DOI: 10.17323/2587-814X.2025.3.7.21

Improving bankruptcy prediction using Data Envelopment Analysis scores

Yuri A. Zelenkov
E-mail: yzelenkov@hse.ru

Graduate School of Business, HSE University, Moscow, Russia

Abstract

Most current bankruptcy prediction models are based on financial ratios, although their usage is not supported by formal theory and their interpretation is problematic. One of the prospects for improving the predictive models is the study of other firm performance measures, such as the data envelopment analysis (DEA) scores. However, this raises the problem of choosing the optimal DEA specification, since it determines the shape of the efficiency frontier and the predictive properties of the model. This paper presents a method for automatically designing DEA models whose scores are then used as features to improve the quality of bankruptcy predictors. The method has two goals. The first is to improve accuracy. The second objective assumes that if DEA scores improve the prediction, then the specification of this model can provide information about failures. At the first step, accounting measures that are potentially suitable for the DEA are selected using hierarchical clustering. The second step explores the causal relationships between the selected measures. The third step calculates pure technical efficiency, scale efficiency and mix efficiency. Experiments with two datasets show that the inclusion of these scores in the list of features improves the AUC-ROC by more than 20%, which is superior to previous works. The analysis of the DEA models provides insight into the reasons for a firm's failure in both stable and crisis periods.

Keywords: bankruptcy prediction, bankruptcy factors, DEA specification, causal modeling

Citation: Zelenkov Yu.A. (2025) Improving bankruptcy prediction using Data Envelopment Analysis scores. *Business Informatics*, vol. 19, no. 3, pp. 7–21. DOI: 10.17323/2587-814X.2025.3.7.21

Introduction

Current bankruptcy prediction models use financial ratios (FR) because bankruptcy is a consequence of the inability of a firm to meet its obligations expressed in monetary form [1]. However, many researchers in the field of financial analysis note that the comparison of firms based solely on a cross-sectional FRs analysis has certain limitations. Use of FRs is based on the following assumptions: constant return to scale, linear relationships between variables, multivariate normal distributions, and so on [2]. Therefore, while FRs are easy to compute, their interpretation is problematic, especially when two or more ratios provide conflicting signals [3].

The authors of machine learning models try to overcome these limitations by simultaneously using many different FRs – typically, several dozen indicators divided into groups of activity, solvency, profitability, liquidity, leverage, etc. [1]. It is assumed that this approach allows training a predictor that combines views of the firm from different points. However, this generalization of different types of ratios is just a convenient means and is not supported by any formal theory [4]. In addition, although complex models perform well on demonstration data [5], practitioners prefer linear methods [6, 7] as they are better interpretable.

One of the prospects for improving the performance of predictive models is the study of other performance scores of a firm, namely, measurements generated using Data Envelopment Analysis (DEA). DEA allows us to assess the performance of decision-making units (DMUs), which use the same types of inputs and produce the same type of outputs against each other. However, DEA model specification, in the form of feature (where the term feature is used to refer to either outputs, inputs or environmental variables) selection, has a significant impact on the shape of the efficient frontier as well as the discriminatory power of DEA models [8].

Another important advantage of DEA is the interpretability of the model. The interpretation should be easy to understand, and the properties of the model should be natural, which facilitates acceptance of the model by stakeholders [8].

Thus, the question arises as to which firm performance indicators should be selected for the DEA model to ensure both informativeness and interpretability. Many authors using DEA scores as predictive features choose to specify the model according to the rule that inputs are the resources to be minimized and outputs are the outcomes to be maximized. However, the use of scores derived from these models yields predictor accuracy gains of between 1–2% [9, 10]. Thus, two related research questions can be formulated:

1. Is it possible to generate a DEA specification using data science methods whose scores would significantly (by 5% or more) improve the accuracy of bankruptcy prediction?

2. If DEA scores would significantly improve prediction accuracy, could the specification of this model provide new information about the causes of bankruptcies?

If the answers to both questions are positive, it means that by using the data science approach to generate the DEA specification we gain new knowledge

about the factors leading to bankruptcy. This knowledge can be used to support decisions to ensure the financial sustainability of the firm.

This paper[1] proposes a formal approach for designing a DEA specification which includes three steps. At the first step, accounting measures that are potentially suitable for the DEA are selected using hierarchical clustering. The second step explores the causal relationships between the selected measures and determines the inputs and outputs of the DEA specification. The third step calculates pure technical efficiency (PTE), scale efficiency (SE) and mix efficiency (ME), which are then used as features to train a classifier that separates bankrupt and successful firms. Experiments with two datasets have shown that the inclusion of PTE, SE, and ME in the list of features improves the AUC-ROC by more than 20%, which is significantly superior to the results of previous works. Finally, we analyze the DEA models developed in this way, which allows us to draw nontrivial conclusions about the reasons why firms fail in both stable and crisis periods.

1. Literature review

Suppose we have a set of N DMUs denoted as DMU_i ($i = 1, ..., N$), which produce outputs y_l ($l = 1, ..., L$) by using inputs x_k ($k = 1, ..., K$). The goal is to find for any DMU denoted as DMU_0 the weights of inputs and outputs that maximizes efficiency. Efficiency is defined as the ratio of weighted outputs to weighted inputs, subject to this ratio being not greater than 1 for any DMU:

$$\max \theta_0 = \mathbf{u}^T\mathbf{y}_0 / \mathbf{v}^T\mathbf{x}_0 \quad (1)$$
$$\text{s.t. } -\mathbf{v}^T\mathbf{X} + \mathbf{u}^T\mathbf{Y} \leqslant 0, \mathbf{v} \geqslant 0, \mathbf{u} \geqslant 0,$$

where $\mathbf{u}$ and $\mathbf{v}$ are vectors of weights to be estimated; θ_0 is efficiency of DMU_0.

There are several types of DEA models, which can be classified firstly in terms of returns to scale and secondly in terms of orientation. First, the Charnes, Cooper and Rhodes (CCR) model [11] was introduced, which assumed a constant returns to scale (CRS). The Banker, Charnes and Cooper (BCC) model [12] estimates the efficiency of DMUs when returns to scale is variable (VRS), thus, it takes account of scale effect and postulates that convex combinations of the observed DMUs form the production possibility set. The CCR and BCC scores, which we denote as θ_{CCR} and θ_{BCC}, are called the Technical Efficiency (TE) and the Pure Technical Efficiency (PTE), respectively. If a DMU has a perfect θ_{BCC} but a low θ_{CCR}, then its operations are locally efficient but not globally efficient, because of the scale size of the DMU. Considering these concepts, the Scale Efficiency (SE) is defined as

$$SE = \frac{TE}{PTE} = \frac{\theta_{CCR}}{\theta_{BCC}}. \quad (2)$$

This decomposition determines the source of global inefficiency as measured by TE, whether due to inefficient operation (PTE) or due to the scale (SE) or both [13].

In terms of orientation, input-oriented model minimizes the inputs $\mathbf{v}^T\mathbf{x}_0$, and the output-oriented model maximizes the outputs $\mathbf{u}^T\mathbf{y}_0$.

There are several problems associated with the use of DEA [2]. First, DMUs being considered should be strictly comparable, i.e., operating in the same environment (industry, market and being exposed to the same risks). Next, inputs and outputs should match; at least, they should be recorded in the same accounting periods. Most fundamentally, there is a problem determining which inputs and outputs should be selected for inclusion in the model. There is no selection procedure like, for example, stepwise regression to guide the analysis. Finally, the DEA relies on a search for outliers to identify the efficient frontier, and analysts should decide whether the specific outlier offers an appropri-

[1] Preliminary results of the study were presented at the XXIV Yasin (April) International Academic Conference on Economic and Social Development (Moscow, 4–14 April 2023), https://conf.hse.ru/en/2023/

ate representation of feasible efficiency, or is it noise. Again, there is no standard procedure which allows us to identify noise data.

In bankruptcy prediction studies, DEA is applied in two different ways. First, the DEA model is used as a classifier to separate bankrupt and successful firms. Second, DEA scores are used as additional attributes when training classifiers based on machine learning algorithms.

Some economic theories assume that bankruptcy is the result of a selection process that eliminates inefficient firms. From this point of view, DEA is an appropriate tool because it compares firms operating under equal conditions and identifies potentially inefficient ones. However, this raises the problem of converting the DEA score (i.e., distance from efficiency frontier) into the probability of bankruptcy. Therefore, many authors introduce additionally an 'inefficiency frontier' [14, 15], and the probability of bankruptcy is estimated in the direction of the 'anti-ideal' DMU, which requires maximum cost to produce minimum results. A separate problem when using DEA as a classifier is the validation of the model on observations outside the training sample [16].

Another line of research is using the DEA score as a predictive variable. In the first stage, DEA evaluates the efficiency; then the score obtained is stored as a feature to train the classifier in the second stage. Researchers use different models to compute such features: CCR [17, 18], BCC [19] and their decomposition [9].

The decomposition of DEA scores in the context of the bankruptcy prediction problem is discussed in [10]. The authors use Slacks-Based Measurements (SBM) presented in [20]. Radial DEA models (CCR and BCC) do not account for possible input and output slack, and therefore possibly overestimate efficiency. The SBM model is a non-radial model that considers slacks in inputs and outputs. For the input-oriented model, the SBM model equals the CCR model if the latter has zero input slacks for every optimal solution, i.e., $\rho_{in} = \theta_{CCR[in]}$, where ρ_{in} is a SBM input-oriented score. So, input Mix Efficiency (ME) can be defined as [13]:

$$ME_{in} = \frac{\rho_{in}}{\theta_{CCR[in]}}.$$

The same reasoning holds true for the output-oriented models. Thus, SBM score can be decomposed as

$$SBM = PTE \cdot SE \cdot ME. \quad (3)$$

Using PTE, SE and ME as features of the second stage classifier improves the performance of default prediction by providing more information about the causes of financial problems [10].

Table 1 summarizes all the works discussed above. The improvement in classification accuracy achieved by incorporating DEA scores into the model is in most cases quite low, less than 2%. The only work showing better results [18] is based on a small dataset.

Perhaps since the inclusion of DEA scores in predictive models does not significantly improve accuracy, researchers' interest in this topic has decreased, with no significant publications after 2019. Thus, we can conclude that the potential of DEA is not fully exploited in bankruptcy forecasting models.

According to [10], there are several issues that need to be addressed: (1) selection of a proper DEA model, (2) type of return to scale, (3) orientation of DEA analysis and (4) suitable DEA inputs and outputs. The works [9,10] provide some guidance on the first three questions. This study focuses on the last issue: we present a method for selecting inputs and outputs that significantly improves model performance.

2. Data

To demonstrate our method, we will use two sets of data on Russian companies. The first contains data from the 2018 financial statements of 5736 companies, of which 525 were declared bankrupt during the following year. The second dataset includes financial statements for 2019 of 6757 companies, of which 1124 went bankrupt in 2020. Both samples contain data for the 14 financial reporting indicators presented in *Table 2* and a class label (1 – bankrupt and 0 – health firm).

Table 1.

Papers that use DEA score as a predictive feature

Paper	Dataset (good/failure)	Inputs	Outputs	DEA score	Prediction model	Prediction improvement
[17]	60/60	TA, TL, Cost of sales	Income from sales	CCR[IO]	DA, LR, SVM	Accuracy: 1.8%
[18]	76/38	R&D expenses, R&D designers, Number of patent and trademarks	Gross profit, Market share	CCR[OO]	SVM, MLP	AUC-ROC: 5.5%
[19]	1142/377 1241/354 2293/344	Capital shares, Number of fulltime employees	Value added	BCC[OO]	LR	N/A
[9]	420/50	Number of employees, Share capital, TC, TA, TL	TS, Total profit, Cash accrued	PTE, SE[IO]	LR	AUC-ROC: 0.3%
[10]	Train 7749/440 Test 3518/236	TL, SF, Number of employees	TS	PTE, SE, ME[IO]	LR	AUC-ROC: 1.4%

DA – discriminant analysis, LR – logistic regression, SVM – support vector machine, MLP – multilayer perceptron.

The data comes from the "Ruslana" database[2], maintained by Bureau van Dijk. The selected firms are manufacturing companies (NACE 2 code C) registered in the Central Federal District of Russia and belonging to small and medium-sized businesses (number of employees not exceeding 50). Industry, region and size selection rules were introduced to meet the requirement of the DEA that the analyzed objects should be comparable.

In addition, the first dataset (2018) refers to a relatively stable period, with GDP growth of 2.2% for 2019. In 2020, GDP fell by 2.7%, largely because of the COVID-19 pandemic. This has been reflected in a sharp rise in bankruptcies. Thus, by comparing the DEA models for these two periods, we can get insights into the factors leading to bankruptcy in both stable and crisis years.

Since DEA requires only positive normally distributed values, all data were transformed in the following way. Firstly, we scaled all values into the interval [1, 10], and secondly, we applied the Box–Cox transformation. To train the classifier, we also computed 32 different FRs characterising the activity, profitability, liquidity and solvency of the firm as well as its financial structure.

3. Proposed approach to the DEA model design

Following [10], we use input-oriented models of CCR, BCC and SBM and decompose the results based on Eq. (3). We will also use raw accounting data (*Table 2*) in the DEA models, as FRs will be used as features for training classifiers in the second step [9].

[2] https://ruslana.bvdep.com

Table 2.

Accounting data

AR	Account Receivables	*PC*	Production Costs
CA	Current Assets	*SF*	Shareholders' Funds
CL	Current Liabilities	*TA*	Total Assets
EBIT	Earnings Before Interests and Taxes	*TL*	Total Liabilities
FA	Fixed Assets	*TS*	Total Sales
LTD	Long Term Debt	*VA*	Value Added
MVE	Market Value of Equity	*WC*	Working Capital

The key issue is the specification of the DEA model. Firstly, a good model should make conceptual sense not only from a theoretical point of view but also from a practical one. In addition, there are technical constraints, in particular the variables (inputs and outputs) must not be highly correlated.

Thus, the objective of our study can be formulated as follows: to find a set of DEA inputs and outputs that: (1) produces scores that significantly increase the predictive properties of the classifier and (2) are well interpreted and allow us to draw conclusions about the causes of inefficiency leading to bankruptcy. Note that while the first requirement is measurable (e.g., through quality metrics), the second is entirely subjective and requires the involvement of experts.

Since we use DEA as a source of additional features for the second-stage classifier, this means that we need to obtain DEA scores that are highly correlated with the target variable – the bankrupt/healthy firm class label. We therefore propose a three-stage process. In the first stage, candidate features with low correlation between them and high correlation with the class label are selected. The second stage involves formal causal analysis and identification of inputs and outputs, which are theoretically validated. In the third stage, scores are computed for CCR, BCC and SBM models and transformed according to Eq. (3). *Figure 1* presents this process as a part of training and validation pipeline for predictive model used in this paper. *Figure 1* also shows software tools (library and programming language) used in each step. The following sub-sections describe this three-stage process in more detail, using the 2019 dataset as an example.

Step one: features selection. The selection process begins with clustering the features according to their correlation. This is done by using an agglomerative clustering algorithm.

Let $d(s, t)$ be the distance between features s and t, and $d(s, t) = 1 - r(s, t)$, where $r(s, t)$ is a correlation between s and t. The agglomerative clustering algorithm begins with a forest of features that will be used to form the hierarchy. At each iteration, it combines nearest features s and t into cluster u, and u becomes the root of s and t. The algorithm stops when only one root remains in the forest. To calculate the distance between the newly formed cluster u and v the Farthest Point Algorithm is used: $d(u, v) = \max(d(u_i, v_j))$. The results can be presented as a hierarchical diagram (*Fig. 2*).

A distance threshold is then selected so that the number of clusters with a root below this threshold is a reasonable value (e.g., 4–6). This is necessary to obtain an easily interpretable DEA model. For each

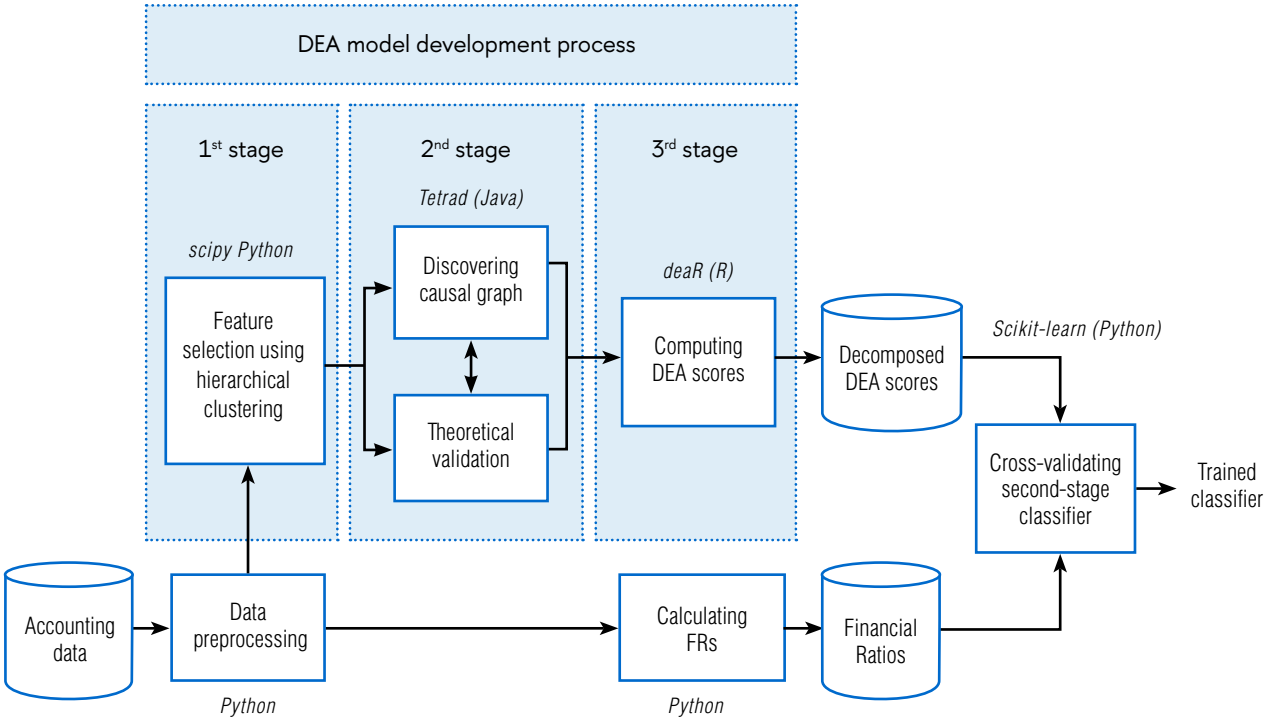


Fig. 1. Predictive model training and validation pipeline.

cluster with a root below the threshold, one feature is selected that has the highest correlation with the class label.

In *Fig. 2*, the absolute value of the Spearman's correlation coefficient ρ with the target variable (*Class*) for each attribute is presented in labels on the horizontal axis. The threshold distance is set at 0.55 (horizontal red dashed line). Using this presentation, we can select candidate features. For example, we select *VA* from the *EBIT* and *VA* pair forming the first cluster with root below threshold because $\rho_{VA,\ Class} = 0.229$ is greater than $\rho_{EBIT,\ Class} = 0.151$. Continuing with this reasoning, we get a set of potential variables for DEA: {*VA*, *SF*, *TA*, *FA*, *LTD*}.

Step two: causal analysis and theoretical validation. The next step is the specification of the DEA model, i.e., defining its inputs and outputs. We propose to base the DEA model specification on a causal structure underlaying selected features.

Formal definition of causality can be formulated as following: for two distinct events X and Y, X causes Y if a change in X corresponds to a change in Y, all else being equal [21]. A causal model is usually represented as a directed acyclic graph (DAG), which allows a compact and visual representation of the structure of the system under study [22]. The nodes of a DAG correspond to variables, the edges correspond to relationships between them, and the direction of the edges corresponds to causal relationships.

The ability of DAG to encode causal relationships is based on two assumptions [23]. The first is based on the graphical criterion of d-separation, which corresponds to the conditional independence of the variables in the dataset. In other words, for any three non-overlapping subsets of variables (X, Y, Z), if variables X and Y are conditionally independent given Z in the joint distribution $\mathcal{P}$ then they will be d-separated in graph $\mathcal{G}$ (Markov condition): $(X \perp\!\!\!\perp_{\mathcal{P}} Y)|Z \Rightarrow (X \perp\!\!\!\perp_{\mathcal{G}} Y)|Z$.

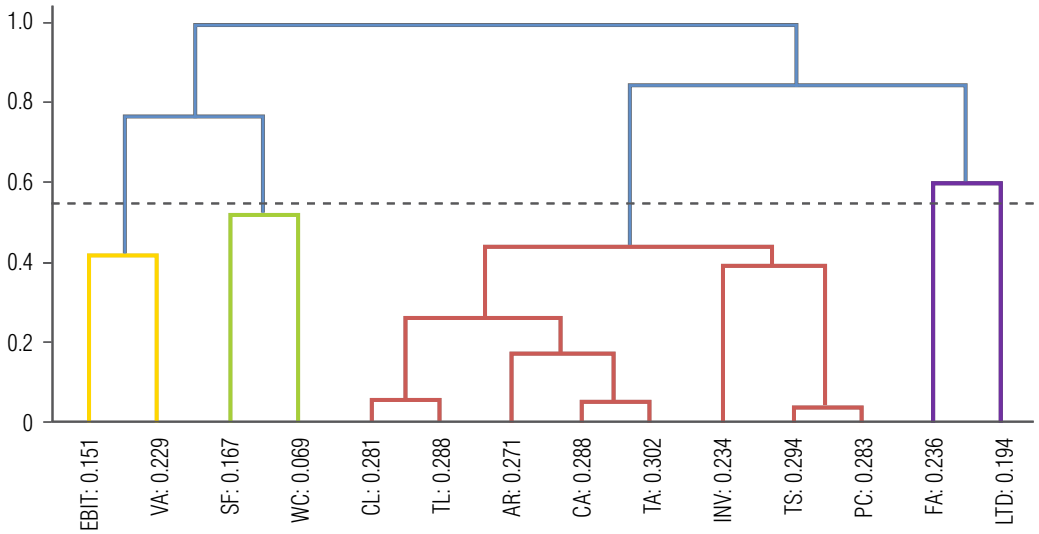


Fig. 2. Hierarchical clustering of accounting indicators. The abbreviations are explained in *Table 2*.

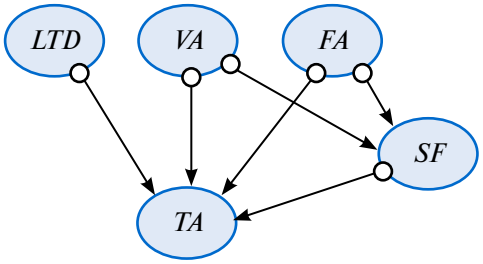
The second assumption is that there are no unobserved confounders (confounding variables affecting cause and effect simultaneously) and the sample is unbiased. Note that in practice this condition is extremely difficult to fulfil, so there are several algorithms (e.g. Fast Casual Inference – FCI) for which these restrictions are omitted, so we will use this algorithm. The result of FCI is a partial ancestral graph (PAG) that generalizes all equivalent DAGs that can be obtained from the dataset, so not all its edges are uniquely oriented [24].

FCI consists of two phases [21]. The structure search starts with a fully connected undirected graph, and then a sequence of conditional independence tests is performed. This removes edges between variables that are conditionally independent while blocking other variables. The second phase of FCI is the orientation procedure based on identification of possible *d*-separations.

Figure 3 shows the DAG discovered for the 2019 dataset using the Tetrad software toolkit[3] and a description of the edge types. As can be seen, the variables included in the set $S = \{LTD, VA, FA, SF\}$ are adjacencies with *TA*, and each variable in *S* is either the cause of *TA*, or there is an unmeasured confounder of S_i and *TA*, or both. On this basis, we will consider the variables *LTD, VA, FA, SF* as inputs to the DEA models and *TA* as output. Note that this specification is non-trivial and can hardly be derived from theoretical analysis. Therefore, it needs validation in terms of sense.

Recall that only small manufacturing firms are considered, and the features are selected according to their correlation with the class label (bankrupt/non-bankrupt). According to this specification, the firm uses resources such as long-term debt (*LTD*), value added (*VA*), shareholders' funds *(SF*) and fixed assets (*FA*) to produce total assets (*TA*). Total assets in addition to fixed assets (plant, equipment, and expensive

[3] https://www.ccd.pitt.edu



Graph Edge Types (Spirteset al., 2001)

Edge Types	Present Relationships	Absent Relationships
$X \rightarrow Y$	X is a cause of Y	Y is not a cause of X
$X \leftrightarrow Y$	There is unmeasured confounder of X and Y	X is not a cause of Y Y is not a cause of X
$X \circ\!\!\rightarrow Y$	Either X is a cause of Y or there is an unmeasured confounder of X and Y or both	Y is not a cause of X
$X \circ\!\!-\!\!\circ Y$	One of the following: (a) X is a cause of Y; (b) Y is a cause of X; (c) there is an unmeasured confounder of X and Y (d) both (a) and (c) (e) both (b) and (c)	

Fig. 3. Directed graphical causal model for 2019 dataset.

tools) include cash, accounts receivable, inventories, and intellectual assets that can be measured through accounting (such as patents, software, etc.) In the context of predicting bankruptcy, considering *TA* as an outcome makes sense because increasing the part of assets that is not fixed and therefore has more liquidity allows liabilities, both long-term and short-term, to be met. Moreover, the accumulation and utilization of intangible assets can provide a firm with additional capabilities, which is particularly important in crisis. Thus, this specification of the DEA model reflects one of possible scenarios for a firm's behavior in a crisis, which aims to increase the more liquid assets. Note also that current assets (*CA*), according to our model, are in the same cluster as *TA*, but have a lower correlation with the class label (*Fig. 2*). At the same time, experiments with setting the model output in the form of *CA* rather than *TA* showed that the estimates of such a model have less predictive power.

To compare the performance of the proposed approach (hereafter referred to as the DEA-1 model) with other, more obvious DEA specifications based on domain knowledge, let us also consider several other models. The first alternative model uses the same features as chosen in the first step. In general, inputs should represent the resources that are utilized to produce the firm's output, and the acquisition of these inputs represents a cost to the firm. Outputs, on the other hand, represent goods or services, whose sale generates revenue [13]. Going back to the basic idea of DEA – effective is the DMU that provides the maximum result with the minimum of resources, we can propose a DEA-2 specification (*Table 3*). From a financial management perspective, to maximize profit, *LTD*, *FA*, *SF, TA* can be minimized as inputs, and *VA* can be maximized as an output. This view identifies a technically efficient firm as using a minimum of resources yet producing a maximum of value added.

Another specification (DEA-3) can be derived using the approach of Feroz et al. [3], who considered the decomposition of Return on Equity (ROE) for profit-producing organizations on the basis of the DuPont model. Feroz et al. [3] study the DEA model that includes three inputs: costs, total assets and common equity, and total revenue as output. By selecting the features available in our dataset (*Table 2*) that are closest in meaning to those used in [3], we can define the DEA-3 inputs as *TA*, *SF*, *PC*, which must be minimized to maximize profit and output to be maximized as *EBIT*.

Table 3.

Alternative DEA models

Model	Inputs	Output	Spearman's rank correlation coefficient $\rho_{score,\ Class}$		
			PTE	SE	ME
DEA-1	LTD, VA, FA, SF	TA	–0.238	0.300	–0.243
DEA-2	LTD, FA, SF, TA	VA	–0.271	0.056	–0.235
DEA-3	TA, SF, PC	EBIT	–0.309	–0.229	0.253

Step three: computing DEA scores. In this step CCR, BCC and SBM scores are calculated; for this operation we used the R package deaR[4]. For the DEA-1 specification, according to the CCR model, 15 out of 6757 firms are technically efficient ($\theta_{CCR} = 1$)) meaning that they exist on the efficient frontier and there is no room for improvement in their output without increasing the inputs by the same proportion. These 15 firms are also mix efficient (SBM score is 1) as their slacks in inputs and outputs are zero. BCC results for the same specification show 38 firms to be technically efficient. 15 firms are those which are CCR efficient, and 23 firms are CCR inefficient but become BCC efficient, which is due to convexity condition in the BCC model.

Next, we transform resulting DEA scores according to the Eq. (3). Since the DEA scores are expected to be used as additional features in training a classifier that predicts bankruptcy, it is of particular interest to analyze the density of distribution of the scores obtained conditioned by the target (class label). *Figure 4* presents these distributions for the DEA-1 model. The header of each graph shows the Spearman's rank correlation coefficient of score with the target variable $\rho_{score,\ Class} = 0.1$. These correlation coefficients for all DEA models tested are also shown in *Table 3*.

Since target variable encoding firm status has a value of 1 for bankrupts and 0 for successful firms, it can be assumed that DEA scores should have a negative correlation with the class label. However, according to the results (*Fig. 4*) this assumption holds only for PTE and ME, SE showing a positive correlation. This requires a theoretical explanation.

In the input-oriented model, the PTE measures the extent to which a firm can reduce its resources while maintaining the same output to move to the VRS frontier. In other words, it measures the DMU's overall success in utilizing its inputs. The negative correlation between PTE and the target variable means that bankrupt firms are more likely than successful firms to use excess resources for the same output. This is confirmed by PTE distribution presented in *Fig. 4*.

Technical inefficiencies can arise not only from inappropriate output/resource ratios, but also from other sources [13]. Firms operating at inappropriate size (too big or too small) may display 'scale inefficiencies,' while others may use their inputs (or produce their outputs) in the wrong proportions ('mix inefficiencies'). Thus, ME measures the extent to which a DMU can benefit from a change in the balance of its

[4] https://CRAN.R-project.org/package=deaR

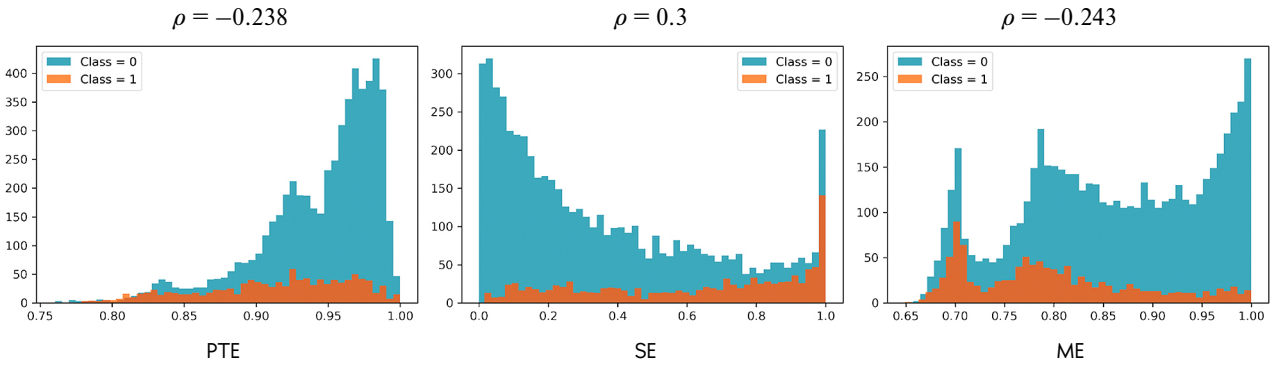


Fig. 4. Distribution density of PTE, SE, and ME.

inputs and outputs. *Figure 4* shows that successful firms maintain this balance more often than bankrupts.

In terms of SE, a significant number of both bankrupt and successful firms have the size to achieve the highest efficiency (SE is close to 1 in *Fig. 4*). For example, 296 of the 1124 bankrupt firms display SE $\geqslant$ 0.9. However, a large proportion of successful firms are scale inefficient. Perhaps this explains the ability of these firms to improve efficiency by adjusting scale to changing conditions during a crisis. Meanwhile, the rapid development of the crisis caused by the COVID-19 pandemic prevented those firms that were scale efficient but had low PTE and ME from optimizing through changes in scale. This is because reducing resources while maintaining output or rebalancing resources requires much more effort and time than reducing scale.

Note that the DEA-2 model has the same signs of correlation coefficients with target (*Table 3*), but the SE correlation is much weaker. It can be assumed that the use of DEA-2 scores as classifier features will give a smaller performance gain than DEA-1.

The DEA-3 specification, built using domain knowledge, offers a different perspective on efficiency as it looks at converting costs into profits. In this case, ME correlates positively with target (*Table 3*). This can be interpreted similarly to the above. Firms that maintained an optimal balance of resources (in this case *TA*, *SF* and *PC*) were limited in their choice of responses to the crisis and were therefore more likely to go bankrupt. Since the ρ values for this specification are higher than for DEA-2, we can expect that using these scores will produce a more accurate predictive model.

4. Results

All proposed models (*Table 3*) were tested on 2019 data. We used as predictors not only LR, which is considered in most of the previous studies (cf. *Table 1*), but also two ensemble techniques, AdaBoost (AB) and Random Forest (RF).

AB implements a sequential scheme in which each new estimator added to the ensemble is trained on the errors of the previous configuration. We used Decision Tree (DT) as the base estimator with a depth of 1. Other hyperparameters: the number of estimators is 100; the learning rate is 0.1.

In the RF model, each base estimator is trained in parallel on a subset of the data, with both the number of observations and the number of features chosen at random. The ensemble decision is an averaging of the predictions of all the estimators. The RF hyperparameters used in our experiments are as follows: the number of estimators is 100, the number of features considered in the search for the best split is calculated as

the square root of the total number of features, and the base estimator is DT without depth limit.

The datasets should allow comparison of model performance on different data configurations. Therefore, we consider four datasets: FR – the dataset includes financial ratios and DEA-*i* (i = 1, 2, 3) – the dataset includes FRS and scores of models DEA-1, DEA-2, DEA-3 correspondingly.

Since all datasets are unbalanced, the quality metric is the Area Under the Receiver Operating Characteristics Curve (AUC-ROC). *Table 4* shows the average AUC-ROC obtained by 10-folds cross-validation and its standard deviation (in brackets). The best values according to the Wilcoxon test are shown in bold.

Analyzing the *Table 4*, the following conclusions can be drawn. The inclusion of DEA scores in the list of features significantly improves performance. The overall improvement in AUC-ROC obtained with DEA-1 scores is 20.3% for AB, 15% for RF and 42.9% for LR compared with results achieved with FRs alone. Among the DEA specifications considered, the best results are achieved with the model DEA-1 proposed here. Note also that the other two DEA specifications display not only to a larger bias, but also a larger variation (i.e., standard deviation).

The proposed method has also been applied to the 2018 data, which corresponds to a more stable economic period. The list of features selected using agglomerative clustering has changed slightly compared to 2019, *TS* is included instead of *TA*.

Figure 5 shows the DAG created for 2018. As we can see, there is no edge between *LTD* and any other feature, so we excluded it from the final DEA specification. The available data do not allow us to define the exact type of relations between *FA*, *SF* and *TS*. However, that is more important in the DEA specification context; all of them are either the cause of *VA*, or there are unspecified confounders that determine joint behaviour of these variables and *VA*. Therefore, we define *FA*, *SF* and *TS* as model inputs and *VA* as output. This means that a DMU, to be efficient, must minimize the various sources of funds (sales revenue, shareholder's equity, and production capacity) to maximize the value created in the production process. This specification is more in line with the domain knowledge traditionally used to design DEA models. Perhaps this type of specification is because we are using data from a stable economic period to determine it.

Table 5 lists the 10-folds cross-validation AUC-ROC results for 2018. As above, the results in the FR column are obtained on data including only the FRs, while the DEA column presents results obtained on data including FRs and PTE, SE, and ME scores. As can be seen, the improvement achieved by using DEA scores, although smaller than in 2019, is significantly better than previous work (*Table 1*), which confirms the efficiency of the proposed method.

Table 4.

Model testing results for 2019 dataset

Second-stage classifier	Dataset			
	FR	DEA-1	DEA-2	DEA-3
AB	0.692 (0.036)	**0.832 (0.041)**	0.778 (0.047)	0.826 (0.057)
RF	0.675 (0.037)	**0.777 (0.047)**	0.756 (0.050)	0.760 (0.056)
LR	0.567 (0.029)	**0.811 (0.049)**	0.757 (0.042)	0.797 (0.060)

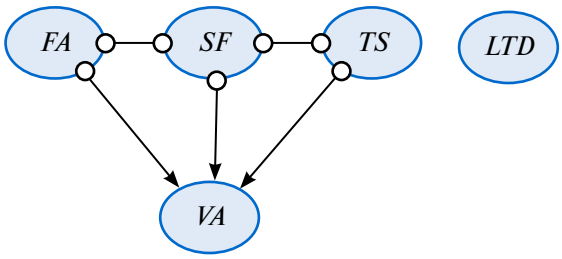


Fig. 5. Directed graphical causal model for 2018 dataset.

The results presented in *Tables 4* and *5* show that this approach improves prediction accuracy. The increase in AUC-ROC for the worst-case model (RF) is over 5% for 2018 and over 15% for 2019. For the best models, this is AB for 2019 and LR for 2018, the improvement in AUC-ROC is 20.3% and 25.8% respectively. This is because the method proposed here allows generation of features with a much higher correlation with the target variable (*Fig. 4*) than FRs (*Fig. 2*).

In addition, since the DEA model allows for significant improvement in forecasting accuracy, its specification contains important knowledge about the factors leading the financial failures of the firm. These factors are discussed above; the most important findings are summarised here:

- DEA specifications that maximise predictive ability differ for crisis and stable periods. The specification for stable periods is more consistent with the idea of a profit-oriented firm.
- Increasing the share of liquid assets provides additional opportunities, (e.g., refocusing or downsizing the business), which is particularly important in a crisis.
- SE scale inefficiencies can be an advantage in times of crisis, as such firms can improve efficiency by adapting the scale of the business. Optimization by reducing resources while maintaining productivity (PTE improvement) or reallocating resources (ME improvement) requires much more effort and time than rescaling. The analysis of feature importance based on permutations (impurity) shows that SE makes an extremely large contribution to improving classification performance for the crisis period.

Conclusion

This article presents a method for developing DEA models whose scores are then used as features to improve the quality of bankruptcy predictors. The method includes three steps: (1) selecting accounting measures, (2) exploring relationships between selected measures to determine DEA inputs and outputs, and (3) calculating DEA scores. Experiments with two datasets containing data from Russian small manufacturing firms showed that the inclusion of PTE, SE, and ME in the list of features during classifier training improves AUC-ROC by more than 20%, which significantly exceeds the results of previous works. Moreover, the analysis of DEA models developed in this way provides insight into the reasons for a firm's failure in both stable and crisis periods. ■

Table 5.

Model testing results for 2018 dataset

Second-stage classifier	FR	DEA	Improvement
AB	0.680 (0.038)	**0.725 (0.059)**	1.066
RF	0.649 (0.044)	**0.684 (0.041)**	1.054
LR	0.591 (0.033)	**0.744 (0.089)**	1.258

References

1. Zhao J., Ouenniche J., De Smedt J. (2024) Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. *Machine Learning with Applications*, article 100527. https://doi.org/10.1016/j.mlwa.2024.100527

2. Smith P. (1990) Data envelopment analysis applied to financial statements. *Omega*, vol. 18(2), pp. 131–138.

3. Feroz E.H., Kim S., Raab R.L. (2003) Financial statement analysis: A data envelopment analysis approach. *Journal of the Operational Research Society*, vol. 54(1), pp. 48–58. https://doi.org/10.1057/palgrave.jors.2601475

4. Fernandez-Castro A., Smith P. (1994) Towards a general non-parametric model of corporate performance. *Omega*, vol. 22(3), pp. 237–249.

5. Zelenkov Y., Volodarskiy N. (2021) Bankruptcy prediction on the base of the unbalanced data using multi-objective selection of classifiers. *Expert Systems with Applications*, vol. 185, article 115559. https://doi.org/10.1016/j.eswa.2021.115559

6. Burova A., Penikas H., Popova S. (2020) Probability of default (PD) model to estimate ex ante credit risk. *Ban*k *of Russia*. Working Paper Series, no. 66. Available at https://www.cbr.ru/StaticHtml/File/116472/wp-66_e.pdf (accessed 01 August 2025).

7. Tangsawasdirat B., Tanpoonkiat S., Tangsatchanan B. (2021) Credit risk database: Credit scoring models for Thai SMEs. *Puey Ungphakorn Institute for Economic Research*. Discussion paper, no. 168. Available at: https://www.pier.or.th/en/dp/168/ (accessed 01 August 2025).

8. Benítez-Peña S., Bogetoft P., Morales D.R. (2020) Feature selection in data envelopment analysis: A mathematical optimization approach. *Omega*, vol. 96, article 102068. https://doi.org/10.1016/j.omega.2019.05.004

9. Li Z., Crook J., Andreeva G. (2014) Chinese companies distress prediction: An application of data envelopment analysis. *Journal of the Operational Research Society*, vol. 65(3), pp. 466–479. https://doi.org/10.1057/jors.2013.67

10. Mousavi M.M., Ouenniche J., Tone K. (2019) A comparative analysis of two-stage distress prediction models. *Expert Systems with Applications*, vol. 119, pp. 322–341. https://doi.org/10.1016/j.eswa.2018.10.053

11. Charnes A., Cooper W.W., Rhodes E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, vol. 2, pp. 429–444.

12. Banker R.D., Charnes A., Cooper W.W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Scienc*e, vol. 30, pp. 1078–1092.

13. Cooper W.W., Seiford L.M., Tone K. (2006) *Data envelopment analysis: A comprehensive text with models, applications, references, and DEA-Solver software.* Springer.

14. Premachandra I.M., Chen Y., Watson J. (2011) DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega*, vol. 39(6), pp. 620–626. https://doi.org/10.1016/j.omega.2011.01.002

15. Shetty U., Pakkala T.P.M., Mallikarjunappa T. (2012) A modified directional distance formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India. *Expert Systems with Applications*, vol. 39(2), pp. 1988–1997. https://doi.org/10.1016/j.eswa.2011.08.043

16. Ouenniche J., Tone K. (2017) An out-of-sample evaluation framework for DEA with application in bankruptcy prediction. *Annals of Operations Research*, vol. 254(1), pp. 235–250. https://doi.org/10.1007/s10479-017-2431-5

17. Xu X., Wang Y. (2009) Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications*, vol. 36(1), pp. 366–373. https://doi.org/10.1016/j.eswa.2007.09.040

18. Yeh C.C., Chi D.J., Hsu M.F. (2010) A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, vol. 37(2), pp. 1535–1541. https://doi.org/10.1016/j.eswa.2009.06.088

19. Psillaki M., Tsolas I.E., Margaritis D. (2010) Evaluation of credit risk based on firm performance. *European Journal of Operational Research*, vol. 201(3), pp. 873–881. https://doi.org/10.1016/j.ejor.2009.03.032

20. Tone K. (2002) A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, vol. 143(1), pp. 32–41. https://doi.org/10.1016/S0377-2217(01)00324-1

21. Spirtes P., Glymour C., Scheines R. (2001) *Causation, prediction, and search*. The MIT Press.

22. Koller D., Friedman N. (2009) *Probabilistic graphical models: principles and techniques*. MIT press.

23. Zelenkov Yu.A., Lashkevich E.V. (2024) Counterfactual explanations based on synthetic data generation. *Business Informatics*, vol. 18(3), pp. 24–40. https://doi.org/10.17323/2587-814X.2024.3.24.40

24. Zhang J. (2008) On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, vol. 172(16–17), pp. 1873–1896. https://doi.org/10.1016/j.artint.2008.08.001

About the author

Yuri A. Zelenkov

Doctor of Sciences (Technology);

Professor, Department of Business Informatics, Graduate School of Business, HSE University, 28/11, Shabolovka Str., Moscow 119049, Russia;

E-mail: yzelenkov@hse.ru

ORCID: 0000-0002-2248-1023