https://doi.org/10.31107/2075-1990-2022-6-74-90

Using Market Indicators to Refine Estimates of Corporate Bankruptcy Probabilities

Daria S. Leonteva
HSE University, Moscow 101000, Russian Federation
dsleonteva@edu.hse.ru, https://orcid.org/0000-0002-4571-7676

Abstract
This study investigates an alternative approach to estimating the probability of default. The introduction of credit spreads as market measures of default into an accounting-based model attempts to enhance the predictive power of classical approach models which analyze only balance sheet data. This paper identifies which of the two market measures of credit spread — the Z-spread or the I-spread — has an advantage in the context of robustness of the bankruptcy prediction models. Using two techniques — logistic regression and a gradient boosting machine approach, as well as a sample of annual series of 80 financial ratios for 385 U.S. listed companies which issue corporate bonds — evidence is obtained that the I-spread has higher predictive power in both techniques. The better performance of the I-spread can be explained by the fact that the accuracy of the Z-spread calculation can be misleading because different methods of interpolation of the yield curve are used. In addition, the predictive power of the chosen techniques is also compared. The up-to-date gradient boosting machine framework performs better on the test sample. These findings may encourage managers to implement additional market characteristics in the analysis and apply modern techniques rather than the classic ones — logistic regressions and multiple discriminant analyses models — to predict inconsistency in corporate performance.

Keywords: *bankruptcy prediction, credit spreads, logistic regression, gradient boosting machine*

JEL: *C53, G33*

Note: *The article is based on the master's paper thesises defended at the Higher School of Economics. Study has been discussed on the XXII April International Academic Conference on Economic and Social Development, Session L-19-2. Corporate Governance and Finance, April 13–30, 2021.*

For citation: *Leonteva D.S. (2022). Using Market Indicators to Refine Estimates of Corporate Bankruptcy Probabilities. Financial Journal, vol. 14, no. 6, pp. 74–90. https://doi.org/10.31107/2075-1990-2022-6-74-90.*

INTRODUCTION

The procedure of bankruptcy prediction may help to identify the symptoms of insolvency and to reduce the risks of bankruptcy and failure in the future. The problem of the enterprise default prediction is a headache not only for creditors and other investors, but for public authorities as well. Stakeholders require continuously updated information on the probability of corporate default risk. According to Bloomberg, in North America alone, the share of bankrupt companies issuing corporate bonds in 2020 is 68%. In this regard, the study of U.S. companies will help to provide an unbiased understanding of the bankruptcy prediction. This study focuses on predictive power rather than explanatory modeling and, therefore, on listed American companies issuing corporate bonds.

The classical approach to bankruptcy prediction is based on the application of accounting ratios. This paper attempts to introduce market indicators into the accounting-based approach and create the hybrid model — a combination of accounting-based model and market spreads. Thus, two types of credit spreads — the Z-spread and the I-spread — are considered as market indicators for analysis. The main reason for the research is to see whether the predictive power of the classical approach can be enhanced.

The purpose of this paper is twofold: to suggest an approach to estimatiing the probability of corporate failure using not only accounting but also market data; and to compare the predictive power of two chosen techniques used in assessing default: the logit model and the gradient boosting approach (machine learning technique).

Hence, the research questions are as follows:

— Does the choice of credit spread specification affect the estimates of the probability of default?

— Does the up-to-date machine learning technique have advantages in predicting bankruptcy compared to the well-known and widely used logit model?

This paper contributes to the existing literature on predicting corporate failures for several reasons. The classical approach based only on accounting data is complemented by an investigation of the advantage of a particular spread among Z-spreads and I-spreads. Besides, the application of a new method based on gradient boosting represents an advancement in the use of new techniques to predict U.S. companies' failures.

LITERATURE REVIEW

The economic literature offers a variety of different models for forecasting enterprises' bankruptcy [Beaver, 1966; Zmijewski, 1984] and is still relevant for further studies. Today, various models of bankruptcy probability assessment are used, based on multiple principles and methods. Among the most commonly used methods are the multiple discriminant analysis (MDA), binary choice models such as logit- and probit-models, and neural networks (NN). These methods have been challenged over time. Altman constructed the first multifactor linear discriminant model [Altman, 1968]. He tried to find a linear combination of factors out of many to predict SMEs (small and medium-sized enterprises) bankruptcy with the highest possible accuracy at that time. This model is considered to be one of the widely used models of bankruptcy prediction. Regardless of its acknowledgement, this LDA (linear discriminant analysis) method was critically analyzed by [Wei, Li and Chen, 2007]. They found that LDA algorithm can misclassify bankruptcy outcomes. Along with their research, the application of LDA and QDA (quadratic discriminant analysis) was also criticized by [Ohlson, 1980] and [Wilson, Sharda, 1994] due to the fact, that financial ratios frequently lack the Gaussian distribution approach which is the main assumption of the modelling. A significant step in the development of enterprise bankruptcy forecasting was the work with binary choice models. Ohlson [Ohlson, 1980] challenged logit- and probit-models and provided ample evidence of their superiority in predictive power compared to the LDA and QDA models. This view is shared by [Makeeva, Neretina, 2013] in their analysis of bankruptcy prediction in the construction industry. Nowadays, many models based on modern economic and mathematical methods, including neural networks approach, artificial intelligence models, and classification trees, are widely used. The application of neural networks is presented in [Tam, Kiang, 1992] and [Makeeva, Bakurova, 2012]. The main drawback of NN is considered to be the problem of overfitting.

The key feature of this paper is the comparison of two approaches: the logit model and the gradient boosting approach. Logistic models require strong traditional assumptions of conventional statistics, such as linearity, normality, and independence between predictor variables [Kim, Kang, 2010]. All these restrictions limit real-world applications and urge researchers to search for new modern techniques. In this study, we apply logistic regression as a traditional binary variable prediction approach. Logit models are still widely used by risk managers in companies. Gradient boosting is a powerful machine learning algorithm. Several recent studies have underlined the superiority of AdaBoost (one of the most popular boosting algorithms) in bankruptcy prediction accuracy over other approaches. Alfaro et al. [Alfaro et al., 2008] show that the AdaBoost algorithm with a classification tree, which is used in data science modelling, decreases the generalization error by about 30% compared to the error obtained with neural networks. Another commonly used approach is extreme gradient boosting (XGBoost). This machine learning approach is used to predict bank failures [Carmona et al., 2019]. The authors show that the XGBoost algorithm outperforms both Logistic and Random Forest methods in the probability of default in the U.S. banking sector.

Comparing logit models and the gradient boosting technique in various practical cases, Dias et al. found that the modern approach yields greater or equal results [Dias et al., 2018]. This can be explained by the fact that gradient boosting has variable exclusion because it represents an inverse interpretation of the business sense. At the same time, logistic regression very often excludes variables for this very reason. Moreover, superior predictive power is achieved with less development effort. We want to compare the traditional and the up-to-date technique to encourage managers to implement novel approaches in their research. Another strong approach is the random forest — one of the most common supervised classification algorithms. It is based on decision tree models. The random forest tries to build multiple tree models with different samples and different variables, creating a forest with a certain number of trees [Carmona et al., 2019]. Zięba et al. compared the predictive performance of conventional approaches and recent artificial intelligence methods. They examined the supremacy of Gradient Boosting approach over the random forest algorithm [Zięba et al., 2016].

Furthermore, besides individual bankruptcy assessment approaches there are findings based on a combined approach to predicting bankruptcy. Trujillo-Ponce and Samaniego-Medina apply both financial ratio data and market data to explain corporate credit risk of credit default swap (CDS) spreads quoted in the European market between 2002 and 2009. They show that the two approaches complement each other, and the hybrid model seems to be the best approach to explain corporate credit risk [Trujillo-Ponce et al., 2012]. Other authors have recognized the combined approach and attempted to implement it in their research. Tinoco and Wilson complement the hybrid analysis with proxies for changes in the macroeconomic environment. The authors offer a comparative analysis of the 'full model' fitted with market variables, 'accounting only' and 'accounting plus macroeconomic variables' models. According to the chosen measure of model efficiency, Area Under Receiver Operating Characteristics Curve (AUC), which incorporates market variables in the accounting model, has a higher AUC [Tinoco, Wilson, 2013]. This indicates that the market measures contain information which is not included in the financial performance ratios. This finding underlines the relevance of the hybrid approach consideration.

Corporate bond spreads are chosen as a market-based measure for predicting corporate bankruptcy because they reflect market movements. Anginer and Yildizhan show that spreads predict corporate defaults better than previously used measures such as bond ratings and accounting variables. They argue that credit spreads have superiority among other measures that are used to predict corporate default in hazard rate regressions [Anginer et al., 2010].

However, the choice of this variable can be controversial, and several questions about spreads endogeneity remain unaddressed. Almeida and Philippon argue that bond spreads contain no historical probabilities of distress. They contain a systematic component of corporate default risk generated by macroeconomic conditions. Thus, corporate bond spreads do not comprise the true probability of default [Almeida, Philippon, 2007].

The literature underlines that credit spreads are not fully explained by expected default. A wide range of papers have attempted to evaluate the share of default risk in the yield spread. For example, Huang and Huang argue that corporate bond spreads of all maturities contain only a small fraction of credit risk [Huang, Huang, 2002].

The above findings regarding the probability of default prompted the author to analyze the effect of introducing credit spreads into the accounting-based approach and to apply a new modern technique that is not inconsistent with existing research.

DATA SELECTION AND MODELLING

This study attempts to predict corporate distress of U.S. companies which falls in one calendar year through the period 2013–2018. As stated in the previous section, the problem of predicting bankruptcy is a classification problem in terms of whether or not a company will go bankrupt. Therefore, the classification problem can be represented as a binary variable, where 0 is a U.S. company which operated during the 2013–2018 period, and 1 is a U.S. company which went bankrupt during the same period. The classification problem is considered to be a supervised learning problem applicable in the data science framework. To test the predictive power of different models — the accounting-based model and the hybrid model with Z-spread and I-spread as its add-ons, two approaches are implemented: logistic modelling and the gradient boosting machine technique.

The analysis is based on data from 385 U.S.-based public companies which attract liquidity by issuing corporate bonds. The panel data cover the period from November 2013 to December 2018. The choice of this time period is due to the limitations of the data sources. Two data sources are considered for analysis: Capital IQ and Bloomberg. Capital IQ contains historically limited information (from 2010 only). As for Bloomberg, its dataset tends to be depleted for early periods (before November 2013). Such data has many missing values (more than 90%). The final sample consists of 203,490 observations and is randomly split into two subsamples — training and test with a split ratio equal to 0.7 [Vatcheva, 2016]. All results were obtained and analyzed by programming in *R*. The choice of the programming language could be conditioned by the possibility of applying the process of binning weights of evidence. Other languages (such as Python) do not have packages for its implementation.

The accounting data and bankruptcy dates were taken from the Capital IQ database. The accounting data consists of approximately 80 annual financial indicators obtained from the companies' balance sheets. This paper takes into account only bankruptcy cases with liquidation dates according to the Capital IQ database. Overall, 14 financially distressed companies are analyzed. The market indicators are taken from the Bloomberg database. The companies' tickers are collected from Capital IQ and their market data are downloaded from Bloomberg. There are also some restrictions imposed on the sample: bonds issued in excess of USD 100 million are included. In order to limit the study to the classical bonds forms and to reduce the stochastic component [Fabozzi, 2007], the analysis excludes the floating coupon type; the convertible, putable, callable maturity types; and the subordinated and junior subordinated collateral types. Data contain weekly bond quotes, maturity and coupon types, coupons and frequencies, ratings, weekly I-spreads and weekly Z-spreads. Spreads are calculated by Bloomberg and reflected in BLP_I_SPRD_MID and BLP_Z_SPRD_MID fields in the Excel Add-in, respectively. The Bloomberg Mid I-Spread is calculated between the selected bond and the interpolated yield curve from the swap curve using Yield and Spread Analysis. It is based on the selected bond's nominal maturity date. The Bloomberg Mid Z-Spread, in turn, reflects the value that must be added to the swap spot curve so that the security's discounted cash flows equal its mid-price, with each dated cash flow discounted at its own rate.

Data on financial ratios and market measures are matched to company tickers. First, tickers taken from Capital IQ are used to search for market characteristics in Bloomberg. The imperfection of the data sources leads to a decrease in the number of observations. Moreover, the result of this procedure is the number of bonds with various ISINs (International Securities Identification Numbers) of one particular ticker. Such ISINs contain minimal number of missing values.

In order to match dimensions (annual accounting data and weekly market data), weekly market measures are transformed into annual by simply taking the arithmetic mean. This study considers only the last 3 months of each year of historical spreads. Thus, the resulting dataset contains annual data.

With all limitations, the sample is representative of the percentage of bankruptcies in the general population. This paper attempts to develop bankruptcy prediction models related to the time of one year before bankruptcy.

METHODOLOGY

The main models of interest are the accounting-based model and the hybrid model with Z- and I-spreads as its add-ons. To test the predictive powers of these models, two approaches are challenged: logistic modelling and gradient boosting machine.

Logistic Regression and Gradient Boosting Machine

The logit model is characterized by logistic distribution:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_{1,i} + \cdots + \beta_k \cdot x_{k,i})}}, \tag{1}$$

where i is a U.S. company in 2013–2018, P_i — a binary target variable (0 — company operates, 1 — company goes bankrupt), $x_{k,i}$ — independent variables (accounting data, Z-spread and I-spread), β_k — parameters of the model.

The Gradient Boosting Machine framework was originally proposed in [Friedman, 2001]. As the target is a binary variable, $y \in \{0,1\}$ — the classification problem is solved. The main principle of GBM is to get an estimate or approximation $\hat{f}(x)$ of the function $f^*(x)$ mapping x to y in such a way that the expected value of some specified loss function $L\big(y, f(x)\big)$ is minimized.

$$f^*(x) = \arg\min_f E_{y,x} L\big(y, f(x)\big). \tag{2}$$

Negative binomial likelihood loss (or Bernoulli loss) function is applied for the classification problem. This paper also chooses *M* = 0.01 as a number of iterations or the total number of trees to fit, and specifies the hyperparameter: the maximum depth of each tree is 2, and shrinkage (learning rate) is 0.01 [Touzani, 2018]. A higher learning rate could lead to greater sensitivity of the algorithm to both the number of iterations and the depth of the decision trees. On this sample, a shrinkage rate of 0.01 leads to an optimal convergence rate. Furthermore, due to the increasing number of iterations and the model complexity, the algorithm starts to over-fit the training data.

In the classification problem, loss function is as follows:

$$L\big(y, f(x)\big) = \log(1 + \exp(-2yf)), \text{ where } f(x) = \tfrac{1}{2}\log\left[\frac{\Pr(y=1|x)}{\Pr(y=0|x)}\right]. \tag{3}$$

According to Friedman and his algorithm for gradient boosting realization, in this case [Friedman, 2001]:

For every iteration $m = \overline{1, M}$:

1) calculate pseudoresponses

$$\widetilde{y_i} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)} = \frac{2y_i}{(1+\exp(2y_i \cdot f_{m-1}(x_i)))}, \tag{4}$$

$i = \overline{1, N}$, (N — number of observations)

2) find new base algorithm as a regression on pseudoresponses $h(x, a)$
3) find the optimal search line

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} \log\left(1 + \exp\left(-2y_i (f_{m-1}(x_i) + \rho h(x_i, a_m))\right)\right). \tag{5}$$

4) update approximation

$$f_m(x) = f_{m-1}(x) + \rho_m h(x, a_m). \tag{6}$$

GBM has become extremely popular over the last decade and has come to dominate the recent conversation about boosting. Moreover, it often performs better than any other algorithm.

Credit Spreads

Two types for credit spreads are analyzed: the I-spread and the Z-spread.

The I-spread, or interpolated spread, is known as the difference between the corporate bond internal rate of return (IRR) and the interpolated yield to maturity of treasury bonds. The method of interpolation can be anything, for example, linear or cubic.

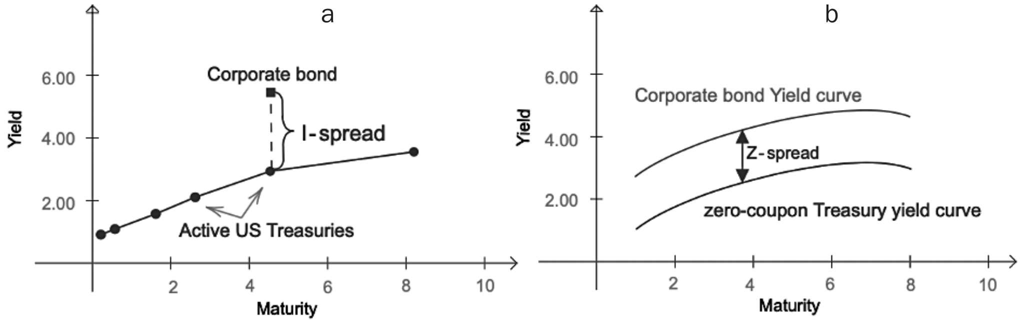
The Z-spread, or zero-volatility spread, measures the spread that an investor will get over the entire Treasury yield curve. It is the spread that must be added to each spot interest rate in order for the price of the bond to equal to the sum of its cash flows.

$$P_{bond} = \sum_k CF_k * e^{-t_k * (z(t_k)+x)}. \tag{7}$$

where x is the Z-spread, CF_k — cash flows of a particular bond, t_k — time to maturity, $z(t_k)$ — the zero spot rates for t_k maturity. In this paper, continuous discounting is assumed.

Figure 1

I-spread (a) and Z-spread (b) illustration



Source: compiled by the author.

There are several methods of fitting zero-coupons (yield curve smoothing). The most commonly used are Bootstrap (standard, blocked, iterative), Spline (polynomial, cubic) and Nelson-Siegel [Lapshin, 2018].

Weight of Evidence Binning

Binning is a widely using technique (especially in credit scoring) for converting continuous variables into categorical variables. This categorization process also deals with missing values — NAs (not available). Such transformation could be done according to the weight of evidence (WOE) technique. WOE is a quantitative method of combining evidence to support a statistical hypothesis [Good, 1985]. It compares the proportion of good-to-bad cases at each attribute level. Then it measures the strength of the attributes of an independent variable for good and bad cases separately. After the binning procedure, the divergence measure as the information value (IV) could be calculated. IV is a numerical value to quantify the predictive power of the independent continuous variable in capturing the binary dependent variable [Zeng, 2013]. Siddiqi proposed the following thresholds for quantifying IV: <0.02 — unpredictive; [0.02, 0.1) — weak, [0.1, 0.3) — medium, ≥0.3 — strong [Siddiqi, 2006].

In this study, the WOE binning transformation is applied in order to cope with missing values and to select the first potential set of predictors for the logit model according to IV.

The Area Under the ROC-curve

The ROC-curve (Receiver Operator Characteristic) is a graphical interpretation of the dependence of the proportion of correctly defined positive classifications on the proportion of falsely defined negative examples when varying the decision rule. This graph allows determining the quality of the binary classification and ranking classifiers to visualize their performance [Fawcett, 2003]. The decision rule is implemented by selecting a decision threshold which separates positive and negative classes.

The construction of the ROC-curve is determined by the following classification matrix identical to the matrix of Type I, II errors:

Table 1

Confusion Matrix

Decision about H0	H0	
	True	False
Reject	True Negative TN	False Negative FN (Type II Error)
Fail to Reject	False Positive FP (Type I Error)	True Positive TP

Source: [Hajian-Tilaki, 2013].

The objective value of any binary classifier is due to the sensitivity and specificity of the model.

$$Se = \frac{TP}{TP+FN} * 100\%, \tag{8}$$

$$Sp = \frac{TN}{TN+FP} * 100\%. \tag{9}$$

Sensitivity (2) is the proportion of positive cases determined by the ratio of true positive cases to the total number of actual positive cases. Specificity (3) shows the proportion of true negative cases that were correctly identified by the model. It is determined by the ratio of correctly defined negative cases to the total number of actual negative cases.

In the analysis, the Area Under the ROC-curve is used as an evaluation metric. In order to match the quality of the model with the obtained AUC value, the exemplary classification is used:

Table 2

AUC Classification

AUC value	Quality of the model
0.9–1.0	Excellent
0.8–0.9	Good
0.7–0.8	Fair
0.6–0.7	Poor
0.5–0.6	Fail

Source: [Gorunescu, 2011].

The closer AUC is to 1, the better, because quality of the model is higher. However, in some cases AUC can be misleading. It is important to look not only at AUC but also how the shape of each curve indicates how model performs across the range of predictions.

RESULTS AND ANALYSIS

Logit Model Fitting

This study attempts to train the logit model. As our sample consists of missing values, it is necessary to perform the process of binning the features according to the weight of evidence. In accordance with the predictors' information value, the chosen features lie within medium and strong information values (≥0.1). For the formed features list, a long list of predictors is chosen for the accounting-based model according to the one-factor analysis of the area under ROC-curve and its correlation. After the long list of suitable parameters, a short list is constructed by excluding features with multicollinearity. The choice of balance sheet parameters is limited to the four main predictors according to the economic sense in terms of the consistency of companies' business conduct which can thus be applied in the model.

Table 3

Logit Model Parameters

Parameter	AUC
Cash and Equivalents	0.7347
Accounts Payable Assets	0.7330
Total Common Equity	0.7159
Unearned Revenue, Current	0.6762

Source: compiled by the author.

The selected parameters have a sufficient AUC value and can certainly be included in the logit model. All predictors have a correlation with each other of no more than 0.7.

Figure 2 illustrates the correlation matrix of the chosen features for the logit model. The predictors are numbered according to the list:

[1] Accounts Payable Assets
[2] Cash and Equivalents
[3] Total Common Equity
[4] Unearned Revenue, Current
[5] Z-spread
[6] I-spread.

Figure 2

Correlation matrix of the Logit model predictors

	1	2	3	4	5	6
1	1	−0.45	−0.22	−0.4	−0.21	−0.21
2	−0.45	1	0.32	0.39	0.27	0.27
3	−0.22	0.32	1	0.24	0.12	0.12
4	−0.4	0.39	0.24	1	0.34	0.34
5	−0.21	0.27	0.12	0.34	1	1
6	−0.21	0.27	0.12	0.34	1	1

Source: compiled by the author.

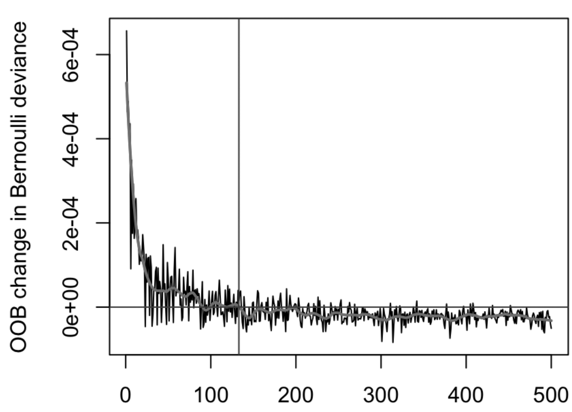
The accounting-based model consists of the first 4 features. The hybrid model using spreads as market measures consists of all the features of the accounting-based model and one of the spreads: the Z-spread or the I-spread. Only one spread is added to the model because in this paper we need to investigate the impact of including a particular spread and analyze the difference in predicted power behind the two spreads. All predictors have no multicollinearity except for the spreads. The perfect correlation between the two spreads can be explained by the fact that they both reflect the market. The main difference between the spreads is in their design. By including spreads in the model, an alternative approach with the market measure implementation is tested. By constructing two hybrid models with different spreads, the question of which spread is more suitable for predicting bankruptcy is answered. Thus, in total, three models are considered in this paper: the accounting-based model and two hybrid models with spreads.

Gradient Boosting Machine Fitting

Another approach in the context of this paper is the gradient boosting machine. The gradient boosting procedure is applied to the binning sample. Multicollinearity and one-factor feature analysis were not tested because the GBM algorithm does not require such restrictions.

Figure 3

“Out-of-Bag” early stopping time



Source: compiled by the author.

First of all, we need to define the early stopping time. The term "early stopping" is used to describe the process of stopping the training process in an iterative algorithm by evaluating the model performance on the entire dataset. In the case of GBM, it may be based on the performance of the "Out-of-Bag" (OOB) algorithm. It is often argued that the OOB error is an unbiased estimate of the true prediction error [Mitchell, 2011]. The ideal time to stop training the model is after the validation error has decreased and then stabilized and before the validation error has increased due to the overfitting.

In Fig. 3, the red curve illustrates the validation error depending on the number of trees, the black graph illustrates the training error. Next, it is identified that the optimal number of trees according to the chosen method of early stopping is 133.

After performing the gradient boosting machine on all the features, four main predictors are chosen according to relative influence parameter. Relative influence is a measure that quantifies how useful certain variables are in training the model.

Table 4

Gradient Boosting Model Parameters

Parameter	Relative influence
Cash and Equivalents	3.4124
Total Common Equity	2.4704
Other Current Assets	0.8247
Total Cash and Short-Term Investments	0.8162

Source: compiled by the author.

These parameters form the accounting-based model for gradient boosting machine. Hybrid models are still assumed with the addition of spreads. Comparing the short list of parameters chosen for the logistic regression and for the gradient boosting machine, two common features can be observed: Cash and Equivalents and Total Common Equity. The other parameters differ. It is important to mention that different types of assets are also added in the list for both techniques.

Model Comparison via ROC Curve & Gini Metrics

In this section, model comparison is performed across all types of models: the logit specification and the gradient boosting machine for both accounting-based models and hybrid models with spreads. The Gini metric is used as the measure of model performance for comparing binary classification models. The Gini is a proportional measure to the AUC.

$$Gini = 2 \cdot AUC - 1 \qquad (10)$$

The choice of the Gini parameter is explained by the frequency of its application in bankruptcy prediction tasks. The test set of predictions of each of the models is used to calculate the Gini of the test set. The model with the highest Gini value is considered to be the best-performing model.

Table 5

Model Performance Measure — Gini metrics

Model Specification	train	test
Logit accounting	95.32%	96.52%
Logit Z-spread	96.43%	96.68%
Logit I-spread	96.47%	96.94%
Gradient Boosting accounting	97.14%	98.05%
Gradient Boosting Z-spread	97.32%	97.93%
Gradient Boosting I-spread	97.53%	98.44%

Source: compiled by the author.

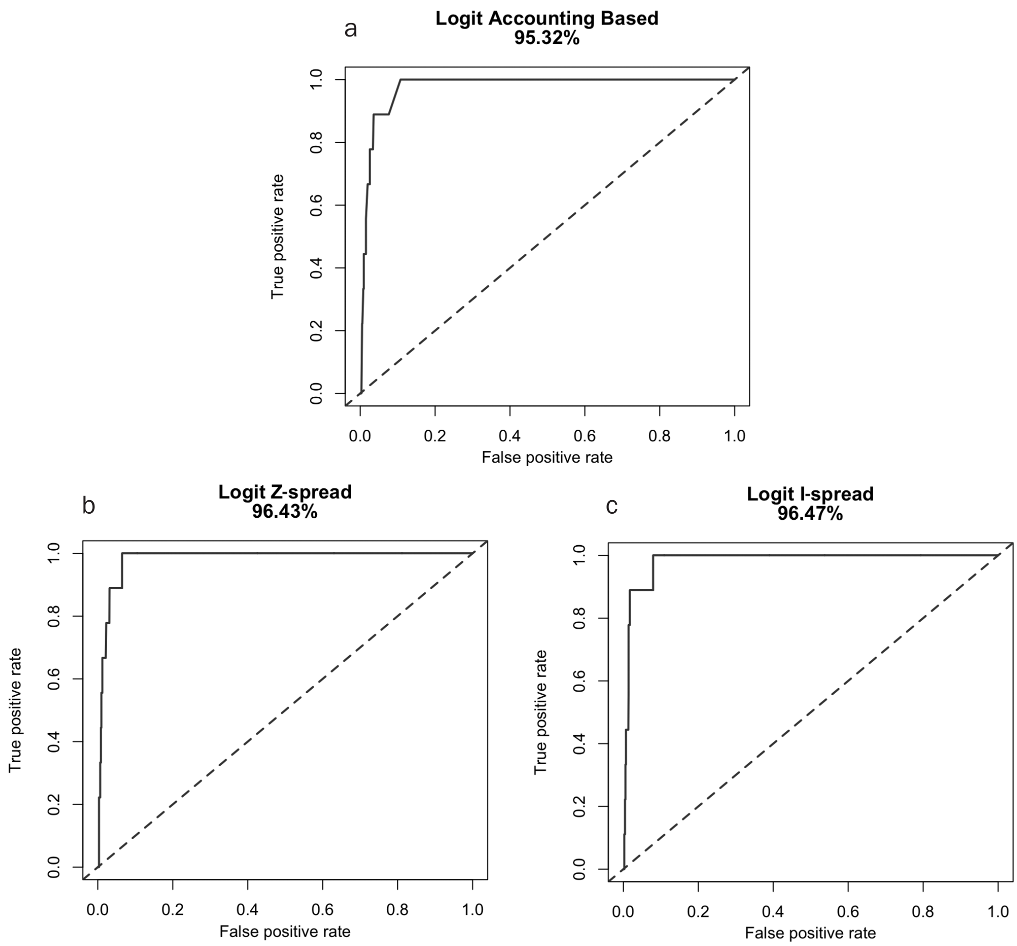
The models' performance on the training and test sets does not differ significantly, which means that our training model avoided overfitting. Comparing the two approaches, the logit model and GBM, the latter performs better in all model specifications. All models have quite high performance in predicting corporate failures. These results are not inconsistent with other studies (e.g. [Carmona et al., 2019]), where the AUC on the logistic regression and extreme gradient boosting test dataset is 0.84 and 0.98, respectively).

However, the main objective of our analysis is to identify the better performance of specific spreads in hybrid models. Hybrid models generally perform better than accounting-based models. Considering separately the Z-spread and I-spread, the latter performs better on the test set for both modelling techniques. However, the difference in the Gini metric between these two market measures is not significant. Thus, this study concludes that the choice between the Z-spread and I-spread is indifferent.

Rare events create challenges for classification models. When one outcome predicts very rare events, the opposite can result in a very high accuracy. ROC Curves for each of the models are also visualized. This provides a better understanding of the models' ability to distinguish between positive and negative predictions.

Figure 4

Logit ROC Curves and training sample Gini metrics: Accounting-Based (a), hybrid with Z-spread (b) and hybrid with I-spread (c)

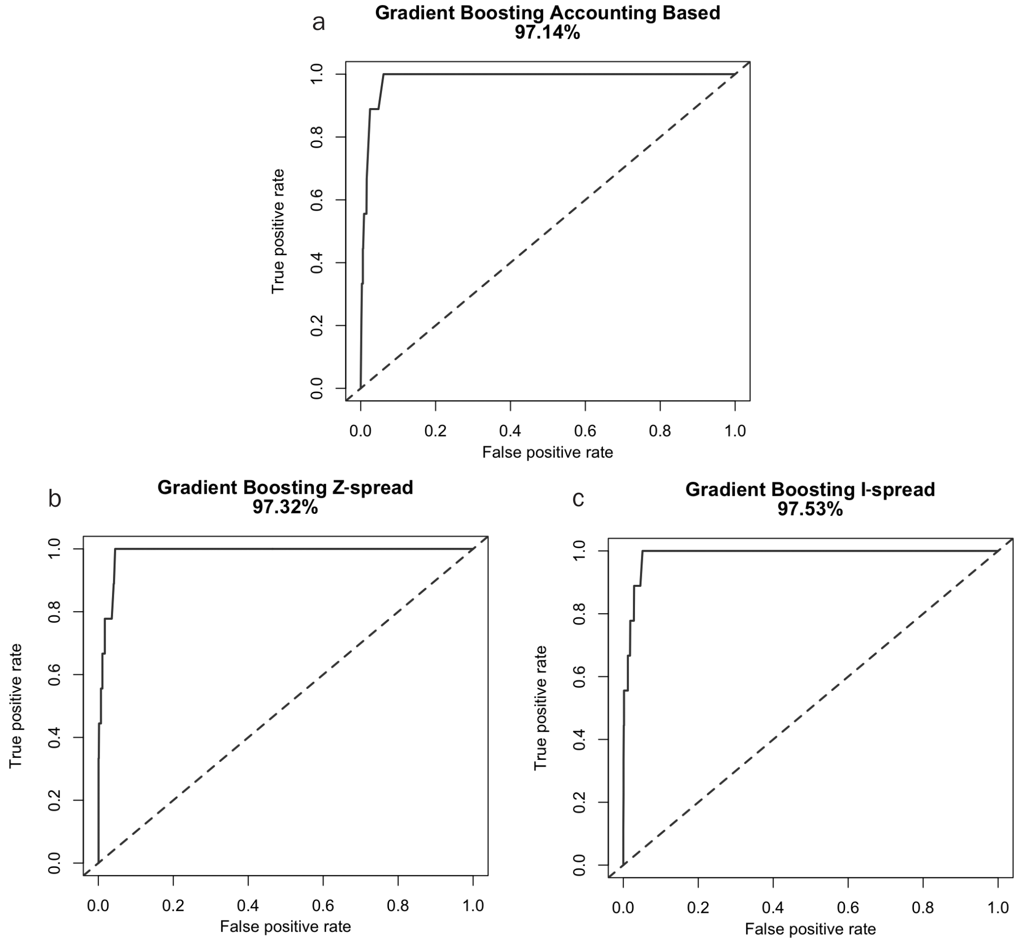


Source: compiled by the author.

The diagonal line in Fig. 4 and Fig. 5 is the baseline performance for a very poor model. The further the solid curve is from this dotted line, the better it performs. The nature of the curves of the two hybrid models is quite similar. Comparing the ROC curves of the hybrid models and the accounting-based model, we note that the latter is less “up and left”. Thus, it performs worse than other models according to the character of its ROC curves.

Figure 5

Gradient Boosting ROC Curves and training sample Gini metrics: Accounting-Based (a), hybrid with Z-spread (b) and hybrid with I-spread (c)



Source: compiled by the author.

Preserving the logic of comparing models by the ROC curves, the hybrid models do not differ from each other either in the character of the ROC curves or in the Gini metric. However, this study again underlines the lower performance of the accounting-based model, as was seen in the logistic specification.

Gini Bootstrap

According to Table 3, the I-spread performs better for logit model than the Z-spread. The difference in Gini metrics for the two market measures is not significant. This paper investigates the imbalanced sample because the corporate failure is considered to be a rare event. However, in such a bankruptcy prediction task, it is appropriate to use the sample as is because logistic regression covers these imbalances. In order to test the results and make sure that the difference between the spreads is not significant, a non-parametric approach is applied. To do this, a bootstrap procedure is performed for different subsamples.

This study performs 1000 iterations to calculate the Gini metric for three types of models. Wilcox suggests 599 iterations for general use. For a better illustration of the kernel density function, a larger number of iterations is chosen [Wilcox, 2010]. After computing the corresponding Gini values, kernel density estimates are determined and the Epanechnikov smoothing kernel function is applied. The Epanechnikov kernel is optimal in the sense of mean square error, although the efficiency loss is small for other kernels (i.e. uniform, triangular, biweight, triweight, normal, etc.) [Wand, 1995].

The bootstrapping procedure is performed as follows: companies with an outcome of 0 are randomly chosen in two proportions to a target value of 1 — 70:30 and 50:50 (Raschka, 2018). Table 6 illustrates descriptive statistics for the implementation of the first ratio for the training subsample.

Table 6

Descriptive Statistics for Gini Bootstrap

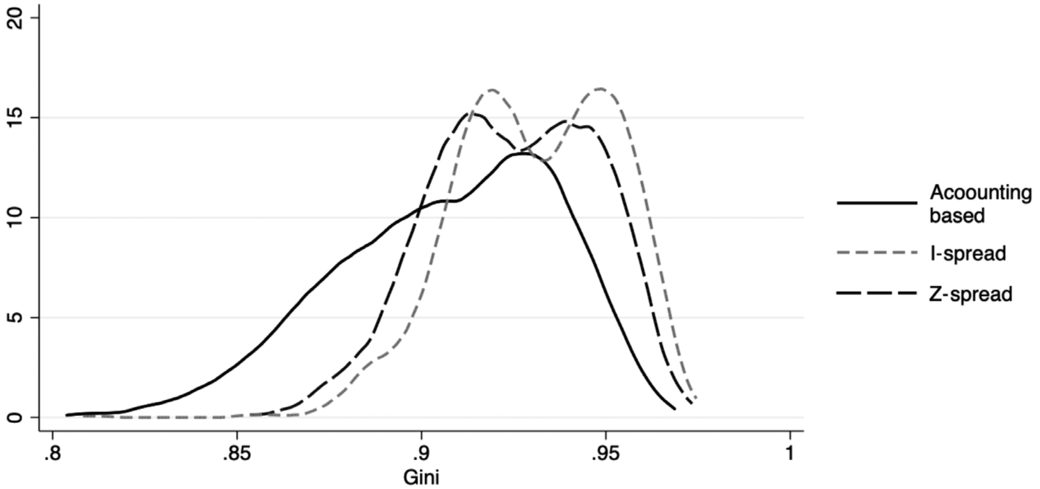
	Accounting-based	Z-spread	I-spread
Min	0.8888	0.9276	0.9295
Median	0.9270	0.9442	0.9445
Mean	0.9266	0.9442	0.9448
Max	0.9424	0.9602	0.9627

Source: compiled by the author.

Figure 6 shows that the density functions for the two types of spreads overlap in the training subsample with 15% of bankrupt firms' implementation for the Gini bootstrapping[1].

Figure 6

Epanechnikov kernel density functions for logit model with 15% of bankrupt firms



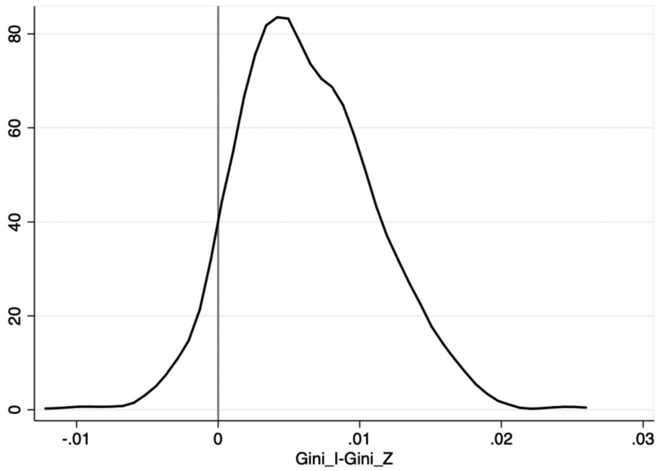
Source: compiled by the author.

[1] *15% ratio makes the bootstrap random with a good overall sample. Bootstrap with 30% bankrupt firms has robust qualitative conclusions.*

According to Figure 7, the mean of the spread difference is biased. Thus, it could be concluded that the I-spread performs better than the Z-spread in the logit model.

Figure 7

Epanechnikov kernel density function of spread difference for logit model with 15% of bankrupt firms

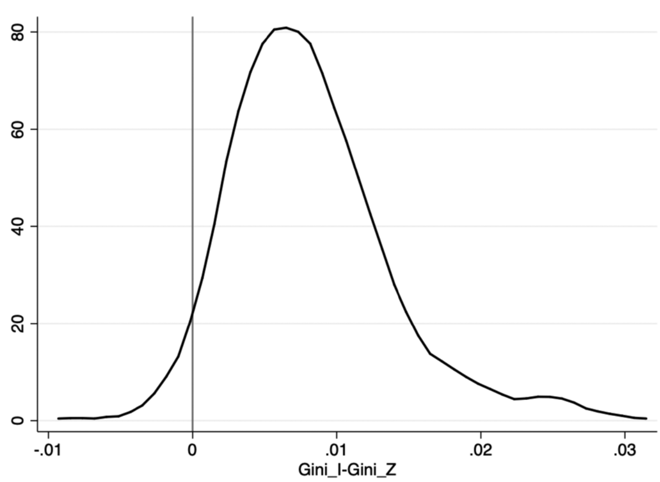


Source: compiled by the author.

Gini bootstrapping with a 15% share of bankrupt firms for the gradient boosting machine approach leads to the same results as for logit modelling: the I-spread outperforms the Z-spread (Figure 8), and hybrid models show better results in general.

Figure 8

Epanechnikov kernel density function of spread difference for GBM with 15% of bankrupt firms

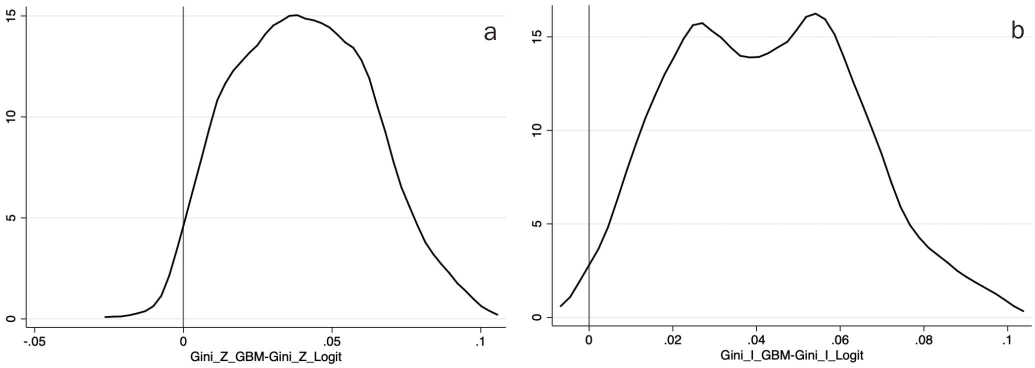


Source: compiled by the author.

Comparing the supremacy of the chosen model in specific spreads, GBM has greater predictive power than the logit model in both the Z-spread and the I-spread (Figure 9).

Figure 9

Epanechnikov kernel density function of model difference with 15% of bankrupt firms for hybrid with Z-spread (a) and hybrid with I-spread (b)



Source: compiled by the author.

As follows from this study, there is almost no difference in the implementation of the Z-spread and the I-spread. However, it can be seen that the I-spread performs better because the empirical distribution of the Gini metric is biased.

CONCLUSION

The main objective of this paper is to investigate the impact of the market measures implemented in an accounting-based model for predicting failure among listed U.S. corporate bond issuers. To do so, an empirical analysis is conducted using logistic regression and the gradient boosting machine in order to compare the predictive power of these techniques.

Upon the investigation, the main findings of the paper are as follows:

1. There is little difference in the effect of the Z-spread and the I-spread on the probability of default estimates. The I-spread performs better both in logistic regression and in the gradient boosting machine approach. This can be explained by the fact that different methods of yield curve interpolation are used to calculate the zero-coupon spread. Thus, the calculation accuracy of the Z-spread may be less precise than the simple difference between the corporate bond internal rate of return (IRR) and the interpolated yield to maturity of the treasury bond (I-spread).

2. The method of gradient boosting machine as the most recent approach to assessing default has an advantage in bankruptcy prediction accuracy compared to logistic modelling for all specifications, including the accounting-based approach and hybrid models.

To sum up, the high predictive power of the gradient boosting machine shown in this paper should encourage managers to favor modern techniques of corporate bankruptcy analysis over classical ones such as logistic regressions and multiple discriminant analysis models. Moreover, they should also take market dynamics into account and consider market measures as additional variables that help in predicting bankruptcy.

A possible extension of further research could be the consideration of various techniques of zero-coupon spread calculation. In this paper, the Z-spread calculated by Bloomberg is investigated. The study could also analyze which particular calculation method (e.g. yield to maturity curve interpolation) better explains estimates of the probability of default.

References

Alfaro E., García N., Gámez M., Elizondo D. (2008). Bankruptcy forecasting: an empirical comparison of AdaBooost and neural networks. *Decision Support Systems*, 45, 110–122.

Altman E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23 (4), 589–609.

Almeida H., Philippon T. (2007). The Risk-Adjusted Cost of Financial Distress. The *Journal of Finance*, 62 (6), 2557–2586.

Anginer D., Yildizhan C. (2010). Is There a Distress Risk Anomaly? Corporate Bond Spread as a Proxy for Default Risk, Working Paper, World Bank and University of Georgia.

Beaver W.H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111.

Carmona P., Climent F., Momparler A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics and Finance, Elsevier*, 61, 304–323.

Dias P. et al. (2018). A comparison of Gradient Boosting with Logistic Regression in Practical Cases. Available at: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum proceedings/2018/1857-2018.pdf.

Fawcett T. (2003). ROC graphs: Notes and practical considerations for data mining researchers. Intelligent Enterprise Technologies Laboratory, HP Laboratories, USA.

Fabozzi F.J. (2007). Bond markets, analysis and strategies. Upper Saddle River, N.J., Pearson Prentice Hall.

Friedman J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29 (5), 1189–1232.

Good I.J. (1985). Weight of Evidence: A Brief Survey. *Bayesian Statistics*, 2, 249–270.

Gorunescu F. (2011). Data mining concepts, models and techniques. Springer, Berlin.

Hajian-Tilaki K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Int Med*, 4, 627–635.

Huang J., Huang M. (2003). How Much of the Corporate-Treasury Yield-Spread is due to Credit Risk? Working paper, Stanford University.

Kim M.J. and Kang D.K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications*, 37, 4, 3373–3379.

Zięba M., Tomczak S.K. et al. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58, 93–101.

Makeeva E., Bakurova A. (2012). Forecasting Bankruptcy Oil and Gas Companies Using Neural Networks. *Journal of Corporate Finance Research*, 23 (3), 22–30.

Makeeva E., Neretina E. (2013). The prediction of bankruptcy in a construction industry of Russian Federation. *Journal of Modern Accounting and Auditing*, 9 (2).

Mitchell M.W. (2011). Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. *Open Journal of Statistics*, 1 (3), 205–211.

Ohlson J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18 (1), 109–131.

Raschka S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. https://arxiv.org/abs/1811.12808.

Siddiqi N. (2006). Credit Risk Scorecards. Developing and Implementing Intelligent Credit Scoring. John Wiley & Sons, 210 p.

Tam K.Y., Kiang M.Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, 38 (7), 926–947.

Tinoco M.H., Wilson N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30, 394–419.

Touzani S., Granderson J. et al. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build*, 158, 1533–1543.

Trujillo-Ponce A., Samaniego Medina R., Cardone-Riportella C. (2014). Examining what best explains corporate credit risk: Accounting-based versus market-based models. *Journal of Business Economics and Management*, 15, 2, 253–176.

Vatcheva K.P., Lee M., McCormick J.B., Rahbar M.H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*, 6, 2, 227.

Wand M.P., Jones M.C. (1995). Kernel Smoothing. London: Chapman & Hall/CRC. 224 p.

Wei L., Li J. et al. (2007). Credit risk evaluation using support vector machine with mixture of kernel. International Conference on Computational Science. Springer Berlin Heidelberg, 431–438.

Wilcox R.R. (2001). Fundamentals of Modern Statistical Methods. Springer.

Wilson R.L., Sharda R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11 (5), 545–557.

Zeng G. (2013). Metric Divergence Measures and Information Values in Credit Scoring. *Journal of Mathematics*, art. 848271. https://doi.org/10.1155/2013/848271.

Zmijewski M.E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 22, 59–82.

Information about the author

***Daria S. Leonteva**, Postgraduate student, HSE University, Moscow*

Article submitted April 12, 2022
Approved after reviewing October 11, 2022
Accepted for publication November 30, 2022

https://doi.org/10.31107/2075-1990-2022-6-74-90

Использование рыночных показателей для уточнения оценок вероятностей банкротств корпораций

***Дарья Сергеевна Леонтьева**, аспирант Научно-исследовательского университета «Высшая школа экономики», Москва 101000, Российская Федерация*
E-mail: dsleonteva@edu.hse.ru, ORCID: 0000-0002-4571-7676

Аннотация
Проблема прогнозирования дефолта корпораций является актуальной не только для кредиторов и инвесторов, но и для органов макроэкономического управления. Однако для точности составления прогнозов дефолта требуется постоянное обновление информации о вероятности риска дефолта корпораций. В данном исследовании рассматривается альтернативный подход к оценке вероятности банкротства компаний. Классический подход к прогнозированию банкротства основан на анализе бухгалтерской отчетности. Гибридная модель предполагает внедрение рыночных мер в подход, основанный на балансовых данных, и подразумевает бóльшую предсказательную силу.
В этой статье определяется, какой конкретный спред среди двух рыночных показателей — Z-спред или I-спред — имеет преимущество при прогнозировании банкротства. С использованием двух методов — логистической регрессии и метода градиентного бустинга, а также панельных данных 80 финансовых коэффициентов для 385 зарегистрированных на бирже американских компаний, выпускающих корпоративные облигации, обнаружено свидетельство того, что I-спред имеет более высокую прогнозную силу в обоих методах. Преимущество внедрения I-спреда может быть объяснено тем, что точность расчета Z-спреда может вводить в заблуждение, поскольку используются разные методы интерполяции кривой доходности. Кроме того, сравнивается предсказательная сила выбранных методов. Современный метод машинного обучения — градиентный бустинг — превосходит классическую логистическую модель в предсказательной силе. Данные результаты могут побудить менеджеров использовать в анализе дополнительные рыночные показатели и применять современные методы моделирования с элементами машинного обучения вместо классических (логистические регрессии и модели множественного дискриминантного анализа) для прогнозирования несостоятельности корпораций.

***Ключевые слова:** прогнозирование банкротства, кредитные спреды, логистическая регрессия, градиентный бустинг*

***JEL:** C53, G33*

***Примечание:** Статья основывается на положениях, выносимых на защиту магистерской диссертации в НИУ ВШЭ. Исследование обсуждалось на XXII Апрельской международной научной конференции по экономическому и социальному развитию, сессия L-19-2. Корпоративное управление и финансы, 13–30 апреля 2021 г.*

***Для цитирования:** Leonteva D.S. (2022). Using Market Indicators to Refine Estimates of Corporate Bankruptcy Probabilities. Financial Journal, vol. 14, no. 6, pp. 74–90.*
https://doi.org/10.31107/2075-1990-2022-6-74-90.

Статья поступила в редакцию 12.04.2022
Одобрена после рецензирования 11.10.2022
Принята к публикации 30.11.2022