in inputs. Additionally, only auxiliary variables are taken into account when evaluating the efficiency of the units' bankruptcy.

Let us consider a set of DMUs (decision-makers), for example, firms. Every $DMU_x \left(x = 1,\ldots,n\right)$ has q input variables and t output variables. The input variable w and the output variable v from $DMU_x \left(x = 1,\ldots,n\right)$ are represented by $x_{wx} \left(w = 1,\ldots,q\right)$ and $y_{vx} \left(v = 1,...,t\right)$, respectively. It thus follows that the SBM model for DMU_0 (the DMU under evaluation) can be written as follows:

$$\rho = \min_{\eta, s^-, s^+} \frac{1 - \frac{1}{q}\sum_{w=1}^{q} \frac{s_w^-}{x_{wo}}}{1 + \frac{1}{t}\sum_{v=1}^{t} \frac{s_v^+}{y_{vo}}}$$

$$\begin{aligned} s.t \quad & \mathrm{x}_{wo} - s_w^{\ -} = \sum_{x=1}^{n} x_{wj}\eta_x \quad \left(w = 1,\ldots,q\right) \\ & \mathrm{y}_{vo} + s_v^{\ +} = \sum_{x=1}^{n} y_{vj}\eta_x \quad \left(v = 1,\ldots,t\right) \\ & \sum_{x=1}^{n} \eta_x = 1 \\ & \eta_x \geq 0 \ \left(\forall x\right), \mathrm{s}_w^{\ -} \geq 0 \ \left(\forall w\right), s_v^{\ +} \geq 0 \ \left(\forall v\right) \end{aligned} \tag{1}$$

where the vector of input/output data of DMU_0 are defined by $x_{wo} = \left(x_{1o},...,x_{qo}\right)^T$ and $y_{vo} = \left(y_{1o},...,y_{to}\right)^T$. The Slack vectors $s^- \in R^q$ and $s^+ \in R^t$ are described as input extra and output shortage, respectively.

In the DEA-DA model, we consider bankruptcy assessment in a situation that includes two groups of firms $\left(G_1, G_2\right)$. The total of both groups has n firms $\left(x = 1,...,n\right)$ that $n = n_1 + n_2$. G_1 a group consisting of non-bankrupt firms and G_2 a group of bankrupt firms. Independent financial variables $\left(k = 1,..,p\right)$ which define any organisation indicate a financial ratio. A function separating two groups is declared by $\sum_{k=1}^{p} \eta_k z_{kx}$. Here η_k is the weight for the financial variable

k. The initial phase of the process involves the identification of areas of overlap between groups of firms, followed by the categorisation of those that do not fall within the aforementioned overlap. The second phase entails the classification of the aforementioned overlapping firms. The initial phase is formulated as follows:

$$\begin{aligned} \min \quad & s \\ \text{s.t} \quad & \sum_{k=1}^{p} \eta_k z_{kx} - h + s \geq 0,\ x \in G_1 \\ & \sum_{k=1}^{p} \eta_k z_{kx} - h - s \leq -\varepsilon,\ x \in G_2 \\ & \sum_{k=1}^{p} \left|\eta_k\right| = 1 \\ & \text{h\&s} : URS \end{aligned} \tag{2}$$

The objective function (2) seeks to minimise the indefinite variable (s), which represents the extent of the overlap between G_1 and G_2. The lower range of G_1, $h-s$, and the the upper range G_2, $h+s$, have been used to define the aforementioned overlap. The value of the scalar d represents the group categorisation resolution score. Both h and s are limitless (URS).

Model (2) includes a small amount ε to distinguish between the two groups. All financial variables $\left(z_{kx}\right)$ from xth are linked together by a partition function $\left(\sum_{k=1}^{p} \eta_k z_{kx}\right)$. In addition, the weights are constrained to ensure that the total of the absolute values η_k for $k=1,...,p$ is equal to 1. Accordingly, model (2) converts each weight into a percentage, allowing us to see which weights are crucial for group classification and which are not. "Normalisation: is the term used to describe the weight constraint. The first stage is to determine the presence of overlap using $s^* \geq 0$ optimality (2). The opposite $s^* < 0$ indicates no overlap.

Model (2) has a new formulation as follows, because we cannot solve (2) directly:

$$\begin{aligned}
&\min \quad s \\
&\text{s.t} \quad \sum_{k=1}^{p}\left(\eta_k^{+}-\eta_k^{-}\right) z_{kx} - h + s \geq 0 \text{ , x} \in G_1 \\
&\qquad \sum_{k=1}^{p}\left(\eta_k^{+}-\eta_k^{-}\right) z_{kx} - h - s \leq -\varepsilon \text{ , x} \in G_2 \\
&\qquad \sum_{k=1}^{p}\left(\eta_k^{+}+\eta_k^{-}\right) = 1 \\
&\qquad \zeta_k^{+} \geq \eta_k^{+} \geq \varepsilon\zeta_k^{+} \text{ , } \zeta_k^{-} \geq \eta_k^{-} \geq \varepsilon\zeta_k^{-}, \zeta_k^{+}+\zeta_k^{-} \leq 1,\ \eta_k^{+}+\eta_k^{-} \geq \varepsilon \ \forall k, \\
&\qquad h, s \text{ : free} \\
&\qquad \zeta_k^{+} \,\&\, \zeta_k^{-} : \ 0,1, \text{ all variables} \geq 0.
\end{aligned} \tag{3}$$

First, we define $\eta_k\left(k=1,...,p\right)$ as $\eta_k = \eta_k^{+}-\eta_k^{-}$, where

$$\eta_k^{-} = \left(\left|\eta_k\right| + \eta_k\right)/2, \eta_k^{+} = \left(\left|\eta_k\right| + \eta_k\right)/2 \tag{4}$$

These even variables maintain the relationship. $\left(\eta_k = \eta_k^{+}-\eta_k^{-}\right)$ and $\left(\left|\eta_k\right| = \eta_k^{+}+\eta_k^{-}\right)$.

To avoid $\eta_k^{+} > 0$ and $\eta_k^{-} > 0$ occurring at the same time, such a transformation requires a nonlinear condition $\left(NLC : \eta_k^{+}-\eta_k^{-}\right)$ for each $\left(k=1,...,p\right)$. To add *NLC*, the equivalent of the reformulated *MIP* is used. Here, consider $\zeta^{+}\left(=0,1\right)$ and $\zeta^{-}\left(=0,1\right)$ two binary variables, then NLC is stated as follows:

$$NLC: \quad \zeta_k^{+} \geq \eta_k^{+} \geq \varepsilon\zeta_k^{+}, \quad \zeta_k^{-} \geq \eta_k^{-} \geq \varepsilon\zeta_k^{-} \quad \text{(k=1,...,p)} \tag{5}$$

$$\zeta_k^{+}+\zeta_k^{-} \leq 1 \quad \text{(k=1,...,p)} \tag{6}$$

Here, (5) shows the upper and lower bounds of η_{k}^{+} and η_{k}^{-}, respectively. In addition, (6) shows that the sum of these binary variables is less than 1. If $\eta_{\mathrm{k}}^{+} \geq \varepsilon > 0$ and $\eta_{\mathrm{k}}^{+} \geq \varepsilon > 0$ happen simultaneously in (5), then it happens in (6) as $\zeta_{\mathrm{k}}^{+}+\zeta_{\mathrm{k}}^{-} = 2$, and consequently it becomes impossible in (6). Therefore, the simultaneous existence of $\eta_{\mathrm{k}}^{+} > 0$ and $\eta_{\mathrm{k}}^{-} > 0$ is removed from the calculation of model (3). All other connections (η_{k}^{+} and η_{k}^{-}) are possible in both (5) and (6) and, therefore, in (3). At the end of the first stage, we need to divide all the firms into subcategories. To describe the category, we assume that $\eta_{\mathrm{k}}^{*} = \left(\eta_{\mathrm{k}}^{+*}-\eta_{\mathrm{k}}^{-*}\right)$, d^{*}

and s^* are an optimal solution to (3). The primary data set (G) is divided into the subgroups listed below:

$$G = C_1 \cup D_1 \cup C_2 \cup D_2$$
$$C_1 = \left\{ x \in G_1 \left| \sum_{k=1}^{p} \eta_k^* z_{kx} > h^* + s^* \right. \right\}$$
$$C_2 = \left\{ x \in G_2 \left| \sum_{k=1}^{p} \eta_k^* z_{k} < h^* - s^* \right. \right\}$$
$$D_1 = G_1 - C_1$$
$$D_2 = G_2 - C_2$$

According to the classification, we specify that firms in C_1 belong to G_1 and firms in C_2 belong to G_2. This is because these enterprises are obviously above or below an overlap identified by the model (3). Overlaps whose DMUs have not yet been identified in phase one are included in the two subsets$(D_1 \cup D_2)$.

In phase 2, the overlap $(G_1 \cap G_2)$ between the two groups means that a recently observed DMU could be owned by both groups. In this case, it is necessary to reclassify all overlapping firms $(D_1 \cup D_2)$, as it is currently unclear which firms form this group and need to be defined. The second is expressed mathematically by the following formula:

$$\begin{aligned}
\min \quad & \sum_{x \in D_1} y_x + \sum_{x \in D_2} y_x \\
\text{s.t} \quad & \sum_{k=1}^{p} \left(\eta_k^+ - \eta_k^- \right) z_{kx} - a + My_j \geq 0, \ \mathrm{x} \in D_1 \\
& \sum_{k=1}^{p} \left(\eta_k^+ - \eta_k^- \right) z_{kx} - a - My_j \leq -\varepsilon, \ \mathrm{x} \in D_2 \\
& \sum_{k=1}^{p} \left(\eta_k^+ + \eta_k^- \right) = 1 \\
& \zeta_k^+ \geq \eta_k^+ \geq \varepsilon \zeta_k^+, \zeta_k^- \geq \eta_k^- \geq \varepsilon \zeta_k^- \ , \zeta_k^+ + \zeta_k^- \leq 1, \eta_k^+ + \eta_k^- \geq \varepsilon \ \forall k, \\
& a\text{:URS} \\
& \zeta_k^+ \ \& \ \zeta_k^- : \ 0,1, \text{ all variables} \geq 0
\end{aligned} \tag{7}$$

The number of incorrectly classified firms is counted by the binary variable (y_x). The objective function is to reduce the number of incorrect classifications as much as possible. It is necessary to set a large number (M) and a small number (ε) in (7). A resolution score (a) from (7) is added as an unbounded infinite variable. The previous differentiation privilege of the first phase in (3) is replaced by the new differentiation privilege for the second phase.

Non-zero condition (NZC): One possibility that must be considered is the simultaneous occurrence of $\eta_k^+ = 0$ and $\eta_k^- = 0$ in (7). The presence of zero in even variables as a computational result (7) does not become a mathematical problem. However, to limit the number of η_k non-zero estimates, we need to add the following non-zero condition to (7):

$$NZC: \quad \eta_k^+ + \eta_k^- \geq \varepsilon \quad (\text{for some } k) \tag{8}$$

Here, (8) prevents the simultaneous occurrence of $\eta_k^+ = 0$ and $\eta_k^- = 0$ with respect to the financial variable *k*. As added in phase (3), it is allowed to have NZC (8) for all financial variables. However, in the second phase (7), the addition of (8) depends on the degree of freedom in the DMUs (firms) within an overlap.

After obtaining an optimal solution (η_k^*, a^*) from (7), the second stage categorises the firms in the overlap as follows:

If $\sum_{k=1}^{p} \eta_k^* z_{kx} \geq a^*$, then firm *k* belongs to group G_1, or if $\sum_{k=1}^{p} \eta_k^* z_{kx} \leq a^* - \varepsilon\, 0$, then this firm belongs to G_2. Thus, in the DEA-DA model, all firms in *G* are classified into one of two groups, G_1 or G_2.

4. Integration of the DEA-DA model with the efficiency obtained from the SBM model

In this step, the obtained ρ from the SBM model (Tone, 2001) was added to the model presented by Sueyoshi and Goto (2009b).

$$\begin{aligned}
&\min \quad s \\
&\text{s.t} \quad \sum_{k=1}^{p}\left(\eta_k^+ - \eta_k^-\right) z_{kx} + \left(\eta_{p+1}^+ - \eta_{p+1}^-\right)\rho_x - h + s \geq 0 \text{ , x} \in G_1 \\
&\qquad \sum_{k=1}^{p}\left(\eta_k^+ - \eta_k^-\right) z_{kx} + \left(\eta_{p+1}^+ - \eta_{p+1}^-\right)\rho_x - h - s \leq -\varepsilon \text{ , x} \in G_2 \\
&\qquad \sum_{k=1}^{p}\left(\eta_k^+ + \eta_k^-\right) + \left(\eta_{p+1}^+ + \eta_{p+1}^-\right) = 1 \\
&\qquad \zeta_k^+ \geq \eta_k^+ \geq \varepsilon\zeta_k^+ \text{ , } \zeta_k^- \geq \eta_k^- \geq \varepsilon\zeta_k^- \text{ ,} \zeta_k^+ + \zeta_k^- \leq 1 \text{ , } \eta_k^+ + \eta_k^- \geq \varepsilon \ \forall k, \\
&\qquad \zeta_{p+1}^+ \geq \eta_{p+1}^+ \geq \varepsilon\zeta_{p+1}^+, \zeta_{p+1}^- \geq \eta_{p+1}^- \geq \varepsilon\zeta_{p+1}^-, \zeta_{p+1}^+ + \zeta_{p+1}^- \leq 1, \eta_{p+1}^+ + \eta_{p+1}^- \geq \varepsilon \\
&\qquad h, s \text{ : free, } \zeta_k^+ \ \& \ \zeta_k^- \ \& \ \zeta_{p+1}^+ \ \& \ \zeta_{p+1}^- \text{ : 0,1, all variables} \geq 0
\end{aligned} \tag{9}$$

If there is an overlap, the second stage DEA-DA model (Sueyoshi & Goto, 2009b) with calculated efficiency from the SBM model (Tone, 2001) is as follows:

$$
\begin{aligned}
\min\quad & \sum_{x\in D_1} y_x + \sum_{x\in D_2} y_x \\
\text{s.t}\quad & \sum_{k=1}^{p}\left(\eta_k^+ - \eta_k^-\right) z_{kx} + \left(\eta_{p+1}^+ - \eta_{p+1}^-\right)\rho_x - a + My_j \geq 0 \text{ , x} \in D_1 \\
& \sum_{k=1}^{p}\left(\eta_k^+ - \eta_k^-\right) z_{kx} + \left(\eta_{p+1}^+ - \eta_{p+1}^-\right)\rho_x - a - My_j \leq -\varepsilon \text{ , x} \in D_2 \\
& \sum_{k=1}^{p}\left(\eta_k^+ + \eta_k^-\right) + \left(\eta_{p+1}^+ + \eta_{p+1}^-\right) = 1 \\
& \zeta_k^+ \geq \eta_k^+ \geq \varepsilon\zeta_k^+ \text{ , } \zeta_k^- \geq \eta_k^- \geq \varepsilon\zeta_k^- \text{ , } \zeta_k^+ + \zeta_k^- \leq 1,\ \eta_k^+ + \eta_k^- \geq \varepsilon\ \forall k, \\
& \zeta_{p+1}^+ \geq \eta_{p+1}^+ \geq \varepsilon\zeta_{p+1}^+, \zeta_{p+1}^- \geq \eta_{p+1}^- \geq \varepsilon\zeta_{p+1}^-, \zeta_{p+1}^+ + \zeta_{p+1}^- \leq 1,\ \eta_{p+1}^+ + \eta_{p+1}^- \geq \varepsilon \\
& a\text{:URS, } \zeta_k^+ \,\&\, \zeta_k^- \,\&\, \zeta_{p+1}^+ \,\&\, \zeta_{p+1}^- :\ 0,1, \text{ all variables} \geq 0.
\end{aligned}
\tag{10}
$$

The bankruptcy assessment algorithm is described as follows:

Step 1: Apply the SBM model to each of the considered firms.

Step 2: Using the efficiency obtained in the previous step, model (9) is run first, and if there is an overlap, model (10) is run.

Step 3: Based on the values obtained, the firms are divided into two categories: bankrupt and non-bankrupt.

5. Results and discussion

In this section, we illustrate the performance of the proposed algorithm on a real dataset of corporate bankruptcies in the US electricity industry. The data used here is from Sueyoshi (2006). The data includes 61 non-bankrupt firms (from DMU1 to DMU61) and 22 bankrupt firms (from DMU62 to DMU83). The efficiency of all firms is measured by 13 financial ratios. Non-bankrupt firms supplied electricity to the US electricity market in 2003, and all bankrupt firms faced bankruptcy between 1996 and 2002.

First, the model (3) is implemented. The results are given in Table 1.

Table 1. The obtained values from the model (3)

	η^+	η^-	ζ^+	ζ^-
k = 1	0	0.10315	0	1
k = 2	0.006352	0	1	0
k = 3	0	0.024577	0	1
k = 4	0.00001	0	1	0
k = 5	0.248452	0	1	0
k = 6	0	0.000019	0	1
k = 7	0.0123735	0	1	0
k = 8	0.0000599	0	1	0
k = 9	0.0122374	0	1	0

	η^{+}	η^{-}	ζ^{+}	ζ^{-}
$k = 10$	0	0.00044	0	1
$k = 11$	0.5293723	0	1	0
$k = 12$	0.045471	0	1	0
$k = 13$	0	0.0174856	0	1

Source: Authors' own creation.

The DMU21, DMU48, DMU56, DMU69, DMU70 and DMU83 were situated within the overlap area, and the model (7) is implemented for them. The results are presented in Table 2.

Table 2. The obtained values from the model (7)

	η^{+}	η^{-}	ζ^{+}	ζ^{-}
$k = 1$	0.00001	0	1	0
$k = 2$	0	0.00001	0	1
$k = 3$	0.00001	0	1	0
$k = 4$	0.00001	0	1	0
$k = 5$	0	0.00001	0	1
$k = 6$	0	0.0036	0	1
$k = 7$	0	0.00001	0	1
$k = 8$	0.004706	0	1	0
$k = 9$	0	0.00001	0	1
$k = 10$	0.000094	0	1	0
$k = 11$	0	0.00001	0	1
$k = 12$	0.98987	0	1	0
$k = 13$	0.00895	0	1	0

Source: Authors' own creation.

In order to obtain the performance of each DMU with the SBM model (Tone, 2001), it is necessary to divide the data into two categories: input and output. In this analysis, we consider the variables and financial ratios with a negative correlation with the bankruptcy debate as input, and those with a positive correlation as output.

Input variables: RE/TA= Retained Earning/Total Assets, NWC/TA= Net Working Capital/Total Assets, SE/TA= Shareholder Equity/Total Assets, S/TA= Sales/ Total Assets, NI/TA= Net Income/Total Assets, Mk= Market to Book Ratio, ROE= Return to Equity, PE= Price over Earning, EPS= Earnings per Share, Price= Share Price.

Output variables: C/TA= Cash/Total Assets, LTD/TA= Long-Term Debit/Total Assets, BETA= Covariance of an asset's return with the benchmark's return/ Variance of the benchmark's return over a given period.

Now, the performance of each DMU is calculated using the model (1). The results are shown in Table 3.

Then, using the efficiency obtained from model (1), model (9) is implemented; the results are presented in Table 4.

The DMU67, DMU69, and DMU83 were placed in the overlap area, and we implemented model (10) for them and the following values were obtained.

The DMU67, DMU69, and DMU83 were situated within the overlap area, and model (10) was implemented for them, resulting in the values being obtained (Table 5).

The incorporation of DMU efficiency into the model (9) results in a reduction in the number of DMUs situated within the overlap region. The incorporation of the efficiency values derived from the SBM model into the financial characteristics under consideration has the effect of enhancing the precision of the bankruptcy prediction.

Table 3. Efficiency of firms by SBM Model

DMU	ρ	DMU	ρ	DMU	ρ
DMU01	1.00	DMU31	0.00	DMU61	-44.54
DMU02	0.00	DMU32	-34.83	DMU62	1.00
DMU03	-25.65	DMU33	-10.09	DMU63	1.00
DMU04	-46.10	DMU34	-9.30E+18	DMU64	1.00
DMU05	-22.42	DMU35	-34.42	DMU65	1.00
DMU06	-30.39	DMU36	-58.20	DMU66	1.00
DMU07	-8.94E+18	DMU37	-30.47	DMU67	1.10
DMU08	0.00	DMU38	-33.52	DMU68	1.00
DMU09	-18.58	DMU39	1.00	DMU69	-64.66
DMU10	-15.18	DMU40	-6.06E+18	DMU70	-43.49
DMU11	-9.18E+18	DMU41	1.00	DMU71	0.00
DMU12	-20.52	DMU42	0.00	DMU72	0.00
DMU13	-25.02	DMU43	-25.76	DMU73	0.16
DMU14	-24.45	DMU44	0.00	DMU74	1.00
DMU15	0.52	DMU45	-27.19	DMU75	0.00
DMU16	-11.20	DMU46	0.00	DMU76	0.34
DMU17	-33.53	DMU47	-28.73	DMU77	1.00
DMU18	-121.97	DMU48	-44.27	DMU78	1.00
DMU19	-47.75	DMU49	-26.42	DMU79	0.00
DMU20	0.00	DMU50	-7.39E+18	DMU80	-3.92
DMU21	0.00	DMU51	-9.29E+18	DMU81	1.00
DMU22	-53.11	DMU52	-2.01E+19	DMU82	1.00
DMU23	-24.97	DMU53	-26.39	DMU83	0.00

DMU	**ρ**	**DMU**	**ρ**	**DMU**	**ρ**
DMU24	1.00	DMU54	-19.08		
DMU25	-25.96	DMU55	-36.38		
DMU26	-80.15	DMU56	1.00		
DMU27	-29.98	DMU57	-27.82		
DMU28	-46.67	DMU58	-16.04		
DMU29	-22.86	DMU59	-56.06		
DMU30	-9.68	DMU60	-39.18		

Source: Authors' own creation.

Table 4. The obtained values from the model (9)

	η^+	η^-	ζ^+	ζ^-
$k=1$	0	0.050477	0	1
$k=2$	0.002505	0	1	0
$k=3$	0	0.103091	0	1
$k=4$	0.240773	0	1	0
$k=5$	0.161737	0	1	0
$k=6$	0.000001	0.000009	0	1
$k=7$	0	0.00001	0	0
$k=8$	0.000099	0	1	0
$k=9$	0.0044	0	1	0
$k=10$	0.000119	0	1	0
$k=11$	0.384841	0	1	0
$k=12$	0.035993	0	1	0
$k=13$	0	0.015935	0	1
ρ	0.000005	0.000005	1	0

Source: Authors' own creation.

Table 5. The obtained values from the model (10)

	η^+	η^-	ζ^+	ζ^-
$k=1$	0.99999	0	1	0
$k=2$	0.00001	0	1	0
$k=3$	0	0.00001	0	1
$k=4$	0.00001	0	1	0
$k=5$	0	0.00001	0	1
$k=6$	0	0.00001	0	1
$k=7$	0.00001	0	1	0

	η^{+}	η^{-}	ζ^{+}	ζ^{-}
$k=8$	0	0.00001	0	1
$k=9$	0	0.00001	0	1
$k=10$	0.00001	0	1	0
$k=11$	0.00001	0	1	0
$k=12$	0.00001	0	1	0
$k=13$	0.00001	0	1	0
ρ	0	0.00001	0	1

Source: Authors' own creation.

6. Conclusions

In view of the critical importance of bankruptcy prediction in the context of today's volatile economic environment, this paper introduces a novel algorithm that integrates DEA-DA techniques. The SBM model is particularly advantageous in that it accommodates negative values, thus enabling a more nuanced evaluation of each DMU. By employing this approach, we meticulously calculate the efficiency of each DMU, after which we determine the bankruptcy status of the entities in question using the DEA-DA model. The results demonstrate a high degree of precision in the prediction of bankruptcy through this model, thereby substantiating its efficacy and dependability. In light of the increasing necessity for reliable predictive instruments in financial administration, we strongly advocate the implementation of this algorithm in both academic and practical contexts. The potential of this approach to enhance decision-making processes in a range of industries is considerable, making it a valuable contribution to the field of financial analysis and risk management.

References

[1] Boudaghi, E., Saen, R.F. (2018), *Developing a novel model of data envelopment analysis–discriminant analysis for predicting group membership of suppliers in sustainable supply chain*. *Computers & Operations Research*, 89, 348-359.

[2] Charnes, A., Cooper, W.W., Rhodes, E. (1978), M*easuring the efficiency of decision making units*. *European journal of operational research*, 2(6), 429-444.

[3] Cheng, C.-B., Chen, C.-L., Fu, C.-J. (2006), *Financial distress prediction by a radial basis function network with logit analysis learning*. *Computers & Mathematics with Applications*, 51(3-4), 579-588.

[4] Cielen, A., Peeters, L., Vanhoof, K. (2004), *Bankruptcy prediction using a data envelopment analysis*. *European journal of Operational Research*, 154(2), 526-532.

[5] Farzipoor Saen, R. (2013), *Using cluster analysis and DEA-discriminant analysis to predict group membership of new customers*. *International Journal of Business Excellence*, 6(3), 348-360.

[6] Fisher, R.A. (1936), *The use of multiple measurements in taxonomic problems*. *Annals of eugenics*, 7(2), 179-188.

[7] Freed, N., Glover, F. (1981), *Applications and implementation: A linear programming approach to the discriminant problem*. *Decision Sciences*, 12(1), 68-74.

[8] Glover, F. (1990), *Improved linear programming models for discriminant analysis*. *Decision sciences*, *21*(4), 771-785.

[9] Horvathova, J., Mokrisova, M. (2023), *Overview of business bankruptcy models*. *Economic and Social Development: Book of Proceedings*, 257-273.

[10] Hussain, K.H., Uraibi, H.S. (2023), *Using the High Robustness Discriminant Analysis in Classification and Predictions (A Comparative Study)*. *Journal of Survey in Fisheries Sciences*, 10(3S), 4555-4569.

[11] Khalafi, F., Matin, R.K. (2021), *Potential Improvements in a Non-Radial Inefficiency Measurement*. *Economic Computation & Economic Cybernetics Studies & Research*, 55(3), 155-170.

[12] Kumar, P.R., Ravi, V. (2007), *Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review*. *European journal of operational research*, 180(1), 1-28.

[13] Mokrišová, M., Horváthová, J. (2023), *Domain Knowledge Features versus LASSO Features in Predicting Risk of Corporate Bankruptcy—DEA Approach*. *Risks*, 11(11), 199.

[14] Mousavi, M.M., Ouenniche, J., Tone, K. (2023), *A dynamic performance evaluation of distress prediction models*. *Journal of Forecasting*, 42(4), 756-784.

[15] Navidi, S., Rostamy-Malkhalifeh, M., Tohidi, G., Behzadi, M. (2023), *Using Data Envelopment Analysis-Discriminant Analysis for predicting the congestion*. *International Journal of Industrial Mathematics*, 15(4).

[16] Newton, G.W. (1998), *Bankruptcy Insolvency Accounting Practice and Procedure*. *1*, 4-21.

[17] Premachandra, I.M., Bhabra, G.S., Sueyoshi, T. (2009), *DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique*. *European Journal of Operational Research*, 193(2), 412-424.

[18] Shi, H., Liu, X., Chen, S. (2024), *Decision-Making Conflict Measurement of Old Neighborhoods Renovation Based on Mixed Integer Programming DEA-Discriminant Analysis (MIP DEA–DA) Models*. *Buildings*, 14(3), 785.

[19] Staňková, M., Hampel, D. (2023), *Optimal threshold of data envelopment analysis in bankruptcy prediction*. *SORT-Statistics and Operations Research Transactions*, 129-150.

[20] Sueyoshi, T. (1999), *DEA-discriminant analysis in the view of goal programming*. *European Journal of Operational Research*, 115(3), 564-582.

[21] Sueyoshi, T. (2001), *Extended DEA-discriminant analysis*. *European Journal of Operational Research*, 131(2), 324-351.

[22] Sueyoshi, T. (2004), *Mixed integer programming approach of extended DEA–discriminant analysis*. *European Journal of Operational Research*, 152(1), 45-55.

[23] Sueyoshi, T. (2006), *DEA-Discriminant Analysis: Methodological comparison among eight discriminant analysis approaches*. *European Journal of Operational Research*, 169(1), 247-272.

[24] Sueyoshi, T., Goto, M. (2009a), *Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA–discriminant analysis*. *European Journal of Operational Research*, 196(1), 289-311.

[25] Sueyoshi, T., Goto, M. (2009b), *Methodological comparison between DEA (data envelopment analysis) and DEA–DA (discriminant analysis) from the perspective of bankruptcy assessment*. *European Journal of Operational Research*, 199(2), 561-575.

[26] Sueyoshi, T., Kirihara, Y. (1998), *Efficiency measurement and strategic classification of Japanese banking institutions*. *International Journal of Systems Science*, 29(11), 1249-1263.

[27] Tavassoli, M., Faramarzi, G.R., Saen, R.F. (2015), *Ranking electricity distribution units using slacks-based measure, strong complementary slackness condition, and discriminant analysis*. *International Journal of Electrical Power & Energy Systems*, 64, 1214-1220.

[28] Tone, K. (2001), *A slacks-based measure of efficiency in data envelopment analysis*. *European journal of operational research*, 130(3), 498-509.

[29] Yadollahi, A.H., Kazemi Matin, R. (2022), *Centralized Resource Allocation in Two-Stage Production Systems: A Network DEA Approach.* *Economic Computation & Economic Cybernetics Studies & Research,* 56(3), 279-296.

[30] Yousefi, S., Saen, R.F., Shabanpour, H., Ghods, K. (2024), *An innovative patient clustering method using data envelopment Analysis–Discriminant analysis and artificial neural networks: A case study in healthcare systems*. *Socio-economic planning sciences*, 95, 102054.