Accounting and Auditing with Application

www.aaa.reapress.com

Acc. Aud. Appl. Vol. 2, No. 3 (2025) 216–227.

Paper Type: Original Article

# Forecasting the Financial Bankruptcy of Iranian Listed Companies Using a Hybrid DEA–PCA Approach

**Mohammad Reza Shahriari[1,*], Arash Zare-Talab[2], Mohammad Ali Mahmoudiar[3], Seyyed Abdullah Sajjadi Jagharq[4]**

[1] Department of Industrial Management, South Tehran Branch, Islamic Azad University, Tehran, Iran; Shahriari.mr@gmail.com.

[2] Department of Industries, Faculty of Industrial Engineering and Systems Management, Amirkabir University of Technology, Tehran, Iran; arash_zaretalab@yahoo.com.

[3] Department of Industries, Faculty of Industrial Engineering, Khajeh Nasiruddin Toosi University, Tehran, Iran; mmahmoodyar@mail.kntu.ac.ir.

[4] Department of Economics, Science and Research Branch, Faculty of Economics and Management, Islamic Azad University, Tehran, Iran; asadjady@yahoo.com.

**Citation:**

| | |
|---|---|
| *Received: 17 March 2025*<br>*Revised: 13 June 2025*<br>*Accepted: 16 July 2025* | Shahriari, M. R., Zare-Talab, A., Mahmoudiar, M. A., & Sajjadi Jagharq, S. A. (2025). Forecasting the financial bankruptcy of iranian listed companies using a hybrid DEA–PCA approach. *Accounting and auditing with application,* 2(3), 216-227. |

## Abstract

The topic of predicting company bankruptcy has attracted significant interest among financial researchers and experts. Due to the considerable impact of financial distress on companies' stakeholders, the development of accurate methods and models for forecasting bankruptcy and financial failure remains a key area of financial research. Investors consistently expect their capital to be secure and to receive returns that reflect the risks undertaken. Furthermore, the capacity to predict financial crises in companies in a timely manner in order to prevent capital loss is of critical importance. To address this need, researchers have conducted extensive studies employing various models and methods to evaluate corporate financial performance and forecast bankruptcy. However, it is essential to note that no single method is sufficient on its own; the best outcomes are achieved by combining multiple approaches with expert professional judgement. One technique that has gained increased attention in recent years for facilitating financial decision-making processes is Data Envelopment Analysis (DEA), which has produced acceptable predictive results. In this study, 52 manufacturing companies listed on the Tehran Stock Exchange were selected from three sectors: food and pharmaceuticals, metals, automotive and machinery, and chemicals and petrochemicals. Specifically, the first group included 21 companies (10 bankrupt and 11 healthy); the second group included 18 companies (10 bankrupt and eight healthy); and the third group included 13 companies (7 bankrupt and six healthy). The primary objective of this research is to evaluate the DEA model's ability to predict bankruptcy, i.e., to classify companies according to their financial distress status. To improve the performance of the DEA model, Principal Component Analysis (PCA) was used to reduce the dimensionality of its input variables.

**Keywords:** Data envelopment analysis, Principal component analysis, Stock market, Bankruptcy.

## 1 | Introduction

The topic of predicting company bankruptcy has attracted significant interest among financial researchers. Traditionally, bankruptcy prediction has relied on financial ratios and multiple discriminant analysis models.

Corresponding Author: Shahriari.mr@gmail.com

10.22105/aaa.vi.76

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

The most well-known of these is Altman's [1] Z-Score model. According to one study, Altman's [1] model achieved prediction accuracies of 95% and 83% one and two years prior to bankruptcy, respectively.

Past research has employed artificial neural networks. These models have generally produced more accurate forecasts than other methods [2]. *Table 1* provides a brief overview of the key aspects of previous research into predicting corporate bankruptcy. The following study has been conducted in the field of Data Envelopment Analysis (DEA) in recent years:

Badi Zadeh and Farzipour Saen [3] developed a network DEA model to analyse undesirable outputs and calculate production line efficiency. Hua and Bin [4] examined the relationships between DMUs and sub-DMUs, making the assumption that there are undesirable outputs. They also presented an NDEA model to calculate the total and partial efficiency of DMUs. Badi Zadeh et al. [5] developed a DEA network model to calculate efficiency in optimistic and pessimistic terms. Their proposed model could combine undesirable outputs and rank supply chains according to efficiency scores. They demonstrated the effectiveness of the model using a case study. Najafi et al. [6] presented an integrated framework combining balanced scorecard and DEA models to measure efficiency over time, based on time-lagged Key Performance Indicators (KPIs), for a proposed BSC model. The causal relationships between the BSC model perspectives over time were initially mapped as a dynamic BSC. After identifying the NDEA structure, they developed a new objective function to measure the efficiency of nine refineries affiliated with the National Iranian Petroleum Refining and Distribution Company over time and with strategies [7].

Rahmani et al. [8] used a novel technique based on a hybrid approach to evaluate the efficiency score of two-stage manufacturing processes where the required model data are imprecise. Their proposed method considers the expected value intervals as fuzzy and uses the convex combination of the two endpoints to measure the efficiency of each stage and the overall efficiency score for different values of $\alpha$.

Since 1994, the DEA model has been incorporated into bankruptcy prediction models. It has been found that this model is effective for bankruptcy prediction when used alongside other methods, such as regression [9]. In 2004, this model was used independently for prediction and demonstrated high accuracy. Subsequently, DEA models, which were previously only used to measure the efficiency of identical sets, received more attention due to their high accuracy in predicting bankruptcy in early research and their advantages. *Table 2* presents research conducted abroad in the field of bankruptcy prediction using DEA. Fewer studies have been conducted domestically in the field of bankruptcy prediction than those undertaken abroad. Only one study has been conducted on prediction using DEA. The following is a list of the most important domestic studies.

Other studies have generally used models designed abroad without changing the coefficients, which does not seem very accurate given the environmental differences (i.e., the differences in Iran's economic environment). Previous foreign and domestic studies have analysed all companies together. Given the nature of DEA, which requires all members of the set Decision Making Unit (DMUs) to be homogeneous, this is one of the weaknesses of previous studies. For this reason, this study selects manufacturing companies listed on the stock exchange as the statistical population and places them in three categories: Food and pharmaceutical industries; metal and automotive industries; and machinery, chemical, and petrochemical industries. Manufacturing companies on the stock exchange increase the accuracy and generalisability of the analysis, solving the main problem of previous research. Using the Principal Component Analysis (PCA) technique also helps to reduce the dimension of the inputs in the DEA model and prevent an artificial increase in the efficiency of DMUs.

The advantages of this method include the simultaneous calculation of the efficiency of companies in comparison with each other, their ranking, and the prediction of their bankruptcy. Compared to other methods based on research conducted in recent years, this method is more accurate. Further research is required to examine the power of this model in predicting bankruptcy, and this research has also been conducted in this direction.

## 2 | Statement of the Problem

Many studies have been conducted on predicting corporate bankruptcy, and various results have been obtained. These are discussed in more detail in the research background section. However, financial researchers have always been attracted to the use of new methods and, of course, old models in new formats. The simpler and more understandable these models are, the more they are favoured. One model that has recently received more attention for predicting the financial future of companies is the DEA model Staňková and Hampel [10].

This model provides a prediction of whether a company will be healthy or bankrupt in the future by measuring its efficiency. A key issue with previous studies is that all companies were placed in one group for analysis. Given the nature of DEA, which requires all members of the set DMUs to be homogeneous, this is one of the weaknesses of previous research. This study attempts to address this issue by categorising companies. Categorising allows the predictive ability of each category to be evaluated separately, providing a more comprehensive outcome regarding the effectiveness of this model in predicting the financial status of companies [11].

### 2.1 | Combining Data Envelopment Analysis and Principal Component Analysis

Since its introduction by Charnes et al. [12], DEA has been used as an effective tool for evaluation and modelling. In this method, the relative efficiency of each DMU is calculated by comparing it to similar units, using the weighted ratio of outputs divided by the weighted ratio of inputs [13]. However, the number of units that can be evaluated is limited by the number of input and output variables. Consequently, the greater the number of variables, the less powerful the analysis will be in distinguishing between efficient and inefficient units [14]. In such cases, it is therefore necessary to reduce the number of variables used in the DEA model. Obviously, such a reduction should have the least possible effect on distinguishing between efficient and inefficient units. Based on an empirical formula [15]:

$$(\text{output} \times \text{input}) \times 3 < \text{units under evaluation}. \tag{1}$$

If the DMUs are more than three times the sum of the inputs and outputs, the dimensions of the problem need to be reduced. The outputs need to be reduced on purpose (Otherwise, a large number of units will be on the efficiency frontier, their efficiency will be unity, and incorrect results will be produced, which will lower the quality of the model). To this end, Jenkins et al. [14] employed the partial covariance matrix to eliminate highly correlated variables. Alternatively, Alder et al. [16] used the PCA method to replace the main outputs or inputs entering the DEA model with input- and output-oriented principal components. This approach has also been used to evaluate privatised airline networks, measure airport quality [17], and select DEA variables and models [18]. Bruce et al. [19] used a similar approach to that of Cinca et al. [18] to evaluate performance in the internet banking industry using PCA and DEA methods.

These studies resulted in the development of a combined DEA-PCA model to reduce dimensionality in each category, rank DMUs, and ultimately predict bankruptcy. To achieve this, information relating to inputs and outputs is first extracted. Then, the ratio of single outputs to single inputs is calculated, and the PCA method is applied to this ratio. Selecting the first few principal components appropriately achieves the necessary reduction in the number of variables. The selected principal components are then used as inputs in the DEA model. This method works well in many applications where the number of units to be evaluated is small compared to the number of input and output variables, to the extent that other methods cannot provide a meaningful analysis [20].

### 2.2 | Proposed Methodology

As mentioned in the introduction, previous foreign and domestic research has shown that, in order to predict bankruptcy using the DEA model, all companies are placed in one group and analysed. However, given that DEA requires all members of the set DMUs to be homogeneous, this approach has been criticised as a weakness of previous research. For this reason, this research selects manufacturing companies listed on the stock exchange as the statistical population, placing them into three categories: Food and pharmaceutical companies; metal, automobile, and machinery companies; and chemical and petrochemical companies. The stock exchange increases the accuracy and generalisability of the analysis. This classification is based on production systems and similar risks for each category. For example, food and pharmaceutical companies in the same category have a continuous production system, similar raw material storage (Perishables), and similar product distribution and transportation. Consequently, they have similar financial and non-financial risks. The second category is also similar in terms of the main raw material (Metals) and production, storage, and distribution systems. The third category includes companies not in the first two groups, including chemical and petrochemical factories [21].

These factories are not dependent on metal industries, and most of them have a continuous production system. They have also been empirically proven to be dependent on economic conditions and oil prices in terms of financial risks. Of course, companies listed on the stock exchange are classified in various ways in books and on websites. This general classification of manufacturing companies is just one example. First, inputs and outputs are extracted from existing sources for each category, namely the companies' financial statements (Profit and loss statement and balance sheet). Inputs include operating expenses, current liabilities, and financial expenses, while outputs include net profit, operating profit, and current assets. These inputs and outputs have been selected based on past research and available sources. These inputs and outputs have been selected based on past research and available sources [22].

Then, the ratio of each output to input is obtained separately for each company, ultimately resulting in nine financial ratios. Given the incremental nature of these ratios, where the larger the fraction, the greater the utility, these ratios are considered a type of output. Therefore, the DEA model used in this study has no inputs. For this reason, we set the inputs of all DMU to the same value. In fact, setting any other constant value produces the same result. According to Mehregan's [15] empirical formula, we need to reduce the dimensionality of the problem and deliberately reduce the number of outputs.

Otherwise, a large number of units will be on the efficiency frontier, falsely suggesting that they are all equally efficient, which would lower the quality of the model. Therefore, we use a dimensionality reduction technique to reduce the number of outputs for each category. The best technique for this is PCA. We use the output of the PCA model for the DEA model. Due to the nature of the inputs and outputs, the DEA model used in this study should be output-oriented BCC. Changes in inputs do not cause changes in outputs in the same proportion, and returns to scale vary; therefore, the BCC model is more appropriate. Considering that DEA inputs are fixed in this study and the goal is to maximise output to increase efficiency, an output-oriented model should be used. The following are the multiplicative and enveloping BCC models, respectively.

$$\begin{aligned}
&\min \sum_{i} v_i x_{ip} + v'_p = \phi_{p,} \\
&\text{s.t.} \\
&\sum_{r} u_r y_{rp} = 1, \\
&\sum_{i} v_i x_{ij} - \sum_{r} u_r y_{rj} + v'_p \geq 0 \text{ for all j}, \\
&u_r, v_i \geq \varepsilon, v'_p \text{ free}.
\end{aligned} \tag{1}$$

$$\max \phi_p + \varepsilon \left[ \sum_{i} s_i^- + \sum_{r} s_r^+ \right], \tag{2}$$

$$\sum_{j} \lambda_j y_{rj} - s_r^{+} = \phi_p y_{rp} \text{ for all r,}$$

$$\sum_j \lambda_j = 1, \sum_j \lambda_j x_{ij} + s_i^{-} = x_{ip} \text{ for all i}, \lambda_j, s_i^{-}, s_r^{+} \geq 0, \phi_p \text{ free.}$$

The number obtained in the DEA model is an efficiency score between zero and one. According to the breakpoint obtained in the sample, a prediction is made regarding the company's bankruptcy or financial health. In order to use the calculated efficiency score for companies, a point between zero and one must be introduced that best separates healthy companies from bankrupt ones (The model breakpoint), in such a way that classification error is minimised. If a company's efficiency number is greater than the breakpoint, the company is considered healthy; if it is less, the company is considered bankrupt.

The general assumption is that healthy and efficient companies have a higher efficiency score, close to one. In fact, their score should be higher than that of inefficient companies, which should be close to zero. The goal is to measure the model's ability to predict bankruptcy one and two years before the event occurs. To this end, the "year of occurrence" for each distressed company is first extracted based on law 141 of the Trade. To give healthy companies a specific year of occurrence, too, the year of occurrence of the distressed company is determined, and then, among the healthy companies in the same group, companies of a similar size will also have the same year of occurrence.

Then, the required financial information for one and two years before the year of occurrence is prepared and entered into the model. To obtain the cut-off point, the financial information for the year of occurrence for each of the three groups is entered into the model to obtain the efficiency score for each group separately. A graph showing the distribution of efficiency scores for both the healthy and bankrupt categories is then drawn. The intersection point of these two graphs (Where the least classification error occurs) will be the cut-off point and the basis for prediction. As previously mentioned, if a company's efficiency number is greater than its category's cut-off point, the company is considered healthy; if it is less, the company is considered bankrupt.

In one of the previous studies, the number 0.5 was chosen as the cut-off point without any specific calculation or basis. That is, if the efficiency score for a company was greater than 0.5, the company was considered healthy; if it was less than 0.5, the company was predicted to be bankrupt. However, this study uses the same method (For higher accuracy).

This comparison is made for one and two years before the year of occurrence, in order to test the model's ability to predict past events. Finally, the accuracy of the model is determined by examining the percentage of bankrupt and healthy companies that the model correctly predicted one and two years before the year of occurrence. Then, a proportion test at the 95% level is used to test the research hypotheses and see whether the model correctly classifies the companies according to the expected percentage.

$$\frac{(p - p_0)}{\left(\frac{p_0 q_0}{n}\right)^{0.5}}. \quad \textbf{(3)}$$

In this formula, p is the percentage of companies that are classified correctly by the model. P0 is the accuracy percentage that is expected, which is set at 0.5 based on previous research. Q0 is defined as 1-P0. At a confidence level of 95%, we examine whether there is a significant difference between the predicted ratios of the three groups. If there is a difference, we can conclude that this model can be helpful for predicting bankruptcy in some groups, but not necessarily in all.

$$\frac{p_1 - p_2}{\left(\left(\frac{p_1 q_1}{n_1}\right) + \left(\frac{p_2 q_2}{n_2}\right)\right)^{0.5}}. \tag{4}$$

In this formula, P1 and P2 correspond to two of the three groups, making a total of three pairwise comparisons for each year. *Fig. 1* shows the general research procedure based on the above explanation.

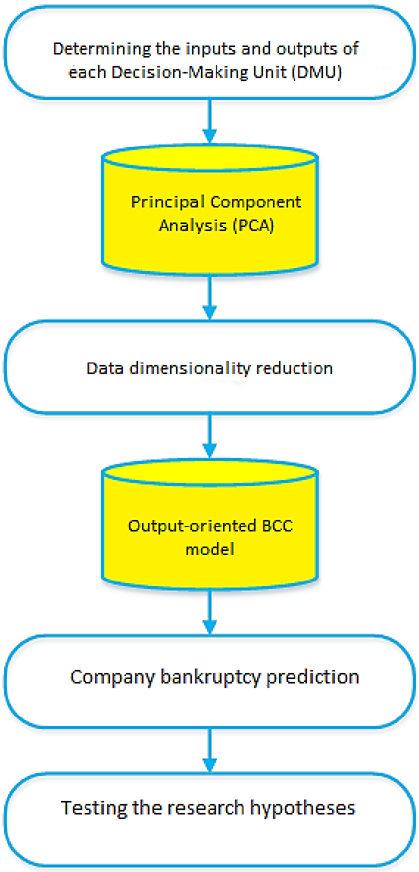


**Fig. 1. Conceptual research model.**

## 2.3 | Research Variables

The independent variables in this study are financial ratios used to classify companies as either bankrupt or healthy. The efficiency score obtained from the DEA model can also be considered an independent variable. Finally, a company's classification as healthy or bankrupt, which is a function of the efficiency score, is considered a dependent variable.

# 3 | Implementation of the Proposed Methodology and Analysis of the Findings

According to the available information, 52 manufacturing companies listed on the Tehran Stock Exchange were selected in three groups: Food and pharmaceutical companies; metal, automobile, and machinery companies; and chemical and petrochemical companies. Twenty-one companies (Ten bankrupt and eleven healthy) were in the first group, eighteen companies (Ten bankrupt and eight healthy) were in the second group, and thirteen companies (Seven bankrupt and six healthy) were in the third group. The names and information of these companies were extracted from the Kodal website.

First, to calculate the cut-off point for the companies in the first group, we entered the financial information for the year in which the companies in this group became bankrupt or healthy into the model. *Fig. 2* shows the distribution of efficiency scores for Financial health and bankrupt companies in the first group.

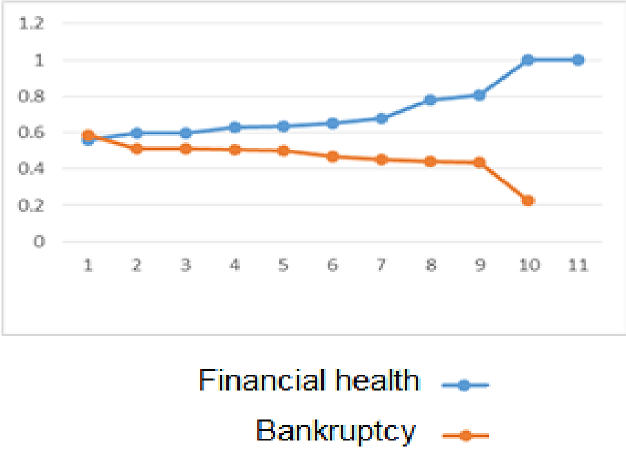


**Fig. 2. Efficiency score distribution chart of the first group of companies.**

According to *Fig. 1*, the cut-off point for this group is 0.59. As can be seen from the chart, as expected, the efficiency score of healthy companies increases and approaches one from the cutoff point onwards, while the efficiency score of bankrupt companies decreases and approaches zero (In fact, the lowest classification error occurs). As stated in Section 2, the process begins with the extraction of inputs and outputs from existing sources, namely the companies' financial statements (Profit and loss statement and balance sheet). Inputs include operating expenses, current liabilities, and economic expenses; outputs include net profit, operating profit, and current assets. The ratio of each output to input is then obtained for each company separately, of which there are nine in total. This number is very small due to the small number of DMUs (Companies), which causes errors in the DEA model. Therefore, they are reduced by PCA and then entered into the output-oriented BCC DEA model. Finally, an efficiency score is obtained, which is a number between 0 and 1.

Based on the calculations, one year before the event occurred, the model correctly predicted 73% of healthy companies and 90% of bankrupt companies. In two years prior to the event, the model correctly predicted 73% of Financial health companies and 80% of bankrupt companies. *Fig. 3* shows the calculation of the breakpoint for the metal, automobile, and machinery companies group.

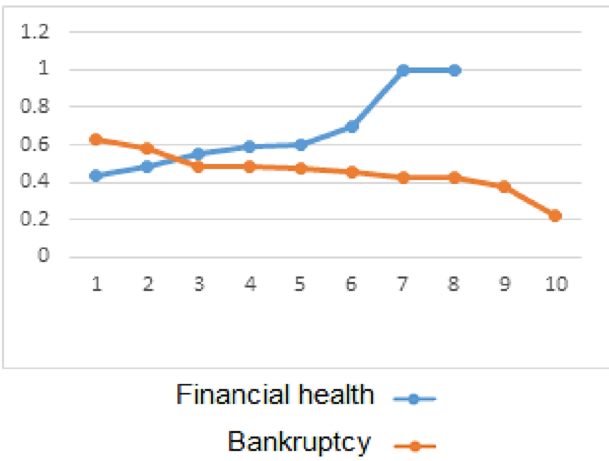


**Fig. 3 Efficiency score distribution chart of companies in the second group.**

As can be seen, the breakpoint value for this group is 0.52. One year before the event, the model correctly predicted 100% of healthy companies and 10% of bankrupt companies. Two years before the event, it correctly predicted 75% of Financial health companies and 80% of bankrupt companies.

*Fig. 4* shows the calculation of the breakpoint for the group of chemical and petrochemical companies.

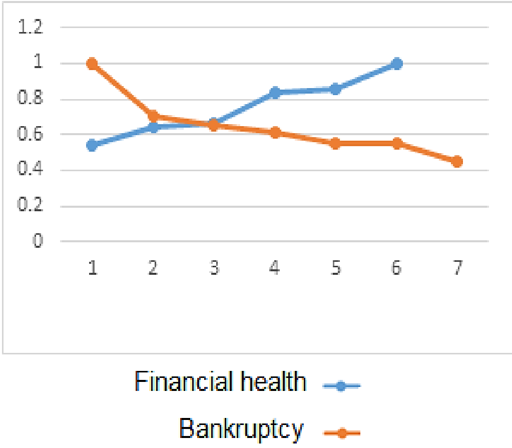


**Fig. 4. Efficiency score distribution chart of third group companies.**

As can be seen, the cut-off value for this group is 0.65. One year before the event, the model correctly predicted 83% of healthy companies and 85% of bankrupt companies. Two years before the event, it correctly predicted 34% of financial health companies and 71% of bankrupt companies.

At this stage, we will test the following hypotheses:

**Hypothesis 1.**

H0: It is not possible to predict bankrupt companies in the food and pharmaceutical industries on the stock exchange using a combination of DEA and PCA models.

H1: It is possible to predict bankrupt companies in the food and pharmaceutical industries on the stock exchange using the DEA-PCA model combination.

This hypothesis was tested using a 95% confidence ratio test, and the results are shown in *Table 4*.

**Table 4. Results obtained for Hypothesis 1.**

| Year | Percentage of Correct Predictions by the Model (Percentage) | Expected Percentage | Z Value | Result |
|---|---|---|---|---|
| One year before the event for healthy companies | 73 | 50 | 1.526 | H0 confirmed |
| One year before the event for bankrupt companies | 90 | 50 | 2.530 | H1 confirmed |
| Two years before the event for healthy companies | 73 | 50 | 1.526 | H0 confirmed |
| Two years before the event for bankrupt companies | 80 | 50 | 1.897 | H1 confirmed |

As can be seen, the model can predict which companies in the first group (On the verge of bankruptcy) will go bankrupt one or two years before it happens. However, it cannot predict which Financial health companies will go bankrupt.

**Hypothesis 2.**

H0: It is not possible to predict bankrupt companies in the metal industry, or in the automobile and machinery sectors of the stock exchange, using the combination of DEA and PCA models.

H1: It is possible to predict bankrupt companies in the metal, automotive, and machinery industries on the stock exchange using the combination of DEA and PCA models.

**Table 5. Results obtained for Hypothesis 2.**

| Year | Percentage of Correct Predictions by the Model (Percentage) | Expected Percentage | Z Value | Result |
|---|---|---|---|---|
| One year before the event for healthy companies | 100 | 50 | 2.828 | Confirm H1 |
| One year before the event for bankrupt companies | 10 | 50 | -2.530 | Confirm H0 |
| Two years before the event for healthy companies | 75 | 50 | 1.414 | Confirm H0 |
| Two years before the event for bankrupt companies | 80 | 50 | 1.897 | Confirm H1 |

According to the second hypothesis, the model can predict critical situations for the second group of companies over the previous two years, but not over the previous year. Regarding the financial health of the companies, the model predicted 100% correctly in the previous year, but could not produce reliable results in the two years prior to the occurrence.

**Hypothesis 3.**

H0: It is not possible to predict which chemical and petrochemical companies will go bankrupt on the stock exchange using a combination of DEA and PCA models.

H1: It is possible to predict bankrupt chemical and petrochemical companies on the stock exchange using the DEA-PCA model combination.

**Table 6. Results obtained for Hypothesis 3.**

| Year | Percentage of Correct Predictions by the Model (Percentage) | Expected Percentage | Z-Value | Result |
|---|---|---|---|---|
| One year before the event for healthy companies | 83 | 50 | 1.746 | Confirm H1 |
| One year before the event for bankrupt companies | 85 | 50 | 1.852 | Confirm H1 |
| Two years before the event for healthy companies | 34 | 50 | -0.847 | Confirm H0 |
| Two years before the event for bankrupt companies | 71 | 50 | 1.111 | Confirm H0 |

For the third group of companies, the model predicted healthy and bankrupt companies accurately in the year before the event, but was unable to distinguish between them in the two years prior to that.

**Hypothesis 4.**

H0: There is no significant difference in the ability to distinguish between healthy and bankrupt companies across the three groups.

H1: There is a significant difference between the results obtained from the three groups when distinguishing between healthy and bankrupt companies.

**Table 7. Results obtained for Hypothesis 4.**

| Year | Comparison | Z-Value | Result |
| --- | --- | --- | --- |
| One year before the event | 1 with 2 | 5.96 | Confirm H1 |
| Year | 1 with 3 | 0.303 | Confirm H0 |
| | 2 with 3 | -4.546 | Confirm H1 |
| One year before the event | 1 with 2 | 0 | Confirm H0 |
| | 1 with 3 | 0.422 | Confirm H0 |
| | 2 with 3 | 0.422 | Confirm H0 |

As can be seen in the table, there is a significant difference in the results obtained from the three groups one year before the event, whereas there is no significant difference in the results obtained two years before the event. The following analyses can be presented considering the results obtained in the three groups:

Due to their structure, food and pharmaceutical companies lose customers continuously over time if they perform poorly and are unable to meet their needs. Losing customers is directly reflected in their financial indicators, so their financial performance can be used to predict their future situation.

It should be noted that, due to fluctuations in basic metal prices and exchange rates, financial crises are somewhat more difficult to predict in this group of companies. Also, the low cut-off point of this group is due to the low efficiency score of the companies in the sample in the year of occurrence. Low efficiency score in the sample means that the current results may differ slightly in other samples. Fluctuations in oil prices and exchange rates strongly impact companies whose raw material is oil or whose competitors are imported goods. The severe fluctuations in oil prices over the ten-year period covered by this study have caused instability among companies in groups two and three, which is perhaps why the model used in the study, which relies mainly on financial indicators, has not been very effective in predicting the financial situation and crises of these two groups.

# 4 | Conclusion

Various methods and models, such as Altman's [1], the multiple diagnostic method, and artificial neural networks, have been used to predict bankruptcy, yielding different results. Some methods have produced successful predictions, while others have failed. Using the DEA model with an efficiency score is a new method of predicting bankruptcy, and there have been mixed results in this area so far. Therefore, more research is needed to investigate the effectiveness of this model.

Although the DEA model is usually used to calculate the efficiency of homogeneous and similar groups, in previous studies, all companies were placed in one group for evaluation. In this study, it was decided to create smaller groups and evaluate manufacturing companies operating in similar fields separately. More financial indicators were also used than in previous studies. However, if the number of DMUs (Here, manufacturing companies in each group) in DEA models is close to the total number of inputs and outputs in the model, this causes problems with the model's results. Therefore, a method was needed to reduce the dimension; this is why the PCA statistical method was employed. The PCA statistical method increased the model's generalisability, as it allows more groups to use it.

Similar previous work, in which all companies were evaluated together using a DEA model, showed that bankruptcy could be predicted up to two years in advance. However, it was unclear which group of companies the model was most effective in predicting, and the results were obtained in a general sense. In the current study, however, we see that the results for different groups vary, and that the BCC DEA model, using the PCA statistical technique, performs better at predicting the bankruptcy of food and pharmaceutical companies, with an accuracy of up to two years (Based on law 141 of the trade). This model can predict the bankruptcy of chemical and petrochemical companies one year before it occurs. However, its performance in predicting the bankruptcy of companies in the metal, automobile, and machinery industries is lower. Instead, it accurately predicts the health of companies in this group. The results obtained in this study show that classifying companies based on their functions and the risks they are involved in, and evaluating them separately using the model, can be beneficial in obtaining acceptable results and providing researchers with more complete information.

It is important to note that one should never rely solely on any bankruptcy prediction method, including the one proposed in this study, even if it has previously produced successful results, because predictions are never guaranteed. To increase the likelihood of a successful prediction, several mathematical and statistical methods and models should always be used alongside fundamental analysis (Interpretation and prediction of the future based on political, economic, military, and social issues).

# Conflict of Interest

The authors declare no conflict of interest.

# Data Availability

All data are included in the text.

# Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# References

[1] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589–609. https://doi.org/10.2307/2978933

[2] Mehrani, S., Mehrani, K., Monsefi, Y., & Karami, G. (2005). An applied study of Zimsky and Shirata bankruptcy prediction models in companies listed on the Tehran Stock Exchange. *Accounting and auditing reviews*, *12*(41), 105-131. (**In Persian**). https://dor.isc.ac/dor/20.1001.1.26458020.1384.12.3.4.5

[3] Badiezadeh, T., & Farzipoor Saen, R. (2014). Efficiency evaluation of production lines using maximal balance index. *International journal of management and decision making*, *13*(3), 302–317. https://doi.org/10.1504/IJMDM.2014.063573

[4] Hua, Z., & Bian, Y. (2008). Performance measurement for network DEA with undesirable factors. *International journal of management and decision making*, *9*(2), 141–153. https://doi.org/10.1504/IJMDM.2008.017196

[5] Badiezadeh, T., Saen, R. F., & Samavati, T. (2018). Assessing sustainability of supply chains by double frontier network DEA: A big data approach. *Computers & operations research*, *98*, 284–290. https://doi.org/10.1016/j.cor.2017.06.003

[6] Najafi, S. E., Aryanezhad, M. B., Hosseinzadeh Lotfi, F., & Ebnerasoul, S. A. (2023). Performance evaluation accounting with inputs non-discretionary factors in an integrated BSC-DEA methodology. *Big data and computing visions*, *3*(3), 111–124. (**In Persian**). https://doi.org/10.22105/bdcv.2023.190170

[7] Akbarian, M., Najafi, E., Tavakkoli-Moghaddam, R., & Hosseinzadeh Lotfi, F. (2015). A network-based data envelope analysis model in a dynamic balanced score card. *Mathematical problems in engineering*, *2015*(1), 914108. https://doi.org/10.1155/2015/914108

[8] Rahmani, A., Rostamy-malkhalifeh, M., & Hosseinzadeh Lotfi, F. (2020). Evaluating performance of two-step networks using fuzzy data envelopment analysis. *Revista gestão & tecnologia*, *20*, 96–105. https://doi.org/10.20397/2177-6652/2020.v20i0.1748

[9] Fernandez-Castro, A., & Smith, P. (1994). Towards a general non-parametric model of corporate performance. *Omega*, *22*(3), 237–249. https://doi.org/10.1016/0305-0483(94)90037-X

[10] Staňková, M., & Hampel, D. (2023). Optimal threshold of data envelopment analysis in bankruptcy prediction. *SORT-statistics and operations research transactions*, *47*(1), 129–150. https://doi.org/10.57645/20.8080.02.3

[11] Gholaminezhad, A., Rostamy-malkhalifeh, M., Hosseinzadeh Lotfi, F., & Allahviranloo, T. (2025). Integrating of data envelopment analysis and discriminant analysis to predict the bankruptcy of firms: Application in electricity industry. *Economic computation & economic cybernetics studies & research*, *59*(1), 22–37. 10.24818/18423264/59.1.25.02

[12] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

[13] Cooper, W. W. (2001). Data envelopment analysis. In *Encyclopedia of operations research and management science* (pp. 183–191). Springer. https://doi.org/10.1007/1-4020-0611-X_212

[14] Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European journal of operational research*, *147*(1), 51–61. https://doi.org/10.1016/S0377-2217(02)00243-6

[15] Mehregan, M. R. (2008). *Quantitative models in organizational performance evaluation (DEA) data envelopment analysis" publications*. Faculty of Management, University of Tehran. (**In Persian**). https://b2n.ir/rm3345

[16] Adler, N., & Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the operational research society*, *53*(9), 985–991. https://doi.org/10.1057/palgrave.jors.2601400

[17] Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. *European journal of operational research*, *132*(2), 260–273. https://doi.org/10.1016/S0377-2217(00)00150-8

[18] Cinca, C. S., & Molinero, C. M. (2004). Selecting DEA specifications and ranking units via PCA. *Journal of the operational research society*, *55*(5), 521–528. https://doi.org/10.1057/palgrave.jors.2601705

[19] Ho, C. T. B., & Wu, D. D. (2009). Online banking performance evaluation using data envelopment analysis and principal component analysis. *Computers & operations research*, *36*(6), 1835–1842. https://doi.org/10.1016/j.cor.2008.05.008

[20] Yazdi, H. D., Movahedi Sobhani, F., Lotfi, F. H., & Kazemipoor, H. (2023). A novel algorithm for complete ranking of DMUs dealing with negative data using data envelopment analysis and principal component analysis: Pharmaceutical companies and another practical example. *Plos one*, *18*(9), e0290610. https://doi.org/10.1371/journal.pone.0290610

[21] Sidhu, D. A., & Katoch, R. (2019). Bankruptcy prediction using Altman z-score model and data envelopment analysis model: A case of public listed realty sector companies in India. *International journal of advanced science and technology*, *28*(13), 399–411. https://ssrn.com/abstract=3904529

[22] Štefko, R., Horváthová, J., & Mokrišová, M. (2020). Bankruptcy prediction with the use of data envelopment analysis: An empirical study of Slovak businesses. *Journal of risk and financial management*, *13*(9), 212. https://doi.org/10.3390/jrfm13090212