Default Prediction for Housing and Utilities Management Firms Using Non-Financial Data

Vladislav V. Afanasev[1], **Yulia A. Tarasova**[2]

[1,2] HSE University, St. Petersburg 190008, Russian Federation
[1] vvafanasev@hse.ru, https://orcid.org/0000-0002-4041-4465
[2] yutarasova@hse.ru, https://orcid.org/0000-0001-9341-3151

Abstract

For many years, financial ratios have been used as predictors of default. However, biases in financial statements of companies in Russia call into question the applicability of this approach. An alternative approach is to use non-financial data in such models.

The purpose of this paper is to find out whether non-financial data, such as information related to court trials, unscheduled inspections and firm age, can significantly improve the accuracy of default prediction in the housing and utilities management industry.

This part of the services sector is chosen as one of the riskiest industries, in which firm default affects not only conventional stakeholders such as banks, shareholders, employees, etc, but also customers.

A dataset of 378 housing and utilities management firms which have faced default and 765 solvent "healthy peers" is used to create and test default prediction models. Logistic regression is used as the classification algorithm.

The results suggest that addition of non-financial data can significantly improve the accuracy of default prediction, and moreover, non-financial data can be used exclusively without any financial ratios to create classification models which show acceptable accuracy.

The paper contributes to the existing literature by providing new evidence on the benefits of using non-financial data in default prediction models. In addition, we were able to collect a unique dataset of unscheduled inspections and use this data for default prediction, which appears to be the first case of this kind.

Keywords: default prediction, credit risk assessment, housing and utilities management firms, non-financial data

JEL: G33, G21

INTRODUCTION

Housing & Utility Management Firms (hereinafter HUMFs), which are responsible for providing housing and utility services and resources to real estate owners, have a reputation of being risky. In 2021, there were registered 1,199 defaults of firms, specializing on operations with real estate (a category, in which HUMFs hold a significant share), accounting for 12% of the total number of bankruptcies in the economy[1]. Defaults by such firms affect not only conventional stakeholders such as counterparties, employees, etc., but also customers (estate property owners) who may face disruption of resources and services. In this regard, it seems important to accurately assess the risk of default of such firms. In other words, to accurately predict defaults.

Default prediction is usually carried out using financial ratios as predictors. Since the 1960s many researchers have shown that financial data can be effectively used to identify risky firms in terms of credit risk, starting with the work of Edward Altman [Altman, 1968] and ending with modern studies, including works of Russian authors [Grigoriev et al., 2019; Jaki et al., 2021; Mai et al., 2019; Makeeva et al., 2020].

Since the first works in this field, default prediction models have evolved both in terms of methodology and in terms of the set of predictors. As for the basic methodology, 60 years ago a simple classification algorithm was used — Multiple Discriminant Analysis [Altman, 1968], later such algorithms as logistic and probit regression [Hunter et al., 2001; Ohlson, 1980] have become popular and are still often used for such research [Kovacova et al., 2017; Sirirattanaphonkun et al., 2012]. In recent times, more powerful (in many cases nonlinear) Machine Learning classification algorithms have come to the fore [Altman et al., 1994; Cao et al., 2020; Coats and Fant, 1993; Mselmi et al., 2017; Odom and Sharda, 1990; Kumar and Ravi, 2007; Zhang et al., 1999], which increases the reported accuracy of default prediction.

As for the variables used to predict defaults, there also have been changes in this area. While Altman [Altman, 1968] used basic static financial ratios, more recent studies add new predictors, e.g. dynamic variables such as income growth rate [Cao et al., 2020] and stock risk measures such as standard deviation of stock returns [Mselmi et al., 2017].

The point which does not change over time is that when it comes to default prediction, financial ratios, usually calculated several periods before default, are always used as basic predictors in such models. However, it seems that in case of a developing economy such as the Russian Federation, financial ratios of legal entities may be biased for several reasons, such as the high level of business disaggregation for tax optimization [Kachalin, 2011] or off-the-books entrepreneurship [Williams et al., 2013]. This statement calls into question the possibility of using exclusively financial ratios to accurately predict default for Russian firms.

One possible way to achieve better accuracy of default prediction is to use non-financial predictors, which can act as proxies for real financial ratios. The evidence that non-financial predictors improve prediction accuracy can be found in prior studies. These non-financial variables can be of any nature: corporate governance measures [Xie et al., 2011], age [Altman et al., 2010], lawsuit-related variables (in terms of the number or value of such lawsuits) [Shumway, 2001], corporate social responsibility indicators [Boubaker et al., 2020], mood level of text used in news or disclosures [Mai et al., 2019; Makeeva and Sinilshchikova, 2020], measures based on audit reports (sentiment, number of comments, etc.) [Blanco et al., 2015], etc.

In this study, we focused on HUMFs and tested whether the use of non-financial data can improve the quality of default prediction. The research questions are as follows:

---

[1] *Fedresurs. URL: https://download.fedresurs.ru/news/Банкротство%20статрелиз%202021.pdf.*

RQ1: How much could prediction accuracy increase if non-financial data were used along with financial ratios in classification models?

RQ2: What level of prediction accuracy can be achieved by using only non-financial data?

The rest of the paper is organized as follows: In the next section we provide a review of the literature related to default prediction, then we discuss the specifics of HUMFs, then describe the data and how they were analyzed, and finally, we present and discuss the results.

LITERATURE REVIEW

Conventional approach to default prediction

The conventional approach to default prediction implies the use of financial ratios as explanatory variables. Starting from the first model of default risk assessment developed by William Beaver in 1966 [Beaver, 1966], proceeding with the works of Edward Altman [Altman, 1968] and James Ohlson [Ohlson, 1980] who are considered the "fathers" of default prediction, and ending with the recent works of foreign [Kovacova and Kliestik, 2017; Mselmi et al., 2017] and Russian researchers [Grigoriev and Tarasov, 2019], one can find a sufficient number of papers devoted to default prediction using financial data. Usually, financial ratios covering profitability, liquidity, capital structure and turnover, calculated a year or several years prior to default, are used together as predictors of default. A firm's poorer financial condition is considered an indicator of future potential non-payment.

However, financial ratios may be poor default predictors in the case of the Russian economy for at least two reasons. Firstly, the financial ratios of a legal entity may not reflect the condition of the whole business, because it is a common situation when a business is disaggregated. As stated by Kachalin [2011], business disaggregation is a way to optimize tax payments. A business represented by several small legal entities can pay less taxes under simplified taxation regimes. Secondly, there is a large share of shadow operations in Russian business, which may also make financial ratios biased. For example, according to the Russian Longitudinal Monitoring Survey (HSE, 2020). 16% of Russian citizens are paid off the books (and half of them get entire salary "in an envelope"). It is important to find an accurate approach to default prediction as an alternative to the conventional one, in order to account for biases in reporting. One possible solution is to add non-financial data as predictors of default.

Default prediction using non-financial data

In an attempt to increase the accuracy of default prediction, many researchers in recent years have begun to add non-financial data to models, forming a new scientific direction. According to Edward Altman [Altman et al., 2010], this area, for SMEs in particular, was not explored at all until 2010. A. Blanco calls the use of non-financial data a "novel trend in this field" [Blanco et al., 2015].

There is a wide range of non-financial variables that can be used in default prediction and the choice of such variables is limited only by common sense. The use of non-financial data to predict defaults is not a widespread approach, but the findings of those researchers who have attempted to explore this area are promising — a significant improvement in prediction accuracy has been reported, e.g. an 8% increase in the area under the ROC curve [Altman et al., 2010]. Table 1 presents the results received by several previous researchers when adding non-financial variables to default prediction models. To form this table, we selected the most cited works which can be found in the Web of Science database using keywords related to the use of non-financial data, such as "non-financial variables" together with "default prediction".

However, the number of studies devoted to non-financial variables as defaults predictors (particularly for the Russian services sector) is still limited, and we were unable to find any studies related to HUMFs.

Selected results of adding non-financial data in default prediction models

Research paper	Quality measure	Quality of model with financial vars only	Quality of model with non-financial data added	Quality growth
[Altman et al., 2010]	AUC ROC	0.74	0.80	8%
[Grunert et al., 2005]	Overall accuracy	88%	91%	3%
[Bandyopadhyay, 2006]	AUC ROC	0.94	0.97	3%
[Xie et al., 2011]	Overall accuracy	78%	83%	6%
[Lugovskaya, 2010]	Overall accuracy	68%	79%	16%
[Laitinen, 2011]	Overall accuracy	74%	87%	18%
[Wilson et al., 2016]	AUC ROC	72%	77%	7%
[Bhimani et al., 2013]	AUC ROC	71%	86%	21%
[Lin et al., 2010]	Overall accuracy	89%	94%	6%

Source: compiled by the authors.

Approaches to defining default

Definitions of default vary from one study to another. The main reason for the differences is that firms do not go bankrupt instantly. Usually, the process of "failure" stretches over time, starting with non-payment and ending with official bankruptcy. Also and it is "extremely important to distinguish between failure and closure" [Altman et al., 2010], because firms can close for reasons unrelated to insolvency. This is why the definition of default is not obvious and needs to be specified.

In some studies the date of default is considered as the beginning of the legal procedure of insolvency [Muñoz-Izquierdo et al., 2020]. A similar approach is to treat firms that have entered into liquidation, administration or receivership procedures as defaulted firms [Altman et al., 2010]. The delisting date can be considered as the date of default for publicly traded firms [Mai et al., 2019]. However, there is a view that default (financial distress, failure) can be identified before the actual non-payment, e.g. when a firm's EBITDA becomes less than the interest obligations [Andrade and Kaplan, 1998].

In this paper we considered a firm to be in default if the following two conditions were met: a creditor's notice of intent to request the court for bankruptcy was filed AND the insolvency proceedings began. We use the date of the creditor's notice of intent as the date of default, because we are interested in the date closest to the actual non-payment date rather than the official start of proceedings, and there is a lag of 9–10 months on average between these two events (see Figure 1).

**Histogram for the distribution of time
between a creditor's notice of intent and the start of insolvency proceedings, months**
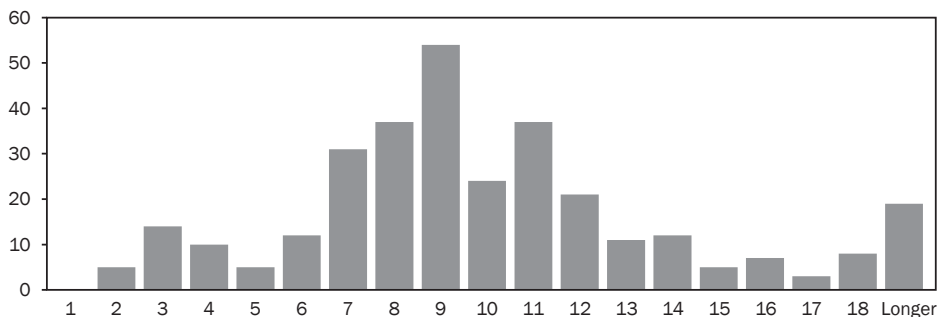
Source: compiled by the authors.

Modelling techniques used for default prediction

All default prediction models can be divided into two large groups: "statistical" and "intelligent" [Kumar and Ravi, 2007]. "Statistical" models are those developed using statistical (and mostly linear) methods, such as multiple discriminant analysis [Altman, 1968], logistic regression [Gruszczyński, 2004; Hunter and Isachenkova, 2001; Kovacova and Kliestik, 2017; Ohlson, 1980; Sirirattanaphonkun and Pattarathammas, 2012], linear discriminant analysis [Lugovskaya, 2010]. "Intelligent" models are those which rely on machine learning algorithms [Ahmadpour Kasgari et al., 2013; Grigoriev and Tarasov, 2019; Mai et al., 2019; Odom and Sharda, 1990].

Machine learning algorithms appear to be more accurate in default prediction compared to "statistical" methods, as shown by Ahmad Ahmadpour Kasgari et al. [2013] or Flavio Barboza et al. [2017]. However, "statistical" methods have the advantage of analyzing the contribution of every variable to the classification result, compared to machine learning algorithms' black-box-like working schemes.

In this paper we used logistic regression, a "statistical" method, because it is crucial to assess the contribution of financial and non-financial variables both in terms of the strength of the contribution to the classification and the direction of this contribution.

SPECIFICS OF HOUSING AND UTILITIES MANAGEMENT FIRMS

HUMFs are firms which provide resources (such as gas, water, energy) and services (e.g. maintenance, cleaning) to the real estate residents. Basically, these firms provide services themselves, but in case of resources they play the role of intermediaries between suppliers and residents. There are almost 50 thousand HUMFs in Russia[2], and the volume of the market for such services is estimated at almost RUB 3000 bn[3]. Thus, it is a huge market which forms about 2% of Russia's GDP.

Basically, HUMFs are intermediaries between suppliers of resources (water, gas, electricity etc.) and residents. Thus, the poor condition of HUMFs creates risks for both residents and suppliers. This is one of the reasons why, since 2018, property owners in Russia have been allowed to enter into direct contracts with suppliers[4].

It was expected that most real estate owners would enter into such agreements. However, there are several limitations to this. First, it is easier for residents to have "one-window" communication with a HUMF, rather than communicate with several suppliers. Second, if any problems with resources occur, it is easier for residents to resolve them through HUMFs, because they have the resources, including the ability to engage professional lawyers, etc. Finally, there may be some problems with shared utility systems, because in the case of direct contracts HUMFs and suppliers tend to shift responsibilities[5]. Thus, HUMFs are still important to residents and defaults by such firms are bad news for customers.

To understand the reasons for defaults and potential explanatory factors, we describe HUMFs from the perspective of three forms of business activity.

HUMFs' operating activity

The first part of the operating activity is how the business earns revenue. The main source of income for HUMFs is payments from residents and renters. The amount of these payments

---

[2] *Reforma GKH. URL: https://www.reformagkh.ru/opendata?gid=2208161&cids=house_management&page=1 &pageSize=10.*

[3] *Умное ЖКХ. URL: https://умное-жкх.рф/article/konsolidatciya-rynka-zhkh-20.*

[4] *DVHUB.RU. URL: https://www.dvnovosti.ru/khab/2018/03/26/80643/.*

[5] *MK.RU. URL: https://www.mk.ru/economics/2018/10/07/pochemu-pryamye-platezhi-za-zhkkh-mogu-vyyti-zhilcam-bokom.html.*

are fixed in the agreements and include fees for the provision of resources, fees for mainte-nance, fees for current and capital repairs. Most payments are fixed monthly, so it appears that HUMFs have recurring and highly predictable revenues. However, as noted by Kvitova and Yavorskaya [Kvitova et al., 2018], payment collection turns to be tough for HUMFs, and this seems to be one of the main reasons for the high number of defaults.

This statement is supported by statistics. The collection period for such firms is an average of 73 days. This means that it takes an average of 3.5 months to collect debts from residents. For comparison, the average turnover of receivables for Russian firms is about 63 days[6].

The major part of HUMFs' operating costs are payments to suppliers. As long as a HUMF is unable to collect money from clients, it will have a large amount of payables, and will be likely to face default. At the same time, suppliers tend to press charges quickly, and this leads to additional costs for a HUMF. These are litigation costs, state fees, etc, and additional interest payments. HUMFs try to improve the process of payment collection, but this again requires additional expenditures, e.g. on professional debt collection services [Kvitova and Yavorskaya, 2018].

Thus, the major problem in HUMFs' operating activity is residents' non-payments, which itself can lead to default, but also imposes additional costs on firms.

HUMFs' investing activity

Basically, HUMFs do not invest much in any fixed assets (e.g. expensive equipment). Thus, if a HUMF faces default and then goes bankrupt, the amount of assets which can be disposed (sold to cover creditors' debts) is very small. The major asset that can be disposed is accounts receivable formed from residents' debts [Kovalenko, 2019]. This fact seems to be bad news for creditors.

HUMFs' financing activity

Usually, the initial shareholder capital is set at RUB 10 ths, and this leads to two major problems [Sukharev et al., 2018]. First, such a small equity capital makes it hard to get debt financing, because the small initial invested capital means a small amount of assets that can be purchased and used as collateral. Second, as the owners in such cases have low responsibility, they do not particularly care about financial stability and high-quality service delivery.

To sum up, HUMFs face problems in all types of activities which either can lead to default (resident non-payments and additional costs), make it difficult to attract financing under stress (difficulties with debt attraction), or make it hard for creditors to receive anything in case of bankruptcy (small amount of assets to be disposed).

The impact of HUMFs' defaults on customers

When a HUMF runs into default and then goes bankrupt, the first affected party (other than creditors) is the clients. HUMFs under insolvency proceedings, which can last several months, tend to stop providing services.

Besides, given that HUMFs collect current repairs fees, which are then stored on their accounts, residents may actually lose money because these funds may be added to the insolvency estate and transferred to creditors. Despite being one of the main stakeholders, residents are unfortunately not considered to be first-priority creditors.

---

[6] *SPARK Interfax.*

Finally, suppliers tend to stop providing resources (electricity, water) when a HUMF faces financial difficulties[7].

Customers are definitely not the only ones facing the consequences of HUMFs' defaults. Counterparties, employees, management and shareholders are also unhappy with the firms' financial distress. However, the effect on clients distinguishes HUMFs from firms in other industries, which makes accurate default prediction for such firms even more important.

METHODS AND DATA

Methods

Default prediction models are classification models which help to classify firms into defaulted and non-defaulted. Following Ohlson [Ohlson, 1980] and many contemporary researchers e.g. [Kovacova and Kliestik, 2017; Sirirattanaphonkun and Pattarathammas, 2012], we used logistic regression as a statistical tool to create classification models. While ordinary linear regression works with any kind of dependent variable, logistic regression is designed specifically to predict binary variables (in this case 1 — default occurred, 0 — default did not occur). The expression underlying the logistic regression is as follows:

$$P(x) = \frac{e^{\wedge}B0+B1X1+\cdots+BnXn}{1+(e^{\wedge}B0+B1X1+\cdots+BnXn)}$$

P(X) — estimated probability of default, B0–Bn — regression coefficients. The regression is fitted by maximizing the logarithm of the maximum likelihood function:

$$LF = \prod \left(P(x_i)^{y_i} \cdot (1 - P(x_i)^{1-y_i}) \right), i \in (1; n)$$

In other words, what is maximized is the product of the estimated default probabilities for defaulted firms multiplied by the product of the estimated (1- probability of default)s for non-defaulted firms.

Data

The data were collected from the SPARK Interfax database and consist of 378 HUMFs which faced default between 2017 and 2021 and 756 firms which are successfully operating nowadays (matched to defaults using the value of total assets), following the approach presented in [Sirirattanaphonkun and Pattarathammas, 2012]. The 90 firms facing default in 2021 and 180 "healthy" peers were used as a test dataset to assess the quality of the classification models.

The pool of financial independent variables consists of ratios covering firms' business activity, liquidity, profitability, and capital structure. The list of financial variables is given in Table 2.

Table 2

Financial independent variables

Variable name	Description
WCTA	Working capital / Total assets
CLR	Current liquidity ratio (Current assets / Current liabilities)
QLR	Quick liquidity ratio (Cash + Receivables / Current liabilities)
ROA	Return on assets (Net income / Total Assets)

---

[7] *Народный контроль в сфере ЖКХ. URL: https://nkgkh.ru/novosti/raz-yasneniya/804-komu-dolzhen-vsem-proshchayu-chto-budet-s-dengami-na-litsevykh-schetakh-domov-posle-bankrotstva-uk.*

Variable name	Description
ROE	Return on equity (Net income / Equity)
ROS	Return on sales (Net income / Revenue)
ART	Receivables collection period in days
APT	Payables credit period in days
AT	Assets turnover
ARTA	Receivables / Total assets
TLE	Total liabilities / Equity
GROS	ROS growth in recent 2 years (%)
GREV	Revenue growth in recent 2 years (%)

Source: compiled by the authors.

The financial ratios used as independent variables are calculated for the year which precedes the year of default. The year of default is estimated using the date of the "creditor's message of intent to request the court for bankruptcy".

Summary statistics for financial independent variables are presented in Table 3.

Table 3

Summary statistics for financial independent variables

Summary statistics for non-defaulted firms

	N	Mean	Median	SD	Min	Max
WCTA	722	0.2	0.2	1.2	−22.9	1.0
CLR	756	7.6	1.3	54.6	0.01	1,325.9
QLR	756	7.4	1.2	54.5	0.01	1,325.9
ROA	732	10%	3%	57%	−527%	1,206%
ROE	704	250%	18%	2,802%	−2,486%	60,610%
ROS	734	−137%	2%	2,095%	−42,750%	187%
ART	735	307.7	121.7	624.6	4.2	3,650.0
APT	756	57,166.3	90.7	937,097.5	0.4	23,267,108.0
AT	756	22,848.0	207.5	547,960.3	13.0	15,059,393.0
ARTA	754	0.7	0.8	0.3	0.0	1.0
TLE	716	147.0	1.2	2,848.3	−645.5	75,789.0
GROS	681	−87%	−37%	1,780%	−28,585%	12,000%
GREV	708	1,944%	4%	34,378%	−98%	785,106%

Summary statistics for defaulted firms

	N	Mean	Median	SD	Min	Max
WCTA	371	−0.7	−0.1	3.5	−60.1	1.0
CLR	377	1.1	0.9	2.3	0.02	39.0
QLR	376	1.0	0.9	2.3	0.01	39.0
ROA	371	−14%	−4%	55%	−389%	531%
ROE	361	−13%	15%	3,006%	−25,377%	34,318%
ROS	370	−66%	−3%	874%	−16,680%	369%
ART	362	503.7	228.1	791.3	1.5	3,650.0
APT	355	1,273.6	269.6	9,713.7	1.0	180,144.6
AT	376	2,111.8	288.7	11,653.2	0.9	173,809.5
ARTA	374	0.9	1.0	0.2	0.0	1.0
TLE	368	120.6	−1.4	749.1	−1,876.6	8,157.7
GROS	362	−181%	−56%	13,859%	−81,700%	164,900%
GREV	378	83,982%	−10%	1,631,320%	−100%	31,716,567%

Source: compiled by the authors.

Non-financial independent variables

To test whether the use of non-financial variables can help to improve the accuracy of prediction, we collected information on the age of the firm in the year before default, data on almost 11 thousand arbitration proceedings, in which firms participated in the 2 years before the year of default (or the same two years for "healthy peers") from the SPARK Interfax database. We also managed to obtain a unique dataset of more than 100 thousand inspections which defaulted firms faced in the 2 years before the year of default (or the same two years for "healthy peers") using web-scrapping to get data from the Russian Business Center website (https://vbankcenter.ru). A list of non-financial variables with reasons for inclusion is given in Table 4.

Table 4

Non-financial independent variables

Variable name	Description	Hypothesis / Reason to include
Age	Age of the firm in years	Younger firms are riskier due to lower value of assets and poorer networking
N trials I2	Number of arbitrage court trials in the last 2 years before the default	Firms which face more pressure from counterparties are riskier due to higher payables value and higher court costs
Sum trials I2 to TA	Sum of claims in these trials divided by total assets	
N inspections I2	Number of unscheduled inspections in the last 2 years before the default	One of the reasons a firm faces more inspections and violates more is the lack of resources to maintain effective services delivery, so the inspections-related variables can be proxies for financial stability measures
N viol I2	Number of unscheduled inspections with violations identified	
Sh viol I2	Share of unscheduled inspections with violations identified	

Source: compiled by the authors.

The probability of default is expected to be lower for relatively old firms (all other things being equal), because such firms seem to have both more market experience and assets. Age is widely used in default prediction models and turns to be a significant predictor e.g. [Altman et al., 2010].

As discussed earlier, when a HUMF stops paying resource providers, it immediately leads to litigation costs. The more lawsuits a legal entity has and the higher the cumulative cost of these lawsuits, the higher the probability of default. This factor works both as a proxy for the firm's unpaid debts and as an indicator of high litigation costs.

Variables related to unscheduled inspections and identified violations seem to be good proxies for the financial condition of the business, because unscheduled inspections are often caused by residents' complaints. And these complaints seem to be caused by inability to provide quality services, including due to poor financial condition.

Summary statistics for the non-financial independent variables are presented in Table 5. It appears that defaulted HUMFs tend to have more court trials in the 2 years prior to default, and the average and median size of the lawsuit is significantly higher. In addition, defaulted firms tend to undergo more unscheduled inspections, and these inspections more often lead to the discovery of violations.

Table 5

Summary statistics for non-financial independent variables

Summary statistics for non-defaulted firms

	N	Mean	Median	SD	Min	Max
Age	756	10	9	7	0	57
N trials I2	756	4	1	7	0	85

	N	Mean	Median	SD	Min	Max
Sum trials I2	756	9,170,610	43,130	51,116,446	0	1,097,000,000
Sum trials I2 to TA	756	0.2	0.002	1.5	0	40.1
N inspections I2	756	27	6	73	0	1,060
N viol I2	756	7	1	18	0	306
Sh viol I2	756	21%	14%	24%	0%	100%

Summary statistics for defaulted firms

	N	Mean	Median	SD	Min	Max
Age	378	9	8	6	1	57
N trials I2	378	21	14	27	0	367
Sum trials I2	378	85,638,244	18,045,436	244,900,000	0	2,006,000,000
Sum trials I2 to TA	378	2	1	11	0	143
N inspections I2	378	68	17	203	0	2,743
N viol I2	378	21	4	77	0	1,026
Sh viol I2	378	28%	25%	24%	0%	100%

Source: compiled by the authors.

RESULTS

Default prediction using the conventional approach (financial variables only)

First, we ran a logistic regression on the training dataset using only financial variables. To choose the variables to be included in the model, we performed a Mann-Whitney rank sum test, which shows whether the differences in the mean values of the variables for defaulted and non-defaulted firms are statistically significant. The results of the test are shown in Table 6. We chose a significance level of 5%, which means that if the probability of getting such test statistic — given that the mean values are the same (null hypothesis) — is less than 5%, then we state that the mean values are statistically different.

Table 6

Mann-Whitney rank sum test results for financial variables

	N	Mann-Whitney test p-value	Conclusion
WCTA	829	0.000	Include
CLR	863	0.000	Include
QLR	863	0.000	Include
ROA	840	0.000	Include
ROE	817	0.620	Do not include
ROS	841	0.000	Include
ART	837	0.000	Include
APT	844	0.000	Include
AT	862	0.001	Include
ARTA	860	0.000	Include
TLE	831	0.000	Include
GROS	789	0.572	Do not include
GREV	823	0.000	Include

Source: compiled by the authors.

The next filtering step was to exclude variables that are closely correlated with each other in order to avoid multicollinearity. The quick liquidity ratio was excluded because of its perfect correlation with current liquidity ratio.

<div align="right"><i>Table 7</i></div>

Correlation table for financial variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) D	1.000											
(2) WCTA	−0.183	1.000										
(3) CLR	−0.069	0.044	1.000									
(4) QLR	−0.067	0.043	1.000	1.000								
(5) ROA	−0.192	0.122	0.017	0.017	1.000							
(6) ROS	0.019	0.171	0.005	0.005	0.317	1.000						
(7) ART	0.134	−0.006	0.009	0.010	−0.049	−0.196	1.000					
(8) APT	−0.034	0.002	−0.005	−0.005	−0.002	0.003	0.030	1.000				
(9) AT	−0.022	0.001	−0.003	−0.003	−0.048	−0.006	0.193	0.092	1.000			
(10) ARTA	0.297	0.014	−0.049	−0.046	−0.016	0.031	0.110	−0.033	−0.074	1.000		
(11) TLE	−0.005	0.002	−0.006	−0.005	−0.005	0.003	−0.002	−0.003	−0.002	−0.036	1.000	
(12) GREV	0.041	−0.013	−0.003	−0.003	−0.118	0.000	−0.013	−0.002	−0.001	0.028	−0.002	1.000

Source: compiled by the authors.

One can notice that there are some missing values in the training data. To avoid missing observations, it was decided to impute the missing values with medians. With this imputation the coefficients of the imputed variables are not affected, but the inclusion of more observations gives more information for the calculation of other variables' coefficients.

Those variables which were decided to be included in the model were used to form Model 1 (Table 8). Not all variables proved to be significant, however the predictive power of this base model is acceptable on the training data: the overall accuracy is about 80%, the area under the ROC curve is 0.84.

However, the sensitivity ratio (the share of correctly classified defaulted firms) is only 63%, which can be considered as low accuracy, while the specificity ratio (the share of correctly classified non-defaulted firms) is close to 90%, which can be considered as very high accuracy. The sensitivity ratio seems to be more important in the case of default prediction, because classifying a near-to-default firm in a healthy group costs creditors more than classifying a healthy firm in a risky group. That is why it was decided to find the optimal cutoff to receive maximum performance in terms of both sensitivity and specificity. Figure 2 shows the dependence of sensitivity and specificity on the cutoff for Model 1 on the training data. The optimal cutoff is about 40% and a decent accuracy of about 76–78% in both sensitivity and specificity can be achieved on the training data. The results turn out to be consistent on the test data as well.

The next step was to reduce the number of variables in an attempt to improve the accuracy of the model. A stepwise forward selection approach was used to form Model 2 and Model 3 in Table 8. A 10% significance level was used for Model 2 and 5% for model 3. No significant improvement appears to have been obtained. Return on sales showed counterintuitive performance in Model 1 by being positively correlated with another regressor, ROA, thus we did not use this variable in Models 2 and 3.

Table 8

Logistic regression results for models with financial variables only

Dependent variable: D Independent variables:	Model 1 b/se	Model 2 b/se	Model 3 b/se
WCTA	−0.546** (0.20)		
CLR	−1.086*** (0.26)	−1.632*** (0.21)	−1.741*** (0.20)
ROA	−0.928** (0.29)	−0.380 (0.21)	
ROS	0.064** (0.02)		
ART	0.000* (0.00)	0.000 (0.00)	
APT	−0.000 (0.00)		
AT	0.000 (0.00)		
ARTA	2.875*** (0.39)	3.027*** (0.38)	3.113*** (0.38)
TLE	−0.000 (0.00)		
GREV	−0.008 (0.01)		
Constant	−1.682*** (0.40)	−1.196*** (0.35)	−1.056** (0.35)
N	864	864	864
BIC criterion	878.891	857.073	852.550

Cutoff = 0.5

In-sample performance			
Overall accuracy	80.79%	78.59%	78.47%
Sensitivity	63.19%	63.89%	66.32%
Specificity	89.58%	85.94%	84.55%
Area under ROC curve	0.8442	0.8406	0.8385
Out-of-sample performance:			
Overall accuracy	82.96%	77.78%	77.04%
Sensitivity	67.78%	66.67%	64.44%
Specificity	90.56%	83.33%	83.33%
Area under ROC curve	0.8612	0.8564	0.8494

Cutoff = 0.4

In-sample performance:			
Overall accuracy	76.85%	75.58%	73.96%
Sensitivity	75.69%	76.39%	76.74%
Specificity	77.43%	75.17%	72.57%
Area under ROC curve	0.8442	0.8406	0.8385
Out-of-sample performance:			
Overall accuracy	78.89%	77.04%	75.93%
Sensitivity	75.56%	78.89%	77.78%
Specificity	80.56%	76.11%	75.00%
Area under ROC curve	0.8612	0.8564	0.8494

* p < 0.05, ** p < 0.01, *** p < 0.001.
Source: compiled by the authors.

Sensitivity and specificity versus cutoff (Model 1, training data)
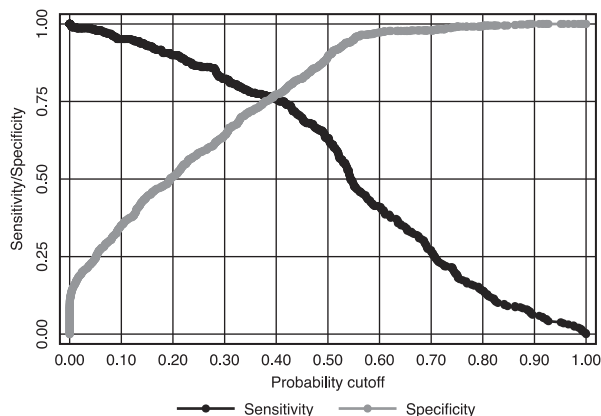
Source: compiled by the authors.

Default prediction using financial and non-financial variables

To choose the non-financial variables to be included in the models, we again ran the Mann-Whitney rank sum test, and it turned out that the mean values for all potential predictors differ significantly for defaulted and non-defaulted firms. The results of the test are shown in Table 9.

Mann-Whitney rank sum test results for non-financial variables

	N	Mann-Whitney test p-value	Conclusion
Age	864	0.020	Include
N_trials_l2	864	0.000	Include
Sum_trials_l2_to_TA	864	0.000	Include
N_inspections_l2	864	0.000	Include
N_viol_l2	864	0.000	Include
Sh_viol_l2	864	0.000	Include

Source: compiled by the authors.

We also excluded the number of unscheduled inspections because this variable is closely correlated with the number of violations detected and the number of court trials, to avoid multicollinearity.

Correlation table for non-financial variables

| Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) D	1.000						
(2) Age	−0.058	1.000					
(3) N_trials_l2	0.426	−0.035	1.000				
(4) N_inspections_l2	0.145	0.019	0.596	1.000			
(5) N_viol_l2	0.143	0.026	0.549	0.924	1.000		
(6) Sh_viol_l2	0.136	0.042	0.087	0.060	0.145	1.000	
(7) Sum_trials_l2_to_TA	0.158	0.007	0.138	0.021	0.005	0.022	1.000

Source: compiled by the authors.

We chose Model 3 as the base model and added all non-financial variables in the first step to form Model 4. Models 5 and 6 were constructed by stepwise forward selection using significance levels of 10% and 5% respectively.

The area under the ROC curve can be used as the main indicator of the quality of classifiers (models). One can notice that the use of non-financial variables led to an increase in this measure from roughly 0.84 to 0.91 (on the training data), which is a 8% increase.

To find the optimal cutoff, we again plotted training data sensitivity and specificity against the cutoff (Figure 3), and the optimal cutoff was close to 32%. Using this cutoff, the overall accuracy is close to 84–85% on both training and test data for models with non-financial variables, compared to 75–78% for models with financial variables only. The sensitivity on test data is lower (about 82–83%), but still much higher than in the case of models with financial variables only (Table 11).

Figure 3

Sensitivity and specificity versus cutoff (Model 4, training data)
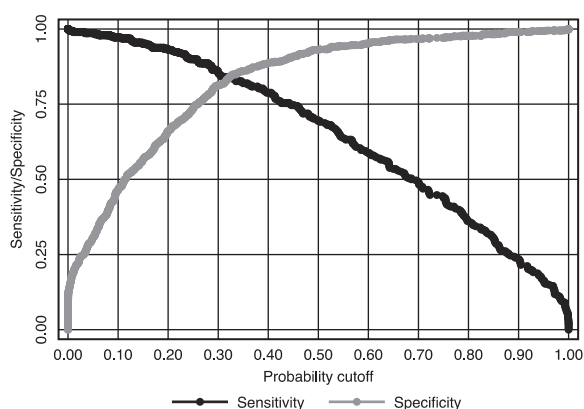
Source: compiled by the authors.

Table 11

Logistic regression results for models with financial and non-financial variables

	Model 4 b/se	Model 5 b/se	Model 6 b/se
Dependent variable: D Independent variables:			
CLR	−1.332*** (0.22)	−1.323*** (0.22)	−1.311*** (0.22)
ARTA	2.406*** (0.44)	2.350*** (0.43)	2.411*** (0.43)
Age	0.008 (0.01)		
N_trials_l2	0.089*** (0.01)	0.088*** (0.01)	0.089*** (0.01)
Sum_trials_l2_to_TA	0.326* (0.13)	0.331* (0.13)	0.330* (0.13)
N_viol_l2	−0.001 (0.00)		
Sh_viol_l2	0.760 (0.42)	0.754 (0.41)	
constant	−2.274*** (0.47)	−2.163*** (0.42)	−2.044*** (0.41)
N	864	864	864
bic			

	Model 4 b/se	Model 5 b/se	Model 6 b/se
Cutoff = 0.5			
In-sample performance			
Overall accuracy	85.30%	85.30%	85.30%
Sensitivity	69.44%	69.79%	69.44%
Specificity	93.23%	93.06%	93.23%
Area under ROC curve	0.9079	0.9077	0.9074
Out-of-sample performance:			
Overall accuracy	82.59%	82.59%	81.11%
Sensitivity	54.44%	54.44%	53.33%
Specificity	96.67%	96.67%	95.00%
Area under ROC curve	0.9092	0.9093	0.9121
Cutoff = 0.32			
In-sample performance:			
Overall accuracy	83.68%	83.80%	84.72%
Sensitivity	84.03%	84.38%	84.38%
Specificity	83.51%	83.51%	84.90%
Area under ROC curve	0.9079	0.9077	0.9074
Out-of-sample performance:			
Overall accuracy	85.93%	85.56%	84.81%
Sensitivity	82.22%	82.22%	83.33%
Specificity	87.78%	87.22%	85.56%
Area under ROC curve	0.9092	0.9093	0.9121

* p < 0.05, ** p < 0.01, *** p < 0.001.
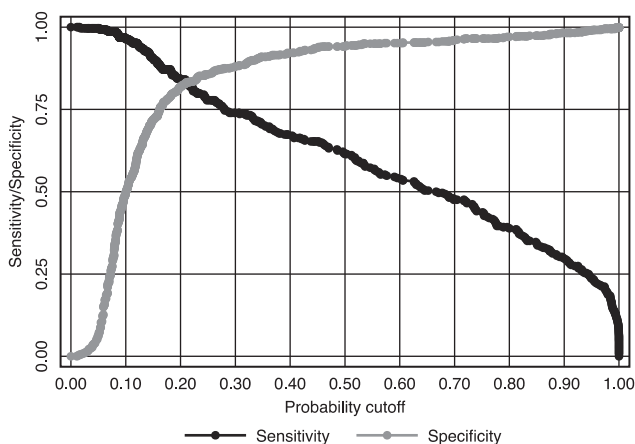*Source: compiled by the authors.*

Prediction using non-financial data only

We went even deeper into the analysis of non-financial variables predictability for defaults and ran several models with only non-financial variables (Table 12). First, we included all non-financial variables from the list (Model 7), then we applied stepwise selection (Model 8).

We again plotted training data sensitivity and specificity against the cutoff (Figure 4) for Model 8 (as the best in terms of the area under the ROC curve), and the optimal cutoff was close to 21%.

Figure 4

Sensitivity and specificity versus cutoff (Model 8, training data)

Source: compiled by the authors.

The results were highly promising. The predictive power of classification models with non-financial variables only is higher than that of models with financial variables only. The overall accuracy, sensitivity and specificity using the 21% cutoff is about 83–84% for the training set. The accuracy for the test data is about 82%, but the sensitivity is about 76%, which is roughly the same as for the financial variables only. This means that it is not necessary to obtain any financial coefficients to estimate the risk of default for HUMFs: the number of lawsuits, sum of legal claims and information related to inspections are enough to classify HUMFs into risky and non-risky groups with decent accuracy.

Table 12

Logistic regression results for models with non-financial variables only

	Model 7 b/se	Model 8 b/se
Dependent variable: D Independent variables:		
Age	−0.028 (0.02)	−0.044* (0.02)
N_trials_l2	0.095*** (0.01)	0.091*** (0.01)
N_viol_l2	−0.004 (0.00)	
Sh_viol_l2	1.466*** (0.39)	1.565*** (0.42)
Sum_trials_l2_to_TA	0.846*** (0.16)	1.362*** (0.21)
Constant	−2.053*** (0.22)	−2.288*** (0.25)
N	864	864
bic	778.374	754.671

Cutoff = 0.5

In-sample performance		
Overall accuracy	82.41%	83.33%
Sensitivity	59.38%	61.46%
Specificity	93.92%	94.27%
Area under ROC curve	0.8937	0.8999
Out-of-sample performance:		
Overall accuracy	79.63%	80.00%
Sensitivity	48.89%	50.00%
Specificity	95.00%	95.00%
Area under ROC curve	0.8711	0.8715

Cutoff = 0.21

In-sample performance:		
Overall accuracy	79.86%	83.33%
Sensitivity	87.15%	83.68%
Specificity	76.22%	83.16%
Area under ROC curve	0.8937	0.8999
Out-of-sample performance:		
Overall accuracy	78.52%	81.85%
Sensitivity	81.11%	75.56%
Specificity	77.22%	85.00%
Area under ROC curve	0.8711	0.8715

* p < 0.05, ** p < 0.01, *** p < 0.001.
Source: compiled by the authors.

CONCLUSION

Traditionally, financial ratios have been considered as predictors of default. Since the late 1960s, there have been numerous attempts to create classification models for default prediction using financial ratios, and many of them have been successful.

However, in the case of service firms in Russia, and HUMFs in particular, financial statements do not in all cases reflect the true condition of the firm because of shadow operations and business disaggregation, which make the ratios of a legal entity biased. Moreover, the full financial statements of private firms are not always obtainable without access to special databases. At the same time, given the specificities of HUMFs, accurate default prediction for such firms is highly relevant due to the unpleasant effect that defaults have on clients. This study sheds light on the possibility of using open access non-financial data to predict default of HUMFs, either in combination with financial data or even separately.

To create a basic classifier, we used financial ratios which cover liquidity, solvency, profitability and business turnover. Then we added non-financial variables: information related to court trials in which the firm participated, information related to unscheduled inspections and detected violations, and the age of the firm.

We have chosen a statistical tool — logistic regression — in order to be able to interpret the strength and direction of the relationship between independent variables and defaults, which is not a common practice while using non-linear machine learning algorithms.

First, a classification model was constructed with only financial variables, which was accurate with the ROC AUC of 0.84 on the training data and 0.86 on the test data. Then non-financial variables were added, and this led to an increase in the ROC AUC up to 0.91 on both the training and test data. A model with only non-financial variables was then constructed and the ROC AUC was about 0.89–0.9 for the training data and 0.87 for the test data, indicating that a model without financial variables is more accurate than a classifier constructed with only financial data.

These results mean that although the conventional classification with financial variables gives acceptable results (ROC AUC of 0.86), the accuracy can be increased significantly (a 8% increase in ROC AUC) by adding non-financial data to the model. Moreover, if a firm's financial statements are not accessible or are known to be biased, it is possible to assess the risk of default for such a firm using only information related to litigation, inspections and age.

The findings of this study correspond to the findings of other researchers in this field [Altman et al., 2010; Blanco et al., 2015; Fernando et al., 2020]. The results obtained can be widely applied for default prediction (credit risk estimation) by credit institutions and any party interested in the assessment of financial stability of a HUMF (clients, counterparties, employees, etc.).

References

Kasgari A.A., Divsalar, M., Javid M.R. and Ebrahimian S.J. (2013). Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications*, vol. 23, no. 3, pp. 927–936.

Altman E.I., Sabato G. et al. (2010). The value of non-financial information in SME risk management. *The Journal of Credit Risk*, vol. 6, no. 2, pp. 95–127. https://doi.org/10.21314/JCR.2010.110.

Altman E.I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, vol. 23, no. 4, pp. 589–609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x.

Altman E.I., Marco G. et al. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, vol. 18, no. 3, pp. 505–529. https://doi.org/10.1016/0378-4266(94)90007-8.

Andrade G., Kaplan S.N. (1998). How Costly is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions that Became Distressed. *The Journal of Finance*, vol. 53, no. 5, pp. 1443–1493. https://doi.org/10.1111/0022-1082.00062.

Bandyopadhyay A. (2006). Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches. *The Journal of Risk Finance*, vol. 7, no. 3, pp. 255–272. https://doi.org/10.1108/15265940610664942.

Barboza F., Kimura H. et al. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, vol. 83, no. C, pp. 405–417. https://doi.org/10.1016/j.eswa.2017.04.006.

Beaver W.H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, vol. 4, pp. 71–111. https://doi.org/10.2307/2490171.

Bhimani A., Gulamhussen M.A. et al. (2013). The Role of Financial, Macroeconomic, and Non-financial Information in Bank Loan Default Timing Prediction. *European Accounting Review*, vol. 22, no. 4, pp. 739–763. https://doi.org/10.1080/09638180.2013.770967.

Blanco A., Diéguez A. et al. (2015). Improving bankruptcy prediction in micro-entities by using nonlinear effects and non-financial variables. *Finance a Uver — Czech Journal of Economics and Finance*, vol. 65, no. 2, pp. 144–166.

Boubaker S., Cellier A. et al. (2020). Does corporate social responsibility reduce financial distress risk? *Economic Modelling*, vol. 91, no. C, pp. 835–851. https://doi.org/ 10.1016/j.econmod.2020.05.012.

Cao Y., Liu X. et al. (2020). A two-stage Bayesian network model for corporate bankruptcy prediction. *International Journal of Finance & Economics*. https://doi.org/10.1002/ijfe.2162.

Coats P.K., Fant L.F. (1993). Recognizing Financial Distress Patterns Using a Neural Network Tool. *Financial Management*, vol. 22, no. 3.

Fernando J.M.R., Li L. et al. (2020). Financial versus Non-Financial Information for Default Prediction: Evidence from Sri Lanka and the USA. *Emerging Markets Finance and Trade*, vol. 56, no. 3, pp. 673–692. https://doi.org/10.1080/1540496X.2018.1545644.

Grigoriev A., Tarasov K. (2019). Corporate Bankruptcy Prediction Using the Principal Components Method. *Journal of Corporate Finance Research*, vol. 13, no. 4, pp. 20–38. https://doi.org/10.17323/j.jcfr.2073-0438.13.4.2019.20-38.

Grunert J., Norden L. et al. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, vol. 29, no. 2, pp. 509–531. http://dx.doi.org/10.2139/ssrn.302689.

Gruszczyński M. (2004). Financial Distress of Companies in Poland. *Department of Applied Econometrics, Warsaw School of Economics, Working Papers*, vol. 10.

Hunter J., Isachenkova N. (2001). Failure risk: A comparative study of UK and Russian firms. *Journal of Policy Modeling*, vol. 23, no. 5, pp. 511–521. https://doi.org/10.1016/S0161-8938(01)00064-3.

Jaki A., Ćwięk W. (2021). Bankruptcy Prediction Models Based on Value Measures. *Journal of Risk and Financial Management*, vol. 14, no. 1, pp. 1–14. https://doi.org/10.3390/su14031416.

Kachalin D.S. (2011). Analysis of Russian models for business disaggregation, which ensure that the scale of the business is appropriate for special tax regimes. *Finansovaya Analitika: Problemi i Resheniya*, vol. 47, no. 5, pp. 52–63 (In Russ.).

Kovacova M., Kliestik T. (2017). Logit and Probit application for the prediction of bankruptcy in Slovak companies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, vol. 12, no. 4, pp. 775–791. https://doi.org/10.24136/eq.v12i4.40/.

Kovalenko Yu. (2019). Some legal problems of invalidation of decisions of estate owners in apartment buildings in case of management organizations insolvencies. *Prolog: Zurnal o prave*, no. 3, pp. 70–77 (In Russ.).

Kumar P., Ravi V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques — A review. *European Journal of Operational Research*, vol. 180, no. 1, pp. 1–28.

Kvitova K., Yavorskaya Ya.S. (2018). Ways of solving the problem of insolvency of management firms (evidence from Mordovia republic). *Ogarev-Online*, vol. 107, no. 2, p. 9 (In Russ.).

Laitinen E.K. (2011). Assessing viability of Finnish reorganization and bankruptcy firms. European *Journal of Law and Economics*, vol. 31, no. 2, pp. 167–198. https://doi.org/10.1007/s10657-009-9136-4.

Lin F., Liang D. et al. (2010). The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology*, vol. 18, no. 4. pp. 504–513.

Lugovskaya L. (2010). Predicting default of Russian SMEs on the basis of financial and non-financial variables. *Journal of Financial Services Marketing*, vol. 14, no. 4, pp. 301–313. https://doi.org/10.1057/fsm.2009.28.

Mai F., Tian S. et al. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, vol. 274, no. 2, pp. 743–758. https://doi.org/10.1016/j.ejor.2018.10.024.

Makeeva E., Sinilshchikova M. (2020). News Sentiment in Bankruptcy Prediction Models: Evidence from Russian Retail Companies. *Journal of Corporate Finance Research*, 14, no. 4, pp. 7–18.

Mselmi N., Lahiani A. et al. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, vol. 50, no. C, pp. 67–80. https://doi.org/10.1016/j.irfa.2017.02.004.

Muñoz-Izquierdo N., Laitinen E.K. et al. (2020). Does audit report information improve financial distress prediction over Altman's traditional Z-Score model? *Journal of International Financial Management & Accounting*, vol. 31, no. 1, pp. 65–97. https://doi.org/10.1111/jifm.12110.

Odom M., Sharda R. (1990). A Neural Network Model for Bankruptcy Prediction. *IEEE International Joint conference on Neural Networks*. pp. 163–168. https://doi.org/10.1109/IJCNN.1990.137710.

Ohlson J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131. https://doi.org/10.2307/2490395.

Shumway T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, vol. 74, no. 1, p. 101. https://doi.org/10.1086/209665.

Sirirattanaphonkun W., Pattarathammas S. (2012). Default Prediction for Small-Medium Enterprises in Emerging Market: Evidence from Thailand. *Seoul Journal of Business*, vol. 18. no. 2, pp. 25–54. https://doi.org/10.35152/snusjb.2012.18.2.002.

Sukharev A.N., Golubev A.A. et al. (2018). The financial mechanism of management organizations in the field of housing and communal services: Deformation problems. *Finance and Credit*, vol. 24, no. 5, pp. 1063–1078 (In Russ.). https://doi.org/10.24891/fc.24.5.1063.

Williams C.C., Nadin S. et al. (2013). Explaining off-the-books entrepreneurship: a critical evaluation of competing perspectives. *International Entrepreneurship and Management Journal*, vol. 9, no. 3, pp. 447–463. https://doi.org/10.1007/s11365-011-0185-0.

Wilson N., Ochotnický P. et al. (2016). Creation and destruction in transition economies: The SME sector in Slovakia. *International Small Business Journal*, vol. 34, no. 5, pp. 579–600. https://doi.org/10.1177/0266242614558892.

Xie C., Luo C. et al. (2011). Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality and Quantity*, vol. 45, no. 3, pp. 671–686. https://doi.org/10.1007/s11135-010-9376-y.

Zhang G., Hu M.Y. et al. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, vol. 116, no. 1, pp. 16–32.

Information about the authors

Vladislav V. Afanasev, *Teacher, Department of Finance, HSE University, St. Petersburg*

Yulia A. Tarasova, *Candidate of Economic Sciences, Associate Professor, Department of Finance, HSE University, St. Petersburg*

Предсказание неплатежеспособности для управляющих компаний ЖКХ с использованием нефинансовых данных

Владислав Викторович Афанасьев, *преподаватель департамента финансов, аспирант Национального исследовательского университета «Высшая школа экономики», Санкт-Петербург 190008, Российская Федерация*
E-mail: vvafanasev@hse.ru, ORCID: 0000-0002-4041-4465

Юлия Александровна Тарасова, *кандидат экономических наук, доцент департамента финансов Национального исследовательского университета «Высшая школа экономики», Санкт-Петербург 190008, Российская Федерация*
E-mail: yutarasova@hse.ru, ORCID: 0000-0001-9341-3151

Аннотация

Традиционно финансовые показатели используются в качестве объясняющих переменных при прогнозировании неплатежеспособности. Однако существуют предпосылки к неточности финансовой отчетности российских компаний, что ставит под вопрос возможность использования такого подхода. Альтернативой является использование нефинансовых данных в подобных моделях.

Цель данной статьи — исследование возможности использования нефинансовых данных, связанных с информацией о судебных спорах, внеплановых проверках и возрасте компании в моделях предсказания неплатежеспособности для существенного увеличения точности предсказания моделей для управляющих компаний ЖКХ.

Рынок управляющих компаний выбран как один из наиболее «рисковых». Неплатежеспособность таких компаний влияет не только на традиционный пул стейкхолдеров (кредиторы, акционеры, сотрудники), но также и на их клиентов.

Выборка из 378 управляющих компаний, попавших в ситуацию дефолта, и 765 «здоровых» компаний была использована для создания и оценки точности модели. В качестве классификационного алгоритма использовалась логистическая регрессия.

Результаты свидетельствуют о том, что добавление в модели нефинансовых переменных может существенно увеличить точность предсказания неплатежеспособности. Более того, модели с использованием исключительно нефинансовых переменных демонстрируют высокую точность предсказания.

Теоретическая значимость исследования заключается в эмпирическом доказательстве обоснованности использования нефинансовых данных в моделях прогнозирования неплатежеспособности. Кроме того, в рамках исследования использована информация о внеплановых проверках предприятий, что представляется первым случаем включения подобных переменных в модели прогнозирования неплатежеспособности.