UNIVERSITY OF TARTU

Faculty of Social Sciences

School of Economics and Business Administration

Keijo Kohv

PREDICTION OF BANK LOAN DEFAULTS WITH FINANCIAL RATIOS, TAX ARREARS AND ANNUAL REPORT DELAYS

Master’s thesis

Supervisor: Senior Research Fellow Oliver Lukason (PhD)

Tartu 2020

Name and signature of supervisor: ....

Allowed for defence on: ....

(date)

I have written this master's thesis independently. All viewpoints of other authors, literary sources and data from elsewhere used for writing this paper have been referenced.

……………………………………..

(author's signature)

Abstract

This thesis aims to compare the accuracy of financial ratios, tax arrears and annual report submission delays in bank loan default prediction. For this, twelve variables from these three domains are used. The study uses whole population dataset from an Estonian commercial bank with 12901 observations of defaulted and non-defaulted firms. The analysis is performed using statistical (logistic regression) and machine learning (multilayer perceptron) methods. Out of the three domains used, tax arrears show high prediction capabilities for loan defaults, while financial ratios and reporting delays do not. Default prediction accuracies were 84% with tax arrears only and 89.1% with all variables combined. To date, it aims to be the first research paper to present the abilities of tax arrears and reporting delays to predict bank loan defaults.

Keywords: failure prediction, loan defaults, tax arrears, reporting delays

CERCS: S181, S190, S192

Contents

1. Introduction

Understanding and predicting business failure is a constantly evolving field of research. The idea is simple – if failure can be predicted, stakeholders can take measures to minimize potential damage or to avoid failure in first place. That is important because when companies fail they can have significant negative social and financial impact on owners of the failed businesses, employees, financiers, clients and other stakeholders, but also to economies and societies in general (Alaka *et al*. 2018; Camacho-Miñano *et al*. 2015; Wu 2010). Business failure in general has a broad range of definitions. For example in their study Dias and Teixeira (2017: 3) analysed 201 journal articles on the topic and found that business failure is most commonly defined as an event of bankruptcy, business closure, ownership change, or failure to meet expectations. In addition business failure could mean bond default, bank loan default, delisting of a company, government intervention and liquidation (Altman and Narayanan 1997). The most commonly used failure definition is bankruptcy and vast majority of studies on business failure have focused on creating bankruptcy prediction models – however, bankruptcy is only one of the outcomes on the large scale of possible endings of business failure process (Balcaen and Ooghe 2006: 73). Weitzel and Jonsson (1989) created a stage model for business failure process, where every stage is seen as failure of some sort. According to the stage model (*Ibid.*: 102), payment default is connected with the crises stage, logically seen as a result of factors such as blindness, inaction and faulty actions from earlier stages. At this point, effective reorganization might save the company, and if not, the company would be dissolved. It can be said that payment default is one of the most serious warning signals that the company has remarkably higher risk of ultimate business failure. One that might end in bankruptcy, merger, absorption, dissolution or liquidation (Balcaen and Ooghe 2006: 73).

From creditor's point of view, in order to avoid such negative consequences, it is vital to assess a company's probability of failure, in order to make sounder credit decisions and to appropriately compensate the risk in expected returns, or to avoid crediting unhealthy firms at first place (Alaka *et al*. 2018; Atiya 2001; Xu and Zhang 2009). Many banks and other credit institutions have set up an automated system giving early warning signals about potential bankruptcy – this provides necessary window for the stakeholders to take action and try to minimize negative consequences (Laitinen 2008). Additionally, most creditors (especially financial institutions, including banks) have implemented failure prediction models, usually in the form of payment default prediction, in their internal credit risk modelling methodology. Indeed, company failure prediction is a quickly growing research domain that affects the whole economy, and thus, its importance is hard to underestimate.

Classical studies of the area include univariate (Beaver 1966) and multivariate (Altman 1968) failure prediction models that used historical accounting data (financial ratios) as independent financial variables. While pioneering studies were based on statistical models, the latest innovations in failure prediction take advantage of artificial intelligence and machine learning techniques. It can be concluded that there are numerous techniques applied in hundreds of studies that mostly use financial ratios to create failure prediction models with high prediction accuracies – for example, see review by Sun *et al.* (2014).

This thesis aims to compare the accuracy of financial ratios, tax arrears and annual report delays in bank loan default prediction. A three-layer analysis is performed using two methods: logistic regression and multilayer perceptron. This avoids single method bias and gives holistic perspective about the prediction accuracies of the three layers: single variables, all variables from a domain, variables from all three domains. To date, there are only a few failure prediction papers that have defined failure via loan default, and thus, used loan default as a dependent variable. This sort of data is usually not available for research purposes. Derived from the latter, the most commonly used firm failure definition (and dependent variable used) is permanent insolvency. As mentioned previously, in failure process, default is located before permanent insolvency, and thus, default prediction capability of financial variables that are commonly used in the context of permanent insolvency, is unknown. Therefore, two novel domains, tax arrears and reporting delays, are included in current study – neither have been studied in prior literature in current setting, i.e. loan default prediction.

The thesis has the following structure. Review of literature consisting of two subsections: first, the theoretical background of company failure, and second, an overview of financial and non-financial variables used in previous research. This is followed by an overview and explanation of data, variables and methods used in the empirical part. Thereupon the results and discussion are presented. The thesis ends with a conclusion in the last chapter.

2. Review of literature

2.1. Definition of loan payment default in the context of firm failure

The term payment default has several definitions. In general it can be said that payment default occurs when a firm is unable to pay its financial obligations as they are due (Altman 1968; Beaver 1966). In other words – obligor is experiencing financial distress, which ultimately may or may not lead to business failure (Höglund 2017: 369). In an SME[1]-specific study, Altman *et al.* (2010) also emphasize the importance to differentiate between firm failure and firm closure. In his study about U.S. firms, Headd (2003) showed that about a third of all closed businesses were financially successful, so not all closures are failures – there can be reasons other than financial distress to close a company.

Firms that are experiencing financial distress may never face legal failure (bankruptcy, for example). Besides bankruptcy, additional outcomes of a financially distressed (i.e. defaulted) company include dissolution, merger, liquidation or sustaining operations (Camacho-Miñano *et al.* 2015: 341). A recovery from payment default is achieved usually through restructuring (reorganization). In case a firm is financially distressed and temporarily unable to pay its obligations, it could avoid bankruptcy by using a reorganization that in principle is available in many countries worldwide, subject to country-specific legal framework (Laitinen 2008). Even though existing practice shows relatively small success rate of reorganizations (Lukason and Urbanik 2013), similarly to failure prediction, the outcome of reorganization can be predicted. In their Estonia-based study, Lukason and Urbanik (2013) showed that financial variables reflecting solvency, profitability and capital structure domains were not remarkably different between failed and non-failed reorganized firms. In contrast, non-financial information, such as economic sector, firm size, shareholder structure, availability of reorganization plan, do greatly impact the prediction accuracy of reorganization outcome (Camacho-Miñano *et*

[1] SME is abbreviation of small and middle-sized enterprises. It should be noted that according to EU definition, SME definition also includes micro companies (European Commission 2020).

al. 2015; Laitinen 2008; Lukason and Urbanik 2013). Therefore, it is essential to include non-financial information also when predicting the success of reorganization.

Since current study uses a dataset from a commercial bank and has main focus on loan payment defaults of credit contracts, the default in current thesis is defined based on Basel II framework that states: "…the obligor is past due more than 90 days on any material credit obligation to the banking group." (Basel Committee… 2016: 8). Thus, unlike bankruptcy, which as a form of business failure is subject to a country's legal and financial frameworks, loan payment default is universally defined through Basel II (Bhimani *et al.* 2010). When company is 90 days overdue in payments, it shows serious financial distress. Nevertheless, in failure prediction studies, existing research about loan defaults is scarce compared to bankruptcy studies. That is probably because loan payment default usually occurs in the context of confidential lender-borrower relationship, making the data itself also strictly confidential. Several countries, such as Estonia and Finland, however do offer free public access to national tax arrears database (a State-firm relationship). Still, to the author's best knowledge there is no public registry in European countries, where one can obtain lender-borrower based public data of payment defaults, which *per se* can be temporary. Therefore, majority of the studies that have previously used defaults as dependent variables, have actually considered bankruptcy or permanent insolvency as default[2] since that information is easily accessible and specifically defined in certain country context.

2.2. Financial and non-financial variables in firm failure prediction

This thesis relies on Iwanicz-Drozdowska et al. (2016) classification of variables into financial and non-financial, where the former are variables calculated by using information from annual financial statements, the latter being therefore all other variables. Even though Balcaen and Ooghe (2006: 79) concluded in their study that there seems to be no superior predictor variables or superior methods when it comes to firm failure prediction, it has been found in previous research that the most used and the most important financial variables are financial ratios that come from liquidity (current ratio, working capital/total assets), solvency (total debt/total assets) and profitability (net income/total assets) categories (Dimitras *et al.* 1996; Bellovary *et al.* 2007). Therefore,

[2] For reasoning, see for example studies by Altman and Sabato (2007) and Ciampi (2018).

liquidity, solvency and profitability seem to best indicate financial distress when a firm deteriorates (Höglund 2017: 369).

Albeit the most used technique in failure prediction uses financial variables in the form of financial ratios calculated from a firm's financial statement, there are several limitations. Besides multicollinearity in between financial ratios that is frequently seen in relevant research, some of the weaknesses of financial variables are firstly the availability of the data in a timely manner, and secondly information asymmetry in the data itself. Researchers assume that the financial figures in financial statements give a true and fair view of the company's situation. However previous studies have shown that is not always the case, especially with failing companies, where financial data might be manipulated in an attempt to hide or postpone the emergence of financial weaknesses (Balcaen and Ooghe 2006; Ciampi *et al.* 2020; Laitinen and Laitinen 2009). Additionally, in case a company is non-audited as smaller companies usually are, its reports are less reliable and often deliberately opaque, setting additional obstacles in using only financial variables in firm failure prediction (Altman *et al.* 2010; Ciampi 2015; 2018). Moreover, since majority of active companies worldwide consists of SMEs[3] (including the dataset of present study), even small movements in absolute figures can lead to exaggerated changes in financial ratios (Ciampi 2015). At the same time SMEs have relatively smaller financial buffers to withstand sudden financial distress (Beaver 1966). Combined with the findings of Lukason and Laitinen (2019) whereby even the latest annual reports might not sufficiently indicate worsening of a firm's situation and the fact that some of the commonly used financial ratios for big companies are completely ineffective in terms of SMEs (Ciampi 2015), occurrences of payment defaults and decline of a firm in general can be sudden and quick (Ciampi *et al.* 2020).

Seeking to overcome aforementioned limitations, recent studies have included non-financial independent variables in order to improve the accuracy of failure prediction models (Ciampi 2015). As explained earlier, in failure prediction literature, the term non-financial variable is commonly used by researchers[4] for various variables other than

[3] In most countries, small and middle sized enterprises (SMEs) are considered as backbone of economies, since they represent usually more than 95% of all businesses and provide roughly two-thirds of jobs in private sector (Altman & Sabato 2007; European Commission 2019; Gordini 2014; Merwin 1942).
[4] See for example Iwanicz-Drozdowska et al. (2016).

financial ratios calculated by using financial statements. Prediction models that include non-financial variables such as previous payment patterns (Back 2005); corporate governance (Ciampi 2015; Süsi and Lukason 2019); as well as reporting and compliance (Altman *et al.* 2010) and tax arrears (Lukason and Andresson 2019) have outperformed classical prediction models based solely on financial variables. Non-financial information, such as previous payment history, holds more updated information compared to financial data (Laitinen 2011). Thus, banks and other financial institutions, whose credit portfolio mainly consists of SMEs, must implement non-financial information in their credit scoring models, since the models that base only on financial ratios would be ineffective (McCann and McIndoe-Calder 2015). In their international study Altman *et al.* (2017: 166) showed that inclusion of non-financial variables generally improves the classical Z-score model's accuracy, however the results vary by countries. Of course, the usage of non-financial variables is always limited to country-specific data availability. Compared to financial variables, non-financial variables are typically less correlated with one another or with financial ratios (Altman *et al.* 2020: 4). Several studies (e.g. Altman and Sabato 2007; Ciampi 2015) have proven that the classical failure prediction models that work well for large companies, are not the best fit for SMEs.

While the literature on general business failure is vast and constantly growing, existing literature focusing on loan payment default prediction (therefore default being the dependent variable) is rare. Even more rare are the studies that have included non-financial variables in the context of business failure prediction, thus remaining largely unexplored research area (Ciampi 2018). Articles where loan default (or proxy of it) is used as a dependent variable or articles where the failure prediction model has incorporated non-financial variables, have been summarized and analysed in Table 1. It should be noted that articles in the area where default is used in the context of hazard models were excluded, because current study aims to compare the loan default prediction accuracy of various variables statistically, and not to disclose, which are the significant variables determining firm survival dynamically.

Table 1. Failure prediction articles using non-financial variables and/or using loan default as a dependent variable

Author	Country	Sample size	Non-financial variables used in prediction models	Dependent variable	Accuracy
Altman *et al.* (2020)	Finland	51 099	firm type; age; industry bankruptcy risk; prior defaults; industry payment default risk; delayed reporting; auditor's report; payment delays >60 days; number of payment delays >60 days; delays/total assets; board characteristics	bankrupt /non-bankrupt	>93% (test sample, in short TS)
Ciampi *et al.* (2020)	Italy	1 200	past due exposures > 60 days (E60D); (E60D)/turnover; (E60D)/EBITDA; (E60D)/cash flow; (E60D)/bank loans; (E60D)/financial debt; number of payment delays > 60 days; number non-remedied payment delays > 60 days; number of cumulative payment delays > 60 days; number of cumulative non-remedied payment delays > 60 days	defaulted/non-defaulted	85.3% (TS)
Lukason and Andresson (2019)	Estonia	4 515	tax arrears	bankrupt /non-bankrupt	91.3% (holdout sample, in short HOS)
Süsi and Lukason (2019)	Estonia	67 058	board size; board gender heterogeneity; board tenure; age of top managers; multiple directorships; ownership concentration; managerial ownership	failure risk	n/a[5]
Ciampi (2018)	Italy	382	corporate social responsibility (CSR) towards employees; CSR towards customers; CSR towards suppliers; CSR towards the local community; CSR environmental aspects	defaulted/non-defaulted	82.8% (HOS)
Altman *et al.* (2017)	Inter-national	2 640 000	year of bankruptcy; size; age; industry; country of origin	failed/non-failed	82.3% (TS)
Höglund (2017)	Finland	768	industry risk of payment defaults; industry risk of bankruptcy;	tax default	73.8% (TS)

[5] The study explored the interconnection between corporate governance variables and failure risk however, specific prediction accuracies were not part of the scope.

Ciampi (2015)	Italy	934	audit committee; board size; CEO turnover; CEO-duality; creditor ownership; director turnover; board member education; number of CEOs and chairpersons; outside directors present but in a proportion lower than 50%; outside directors present but in a proportion equal or higher than 50%; ownership concentration; % held by institutions; % held by managers and directors	defaulted/non-defaulted	87% (HOS)
McCann and McIndoe-Calder (2015)	Ireland	6,745	manager or owner has been with the firm 10 years or more; industry sector	defaulted/non-defaulted	80.6% (TS)
Bhimani *et al.* (2013)	Portugal	17,000	financial support from partners; type of management; ownership of assets; management skill	time to loan default	90.1% (HOS)
Laitinen (2011)	Finland	65,164	industry; age; board characteristics; audit report; number of payment defaults; number of payment delays; number of positive payment signals; firm type; months to the date of last financial reports; length of last accounting period;	viable/ non-viable	89.2% (TS)
Bhimani *et al.* (2010)	Portugal	31,025	size; age; industry; geographic regions	defaulted/non-defaulted	77.9% (HOS)
Altman *et al.* (2010)	UK	5,800, 000	audit information; late filing; age; subsidiary; size; industry; no cashflow statement; country court judgement	failed/non-failed	78% (HOS)
Back (2005)	Finland	3,199	management relation disturbance; management own payment disturbance; payment disturbances; payment delays; age; group membership; size	defaulted/non-defaulted	81.2% (HOS)
Atiya (2001)	USA	1,160	stock price volatility	bankrupt/non-bankrupt	85.5% (HOS)

Source: compiled by the author; Note: in case a paper found several prediction accuracies, for example for different methods, the table reflects only the highest accuracy.

The articles in Table 1 have analysed default prediction based on different non-financial information, such as firm age (Altman *et al.* 2010; Altman *et al.* 2017; Back 2005; Bhimani *et al.* 2013); firm size (Altman *et al.* 2010; Altman *et al.* 2017; Back 2005; Bhimani *et al.* 2013); industry sector (Altman *et al.* 2010; Altman *et al.* 2017; Bhimani *et al.* 2013; Höglund 2017; Laitinen 2011); management characteristics (Back 2005; Bhimani *et al.* 2013; Ciampi 2015; Laitinen 2011; Süsi and Lukason 2019); previous payment history (Back 2005; Ciampi *et al.* 2020; Laitinen 2011; Lukason and Andresson 2019); corporate social responsibility (Ciampi 2018); tax arrears (Lukason and Andresson 2019); year of bankruptcy (Altman *et al.* 2017); country of origin (*Ibid*.); financial support from partners (Bhimani *et al.* 2013); ownership of assets (*Ibid*.); stock price volatility (Atiya 2001); audit information (Altman *et al.* 2010); late filing of reports (*Ibid*.); and country court judgement (*Ibid*.). It must be noted that while some of the variables have actual failure prediction capability, others have been used as additional or control variables. Default prediction accuracies in studies listed in Table 1 range from 73-93%, whereby in all the studies the models that combined financial and non-financial information outperformed the models that used only financial information. For example, in their broad international study, Altman *et al.* (2010) gained 13% improvement in default prediction model's accuracy by including non-financial variables. The most used methods in reviewed articles were logistic regression and artificial neural networks.

Among other findings in existing literature related to non-financial information, a strong link can be found between financial distress and payment history: when the number of past payment delays increases, so does the probability of future financial difficulties (Back 2005; Laitinen 1999; 2011). In their study based on Estonian firms, Lukason and Andresson (2019) showed that up to one year prior to bankruptcy, the prediction model using tax arrears has clearly better failure prediction capability than the model using financial ratios[6], while the best results were obtained by combining financial information with tax arrears. Additionally, several studies have previously focused on non-compliance with regulation indicating a clear link between late filing or non-submission of accounts with increased risk of financial distress and business failure (Altman *et al.* 2010; Lukason 2013; Lukason and Camacho-Miñano 2019).

[6] Note: the accuracies were 89.5% (tax arrears) and 79.5% (financial ratios); however, the accuracies did not differ for the period concerning 13-24 months prior to bankruptcy.

3. Dataset, variables and methods

3.1. Dataset

Time period viewed in current study has been set to 2013-2018. This aims to neglect global financial crises effect from the analysis. It is a period of stable economic growth with average GDP growth of 3.2% per year[7] (Statistics Estonia 2020). At the same time as fresh Estonian dataset as possible is used. Three main sources are employed to gather necessary data for analysis: a) financial data and reporting delays dataset from Estonian Business Register; b) tax payment delays dataset from Estonian Tax and Customs Board; c) whole population dataset from an Estonian commercial bank consisting of firms that had signed at least one credit contract during the period and are grouped as defaulted and non-defaulted. Breakdown of the dataset observations and quartile values of tax arrears and reporting delays can be seen in dataset overview in Table 2 below.

Table 2. Overview of the dataset

Year	Observations		Maximum tax arrears in euros				Reporting delay in days			
			D			ND	D			ND
	D	ND	Q1	Q2	Q3	Q1-Q3	Q1	Q2	Q3	Q1-Q3
2013	5	3808	504	1791	2378	0	0	0	0	0
2014	15	884	229	1945	8492	0	0	0	0	0
2015	40	1198	263	1263	8184	0	0	0	28	0
2016	42	1752	1327	6121	18837	0	0	0	71	0
2017	20	2282	1728	6572	19327	0	0	31	244	0
2018	34	2821	213	1729	7061	0	0	0	120	0
Total	156	12745								

Source: compiled by the author; Note: D means defaulted; ND means non-defaulted; Q1 is the first quartile (25th percentile); Q2 is median value (50th percentile); Q3 is the third quartile (75th percentile). Information about the periods used to calculate respective variables can be seen in chapter 3.3.

In this study, only loan defaults are used, in case of which loan payment has been overdue for at least for 90 days or more. Several earlier studies have used roughly equal-sized samples of defaulted and non-defaulted firms, however this might create a bias, since the non-defaulted sample would not represent the population it originates from (Lukason and Andresson 2019). Thus, present study is based on whole population of firms in the context

[7] Statistics Estonia uses chain-linked growth rates of the GDP.

of an Estonian commercial bank. It includes firms from every economic sector and there are no limitations in terms of company characteristics, albeit more than 90% of the companies are SMEs. The no-limit approach enrichens the dataset and on the example of a specific commercial bank, in essence provides a good representation of Estonian corporate credit landscape in general.

The dataset has a total number of 12901 observations, consisting of 156 defaults and 12745 non-defaults. The observations are based on firms not on credit contracts, meaning that a firm can only have up to one observation per year (albeit it could have signed more than one credit contract)[8]. It must be noted that the defaulted observations are unique and each default observation occurs only once. In practice, in case a borrower has defaulted with any credit obligation, most commercial banks in Estonia have reserved themselves the right to prematurely terminate all the borrower's credit contracts. For instance, a situation where a firm has defaulted with some credit obligations, while continuing timely payments with remaining obligations, is very unlikely. Hence, the firm-based approach is reasonable. The non-defaulted observations are not unique however, as noted they are limited to appear only once per year in their population, i.e. a non-failed company can have a minimum of one and maximum of six observations in the dataset (one per each year during the observed period). Default is being used as an observed dependent variable, taking the form of either 1 for the defaulted or 0 for the non-defaulted firm. Default is the dependent variable in all prediction models.

3.2. Financial ratios

From annual financial statements of firms, only financial ratios have been used in this thesis, while Iwanicz-Drozdowska et al. (2016) also applied (transformed) financial figures (e.g. logarithm of total assets), changes in financial figures (e.g. sales growth rate) and changes in financial ratios. Financial ratios that are used in current study are in accordance with previous extensive literature reviews by Dimitras *et al.* (1996) and Bellovary *et al.* (2007), who showed that most used domains in corporate failure prediction are profitability, liquidity and solvency (solidity)[9]. It has to be taken into

[8] That is to eliminate possible bias by bigger companies who have signed more than one credit contract or have defaulted with multiple credit contracts.
[9] In the literature, one of the most used ratios is the total equity/total assets ratio, that also reflects capital structure – the smaller the ratio is, the more leveraged a firm is.

account that SMEs in Estonia are allowed to report their annual financial statements with only the most essential figures, meaning the reports are not as detailed as for large companies. Hence, it is reasonable to use the most common financial ratios previously used in failure research, in order to retain maximum number of observations with the necessary data available to calculate the ratios. More specifically, top eight ratios have been selected from various financial ratios from the recent study by Lukason and Andresson (2019), who used very similar context to current research: failure prediction of Estonian companies. Selected financial ratios and formulas can be seen below in Table 3.

Table 3. Financial ratios' domains, ratio abbreviations and formulas

Domain	Ratio abbreviation and formula
Liquidity	$CCLA = \frac{(cash - current\ liabilities)}{total\ assets}$
Liquidity	$NWCA = \frac{(current\ assets - current\ liabilities)}{total\ assets}$
Profitability	$NIA = \frac{net\ income}{total\ assets}$
Profitability	$NIOR = \frac{net\ income}{operating\ revenue}$
Financial structure / solvency	$DA = \frac{total\ debt}{total\ assets}$
Activity / efficiency	$ORA = \frac{operating\ revenue}{total\ assets}$
Interest burden / solvency	$FREOR = \frac{(financial\ revenue - financial\ expenses)}{operating\ revenue}$
Interest burden / solvency	$FREA = \frac{(financial\ revenue - financial\ expenses)}{total\ assets}$

Source: Formulas based on Lukason and Andresson (2019: 6), domain names based on synthesis of Lukason and Andresson (2019), Lukason and Laitinen (2016), du Jardin (2017).

In Estonia, companies are obliged to submit their annual reports to Estonian Business Register within six months after the end of the 12-month fiscal period. In accordance with explanation in chapter 3.1, in this study, for the healthy (non-defaulted) companies, all available annual reports from years 2012 to 2017 are in use. All financial ratios are calculated using previous year reports[10]. For the defaulted companies, a relevant annual financial statement prior to default is used. For example, if a firm whose fiscal year ends in December, defaulted in July 2014, then its 2013 annual report submission was due in

[10] Though period under scope in this study is 2013-2018, technically 2012 reports were included to analysis to calculate financial ratios for the year 2013.

June, thus 2013 report was used to calculate the ratios. Respectively, if default occurred in March 2016, then 2015 report's due date had not arrived, and therefore, 2014 report was used for calculations. However, the companies that have higher failure risk tend to delay submitting annual reports or not submit at all (Lukason and Camacho-Miñano 2019) resulting that the required up to date financial data is often unavailable for failing companies. To overcome this potential obstacle and retain as much data as possible, in case a relevant report is unavailable the study reverts to the latest available financial statement before the event of default, to calculate the ratios. Similar approach has been used in previous research, e.g. Back (2005). Firms that had no reports available were excluded from the dataset.

3.3. Variables about tax arrears and annual report delays

Non-financial variables in this thesis come from two main domains: a) tax arrears; and b) reporting delays. The collection of taxes in Estonia is administered by Estonian Tax and Customs Board, while information about unpaid tax debt is publicly available to everyone. Companies in Estonia must pay taxes twice a month (on the 10^{th} and 20^{th} dates in each month). In this study, all tax arrears are considered equal, so there is no classification between value added tax or employer-related taxes. Also, no distinction was made between timed and untimed tax arrears – even if tax arrears are timed, they do reflect a firm was not able to pay them in time. Tax arrears information is applied as a time series of twelve month ends, while for defaulted firms twelve month backwards starting from the pre-default month, and for non-defaulted firms twelve months matching the calendar year. To capture different aspects of the tax arrears, three variables are used in similar approach to Lukason and Andresson (2019) and Valgenberg (2020) studies: maximum (TMAX) and median (TMED) amount of tax arrears that are present on month ends (i.e. on the last day of each month) and also the number of months, where on the last day of each month tax arrears were present (TCOUNT). Based on the motivation in Lukason and Andresson (2019) only tax arrears that are present in month ends and are at least 100 euros are considered.

Reporting delays variable (RDD) was calculated in days that were overdue. Similarly, as with the financial ratios' calculations, for the healthy (non-defaulted) firms all available annual reports from years 2012 to 2017 are in use. RDDs are calculated using last year reports by deducting report due date (legal date that is six months after the end of fiscal year) from report submission date (actual date). Reports that were submitted before due date had negative values, which were considered as zero values (no delay). For the defaulted firms, the calculation principle is the same as with the non-defaulted firms, but only the reports prior to default events are used. If default happened within the first six months of running year t, then the report of fiscal year t-2 is used. If default happened after six months of running year t, then the report of fiscal year t-1 is used. In case the necessary report is missing (not submitted), then the latest available report is used[11].

It must be noted that natural logarithm of maximum and median values of tax arrears were used in order to reduce skewness. In case an observation's TMAX or TMED is zero, then zero value is used instead of natural logarithm. Overview of non-financial variables selection can be seen below in Table 4.

Table 4. Non-financial variables' abbreviations, formulas and explanations

Variable abbreviation and formula	Explanation
$TMAX = ln[\max(x_1 \dots x_{12})]$	Natural logarithm of maximum tax arrears over twelve month ends
$TMED = ln[median\ (x_1 \dots x_{12})]$	Natural logarithm of median tax arrears over twelve month ends
$TCOUNT = \sum_{k=1}^{12} TA_k$ $where\ TA_k = \{1\ if\ tax\ arrears \geq 100; else\ 0\}$	Number of months ending with tax arrears of 100 euros or more over twelve month ends
$RDD = report\ submitted - report\ due$	Reporting delays of annual reports (calculated in days)

Source: Lukason and Andresson (2019); Valgenberg (2020: 15).

Regarding annual reporting, in Estonia, information about non-submitted annual report is also publicly available to everyone via Estonian Business Register. If a company in Estonia fails to submit annual report in time, the register issues a warning of deletion from the register and obliges the company to submit the report within a specified extended term that is at least six months. In addition, the delaying firm can be fined. It has been found that delays over the legal deadline in submitting annual reports reflect higher risk of a firm's bankruptcy (Lukason and Camacho-Miñano 2019). If the company still fails to comply and presents no justified reason for non-compliance, the register may publish a

[11] In this case the delay is at least one year or more.

public notice concerning the company's failure to submit the annual report within the prescribed term and invite creditors to notify their claims against the company and to request the conduct of a liquidation proceeding within six months after the date of public notice (State Gazette…2020: Chapter 60). If no claims are presented, the company would be deleted from the register without liquidation proceeding (*Ibid*.). It has been found to be common for insolvent firms not to submit annual report at all (Lukason 2013). Therefore, in case a firm is using debt, a delayed annual report can *per se* hold valuable information for failure prediction.

3.4. Methods

Before commencing the analysis, the data was checked so that all observations had all the required financial and non-financial information present to calculate the necessary variables used in this study. Therefore, from initial dataset, 448 observations of non-defaults and 21 defaults were excluded because of missing data. In a recent study about bankruptcy prediction on a 10-year horizon, five different commonly used methods were compared against each other, whereby logistic regression and neural networks proved to be superior to other approaches (Altman *et al.* 2020). The two methods would presumably perform well also for bank loan default prediction.

First method used in this thesis is logistic regression (LR), which has been one of the most used methods in earlier relevant failure prediction studies, as well as one of the most practiced methods by banks in their corporate default prediction modelling (Altman and Sabato 2007; Ciampi 2015). Since this study aims to show if and how non-financial variables enhance bank loan default prediction accuracy then the use of classical logistic regression is suitable for that purpose. In context of an Estonian commercial bank, whole population of firms (defaulted and non-defaulted) is used, and thus the observations are very imbalanced. The share of defaulted observations is 1.21%. Using LR method, this would usually cause misclassification errors for the minority group (defaulted firms), so in order to compensate this, a weighting technique was used to equalize the two groups, as is the common practice in previous failure studies (see for example Altman *et al.* 2017; Calabrese and Osmetti 2013). The formulas to calculate weights for both groups are presented in Table 5.

Table 5. Formulas to calculate weights

Weight for defaulted group	Weight for non-defaulted group
$W_d = \dfrac{0.5}{share\ of\ defaulted\ firms}$	$W_{nd} = \dfrac{0.5}{share\ of\ non\text{-}defaulted\ firms}$

Source: compiled by the author.

First, univariate prediction ability of financial and non-financial variables was tested individually. Next, all the variables were categorized into three domains to perform a domain-based multivariate analysis: a) financial ratios; b) annual reporting delays; c) tax arrears. Finally, all variables from the three domains were combined to create the final model. Correlation matrix of used independent variables is located in Appendix 2.

The second method used in current study is a multilayer perceptron (in short, MLP) that is a class of feedforward neural network methodology. MLP is a modern machine learning tool in failure prediction research and thus, is hereby used to verify the initial results obtained by the classical LR method. For the MLP, instead of weighting the two groups, a synthetic minority oversampling technique (SMOTE) was used to equalize the imbalances between the two groups of companies. That is achieved by multiplying defaulted observations to match the non-defaulted observations – 12745 for each group. It must be noted that univariate prediction abilities of variables were calculated using only the logistic regression method – that is because in essence, the other method used in the study, the multilayer perceptron, requires a multivariate setting to perform adequately.

4. Results and discussion

Loan payment default is a rare dependent variable in failure prediction research. There are only a few prior studies that have used payment default as a dependent variable in similar context, however to the best of the author's knowledge, none have used the non-financial variables present in current thesis for predicting loan payment defaults. It means that several findings that are presented in this chapter are unique, i.e. direct comparison with previous studies is not possible.

Descriptive statistics of financial variables can be seen in Table 6.

Table 6. Descriptive statistics of financial variables

		CCLA	NWCA	NIA	NIOR	DA	ORA	FREOR	FREA
Non-defaulted	N	12745							
	Mean	-0.13	0.24	0.12	0.12	0.44	1.92	-0.017	0.002
	Median	-0.14	0.20	0.08	0.06	0.43	1.48	0.002	0.003
	Minimum	-3.00	-3.00	-3.00	-3.00	0.00	0.00	-3.000	-3.000
	Maximum	1.00	1.00	3.00	3.00	3.00	10.00	3.000	2.492
	Std. deviation	0.34	0.34	0.23	0.40	0.27	1.75	0.302	0.068
Defaulted	N	156							
	Mean	-0.26	0.17	0.02	-0.01	0.61	1.86	0.003	0.009
	Median	-0.26	0.08	0.03	0.02	0.61	1.28	0.005	0.010
	Minimum	-1.93	-0.97	-3.00	-2.27	0.00	0.04	-1.205	-0.201
	Maximum	1.00	1.00	0.95	0.70	2.40	10.00	0.651	0.219
	Std. deviation	0.38	0.38	0.40	0.36	0.35	1.85	0.130	0.037
p-value of ANOVA Welch test		0.000	0.019	0.002	0.000	0.000	0.693	0.070	0.025

Source: compiled by the author.

All variables whose p-value of Welch's ANOVA test is ≤ 0.05 are considered statistically significant. The defaulted group has lower mean and median values for liquidity, profitability and efficiency domains' variables and higher mean and median values for solvency domains' variables. Strong differences of minimum and maximum values combined with large standard deviations inside both groups' results indicate that there is no clear reason or path to failure. The presence of different failure processes is reasoned with the use of whole population dataset from an Estonian commercial bank, which in essence makes the dataset heterogenous. The result is also in line with previous failure research, where large datasets were used, for example Lukason and Laitinen (2016). The p-values are statistically significant for all variables, except ORA and FREOR (0.693 and 0.07 respectively). Therefore, ORA variable reflects that in terms of a firm's efficiency, the means of the defaulted and non-defaulted groups are equal. In prior literature, in their default-based Italian study covering years 1999-2002, Bottazzi *et al.* (2011) found that productive efficiency[12] reduced the risk of default, however its importance decreased over

[12] Productive efficiency was measured in terms of value added per employee.

time and was insignificant in the last year before the default – in principle giving the same result as current finding, since present study uses data only one year prior to default.

Descriptive statistics for non-financial variables can be seen in Table 7, where all four variables exhibit large differences of mean values between the defaulted and non-defaulted groups.

Table 7. Descriptive statistics of non-financial variables

<table>
<tr><th colspan="2"></th><th>TMAX</th><th>TCOUNT</th><th>TMED</th><th>RDD</th></tr>
<tr><td rowspan="6">Non-defaulted</td><td>N</td><td colspan="4">12745</td></tr>
<tr><td>Mean</td><td>1.20</td><td>0.57</td><td>0.33</td><td>12.87</td></tr>
<tr><td>Median</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr>
<tr><td>Minimum</td><td>-4.61</td><td>0.00</td><td>-4.20</td><td>0.00</td></tr>
<tr><td>Maximum</td><td>13.59</td><td>12.00</td><td>12.93</td><td>753.00</td></tr>
<tr><td>Std. Deviation</td><td>2.99</td><td>1.89</td><td>1.61</td><td>45.51</td></tr>
<tr><td rowspan="6">Defaulted</td><td>N</td><td colspan="4">156</td></tr>
<tr><td>Mean</td><td>6.90</td><td>5.63</td><td>3.72</td><td>67.78</td></tr>
<tr><td>Median</td><td>8.02</td><td>5.00</td><td>0.00</td><td>0.00</td></tr>
<tr><td>Minimum</td><td>-0.80</td><td>0.00</td><td>-2.53</td><td>0.00</td></tr>
<tr><td>Maximum</td><td>12.03</td><td>12.00</td><td>12.03</td><td>548.00</td></tr>
<tr><td>Std. Deviation</td><td>3.57</td><td>4.55</td><td>4.23</td><td>130.68</td></tr>
<tr><td colspan="2">p-value of ANOVA Welch test</td><td>0.000</td><td>0.000</td><td>0.000</td><td>0.000</td></tr>
</table>

Source: compiled by the author.

It is clearly visible that the defaulted companies have serious issues with tax arrears, while the non-defaulted have almost none. The mean value of TCOUNT for the defaulted firms is 5.6, which is well above the three times threshold – the point, when reached or exceeded, where in his similar payment default setting, Back (2005) discovered an important increase in probability of permanent payment default (*Ibid*.: 861). Back (*Ibid*) used 2.5-year horizon, hence the finding of TCOUNT 5.6 in current study with only one-year horizon, is remarkable. In their bankruptcy based Estonian study, Lukason and Andresson (2019) relevant result for one-year horizon was 7.4. It can be assumed that defaulted firms either try to survive by aggressively evading tax obligations in favour of other creditors (for example, banks and key suppliers), or the firm has been left dormant because of not having perspective of continuing activities, and thus, the unpaid obligations accumulate further (until official insolvency proceedings).

It was confirmed that the mean value of reporting delays was 55 days more for the defaulted firms compared to the non-defaulted firms, though median values for both groups were zero. It shows that a minority of firms with defaults had problems with timely reporting. Additionally, as the minimum and median values for the defaulted group are zero, it explains that in general, delays in reporting would not directly indicate increased risk of payment default. Lukason (2013) found that non-submission of reports in Estonia varies for different insolvency types and was more frequent in bankruptcy proceeding abatement – situation where managers try to hide information and suggest there are not enough resources left to finance bankruptcy proceeding, so the firm would be deleted from the registry. Thus, in this context, the RDD variable could indicate that, as explained in chapter 2.1, a firm that has defaulted, might not end up insolvent – so the incentive to hide financial information is not present.

Next, the univariate prediction abilities of the applied variables are presented in Table 8. In terms of financial variables, the table shows that the best prediction capability comes from solvency variables (DA and FREA). Closely followed by liquidity (CCLA) and profitability (NIA and NIOR) variables. However, almost all non-financial variables outperform every financial variable. Only reporting delays show slightly smaller prediction accuracy than the best performing financial variables. Tax arrears have clearly the highest univariate prediction accuracy, specifically the maximum tax arrears variable (TMAX) with 84%. Therefore, the companies that hold big tax arrears are most likely to default. Lukason and Andresson (2019) arrived exactly at the same conclusion in their different bankruptcy-oriented setting, where also maximum tax arrears variable had the best univariate failure (bankruptcy) prediction accuracy (85.9%).

Table 8. Univariate prediction accuracies of variables (%)

Financial variables				**Non-financial variables**	
CCLA	58.6	DA	61.2	RDD	59.6
NWCA	53.6	ORA	51.3	TMAX	84.0
NIA	57.9	FREOR	50.8	TMED	71.3
NIOR	57.8	FREA	60.0	TCOUNT	78.4

Source: compiled by the author.

For the next step, using LR and MLP methods, the domain-based failure prediction capabilities are tested. The prediction accuracies for all three domains can be seen in Table 9.

Table 9. Domain-based multivariate prediction accuracies (%)

Method	**Logistic regression (LR)**			**Multilayer perceptron (MLP)**		
Domains used	Defaulted	Non-defaulted	Overall	Defaulted	Non-defaulted	Overall
Financial ratios	59	64.8	61.9	71.7	60.2	65.9
Reporting delays	29.5	89.7	59.6	28.9	90.2	59.4
Tax arrears	80.1	85.9	83	80.9	86.2	83.5
All combined	80.8	86.7	83.7	88.6	89.7	89.1

Source: compiled by the author.

Going forward, the study focuses and describes only the highest prediction results achieved by using either of the two methods. Financial ratios show modest prediction accuracies[13], whereby 65.9% of overall accuracy was reached. In a comparable setting, Back (2005) achieved 72%[14]. It also confirms an important finding of prior failure literature: several previous studies, e.g. Altman and Sabato (2007) and Ciampi (2015), have showed that failure prediction models that are based on financial variables and perform well on large public firms, tend to show poor prediction accuracies for SMEs. Additionally, financial reports often fail to indicate problems in the company's financial situation (Lukason and Laitinen 2019) and since empirical evidence show that large companies are in essence more solvent than SMEs, the financial ratios of SMEs and large companies cannot be directly compared (Beaver 1966). This is why non-financial information is being implemented in today's failure studies. This summarizes one of the main findings of current thesis – financial variables in the form of financial ratios are not suitable to predict loan payment defaults.

The reporting delay variable shows almost no prediction capability for the defaulted group with only 29.5%. This variable is not suitable for payment default prediction. Albeit for example, in their study Lukason and Camacho-Miñano (2019) showed that delays in

[13] Between the two methods, MLP classified defaulted firms better (71.7%), while LR did so with the non-defaulted (64.8%).
[14] Back (2005) used multinomial logistic regression with only two financial ratios: return on investment and debt ratio.

reporting indeed do show increased risk of failure. Therefore, the RDD's poor prediction capability in terms of loan defaults is slightly unanticipated.

The tax arrears domain combining TMAX, TMEDIAN and TCOUNT variables strongly outperforms all other domains. Prediction accuracy for the defaulted group is 80.9% and 83.5% overall. The result confirms initial findings in univariate prediction accuracies and is comparable with Back's study using similar dependent variable, where 86.3% default prediction accuracy was achieved using the only non-financial variables model (Back 2005: 859). Lukason and Andresson (2019) reached an even higher result using bankruptcy as dependent variable and the same independent variable as in current study (tax arrears 12 months prior to event date) – obtaining 89.5% bankruptcy prediction accuracy.

Finally, a multivariate model was constructed that included all the aforementioned three domains. Prediction accuracy was 88.6% for the defaulted group and 89.1% overall. As can be seen in Table 9, in terms of methods used in the study, the modern machine learning method (MLP) outperformed the classical statistical method (LR). For the MLP, the most important variable predictors of default were TCOUNT (100% normalized importance rate), FREA (98.2%) and TMED (75.6%)[15]. Due to high multicollinearity between the variables[16], logistic regression models are not presented, since the variables' estimations would be biased. Area under the curve value was 0.951.

The main theoretical implication of current thesis is that tax arrears offer high predictive performance to loan payment defaults. That is an important conclusion since there is no prior literature where tax arrears are used to predict loan defaults. Tax arrears outperform the most common financial ratios previously used in failure prediction literature. It is important to note that one year prior to bankruptcy, as showed by Lukason and Andresson (2019), financial ratios had 79.5% prediction accuracy; while in the context of defaults in this thesis, the accuracy was only 65.9%. This indicates that companies default rather unexpectedly – it would not be seen coming from the companies' financial reports. A potential explanation for this finding was given by Laitinen and Lukason (2014: 827), who discovered that Estonian firms generally lacked financial flexibility to withstand

[15] Overview of MLP method's independent variable importance is located in Appendix 1.
[16] Correlation matrix is located in Appendix 2.

external shocks and other specific external events. Present study’s conclusion in general confirms previous research whereby inclusion of non-financial information does greatly enhance failure prediction accuracy.

The main practical implication of this study is that from a creditor’s point of view, earlier payment disturbances, namely tax arrears, are a clear sign of increased default risk. This information should be considered by the banks when granting credit to borrowers and also in the context of existing loan portfolio – to take necessary measures in a timely manner, in order to minimize potential losses. Evidently the larger the tax arrears and the more frequent they are, the higher the risk of payment default. Since tax arrears are publicly available on daily or monthly basis in many countries, it offers high practical value to creditors. It should help in decision making when financial reporting is delayed or opaque that is especially inherent to companies with increased failure risk. As of today, established financial institutions have already implemented previous payment behaviour component in their credit scoring models, albeit mostly in the form of data about disturbances originating from their own organization. Present study confirms why it is essential to include tax arrears information into credit scoring models in the context of Estonia.

5. Conclusion

The study aimed to compare the accuracy of financial ratios, tax arrears and annual report submission delays in bank loan default prediction. For the analysis, logistic regression and multilayer perceptron methods were applied to a dataset consisting of defaulted and non-defaulted companies originating from an Estonian commercial bank.

The results showed that by including non-financial variables the accuracy of loan default prediction increases remarkably. The study provided several implications. As for the theoretical implication, it was discovered that tax arrears offer excellent prediction capability to loan defaults. At the same time, even though prior research has found that occurrences of reporting delays can effectively indicate an increased failure risk, the phenomenon does not suit to predict loan defaults. As for the practical implication, the findings should help lenders to consider the role of previous payment history, in the form of tax arrears, to loan defaults. It would be rational to implement this sort of information (which is publicly available in some countries) in a credit scoring methodology and also in an early warning system – to get indications of increased default risk in time. This would enable the lender to take timely action to minimize potential financial losses.

This study has filled a gap in modern firm failure research by analysing tax arrears in the context of loan payment defaults. Future research could elaborate the prior payment concept to test how payment history inside the credit institution itself would compare with or accompany tax arrears to predict failure. Additionally, information about other types of defaults could increase the prediction accuracies.

References

1. Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94, 164–184. https://doi.org/10.1016/j.eswa.2017.10.040
2. Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589–609. https://doi.org/10.1111/jofi.12742
3. Altman, E., Sabato, G., Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 1–33. https://doi.org/10.21314/JCR.2010.110
4. Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management and Accounting*, *28*(2), 131–171. https://doi.org/10.1111/jifm.12053
5. Altman, E. I., Narayanan, P. (1997). An International Survey of Business Failure Classification Models. *Financial Markets, Institutions and Instruments*, 6(2), 1–57. https://doi.org/10.1111/1468-0416.00010
6. Altman, E. I., Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus*, 43(3), 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x
7. Altman, E. I, Iwanicz-drozdowska, M., Laitinen, E. K. (2020). A Race for Long Horizon Bankruptcy Prediction. *Applied Economics*, 52(37), 4092-4111. https://doi.org/10.1080/00036846.2020.1730762
8. Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12(4), 929–935. https://doi.org/10.1109/72.935101
9. Back, P. (2005). Explaining financial difficulties based on previous payment behavior, management background variables and financial ratios. *European Accounting Review*, *14*(4), 839–868. https://doi.org/10.1080/09638180500141339
10. Balcaen, S., Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *British Accounting Review*, 38(1), 63–93. https://doi.org/10.1016/j.bar.2005.09.001
11. Basel Committee on Banking Supervision. (2016). *Prudential treatment of problem assets - definitions of non-performing exposures and forbearance*. 1-45.

https://www.bis.org/bcbs/publ/d403.htm
12. Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, 71–111. https://doi.org/10.2307/2490171
13. Bellovary, J.L., Giacomino, D., Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 33, 1-42. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=892160
14. Bhimani, A., Gulamhussen, M. A., Lopes, S. D. R. (2010). Accounting and non-accounting determinants of default: An analysis of privately-held firms. *Journal of Accounting and Public Policy*, 29(6), 517–532. https://doi.org/10.1016/j.jaccpubpol.2010.09.009
15. Bhimani, A., Gulamhussen, M. A., Lopes, S. D. R. (2013). The Role of Financial, Macroeconomic, and Non-financial Information in Bank Loan Default Timing Prediction. *European Accounting Review*, 22(4), 739–763. https://doi.org/10.1080/09638180.2013.770967
16. Bottazzi, G., Grazzi, M., Secchi, A., Tamagni, F. (2011). Financial and economic determinants of firm default. *Journal of Evolutionary Economics*, 21(3), 373–406. https://doi.org/10.1007/s00191-011-0224-6
17. Calabrese, R., Osmetti, S. A. (2013). Modelling small and medium enterprise loan defaults as rare events: The generalized extreme value regression model. *Journal of Applied Statistics*, 40(6), 1172–1188. https://doi.org/10.1080/02664763.2013.784894
18. Camacho-Miñano, M. D. M., Segovia-Vargas, M. J., Pascual-Ezama, D. (2015). Which Characteristics Predict the Survival of Insolvent Firms? An SME Reorganization Prediction Model. *Journal of Small Business Management*, 53(2), 340–354. https://doi.org/10.1111/jsbm.12076
19. Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal of Business Research*, 68(5), 1012–1025. https://doi.org/10.1016/j.jbusres.2014.10.003
20. Ciampi, F. (2018). Using corporate social responsibility orientation characteristics for small enterprise default prediction. *WSEAS Transactions on Business and Economics*, 15, 113–127. https://www.wseas.org/multimedia/journals/economics/2018/a265907-592.php
21. Ciampi, F., Cillo, V., Fiano, F. (2020). Combining Kohonen maps and prior payment behavior for small enterprise default prediction. *Small Business Economics*, 54(4), 1007–1039. https://doi.org/10.1007/s11187-018-0117-2
22. Dias, A., Teixeira, A. A. C. (2017). The anatomy of business failure: A qualitative

account of its implications for future business success. *European Journal of Management and Business Economics*, 26(1), 2–20. https://doi.org/10.1108/EJMBE-07-2017-001

23. Dimitras, A. I., Zanakis, S. H., Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, 90(3), 487–513. https://doi.org/10.1016/0377-2217(95)00070-4
24. du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, 75, 25–43. https://doi.org/10.1016/j.eswa.2017.01.016
25. European Commission. (2019). *Entrepreneurship and Small and medium-sized enterprises (SMEs)*. https://ec.europa.eu/growth/smes_en
26. European Commission. (2020). *What is an SME?* https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en
27. Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*, 41(14), 6433–6445. https://doi.org/10.1016/j.eswa.2014.04.026
28. Headd, B. (2003). Redefining Business Success: Distinguishing between Closure and Failure. *Small Business Economics*, 21(1), 51–61. https://doi.org/10.1023/A:1024433630958
29. Höglund, H. (2017). Tax payment default prediction using genetic algorithm-based variable selection. *Expert Systems with Applications*, 88, 368–375. https://doi.org/10.1016/j.eswa.2017.07.027
30. Iwanicz-Drozdowska, M., Laitinen, E., Suvas, A., Altman, E. I. (2016). Financial and nonfinancial variables as long-horizon predictors of bankruptcy. *The Journal of Credit Risk*, 12(4), 49–78. https://doi.org/10.21314/JCR.2016.216
31. Laitinen, E. K. (1999). Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International Review of Financial Analysis*, 8(2), 97–121. https://doi.org/10.1016/S1057-5219(99)00012-5
32. Laitinen, E. K. (2008). Data system for assessing probability of failure in SME reorganization. *Industrial Management & Data Systems*, 108(7), 849–866. https://doi.org/10.1108/02635570810897964
33. Laitinen, E. K. (2011). Assessing viability of Finnish reorganization and bankruptcy firms. *European Journal of Law and Economics*, 31(2), 167–198. https://doi.org/10.1007/s10657-009-9136-4
34. Laitinen, E. K., Laitinen, T. (2009). Audit Report in Payment Default Prediction: A Contingency Approach. *International Journal of Auditing*, 13(3), 259–280. https://doi.org/10.1111/j.1099-1123.2009.00396.x

35. Laitinen, E. K., Lukason, O. (2014). Do firm failure processes differ across countries: evidence from Finland and Estonia. *Journal of Business Economics and Management*, 15(5), 810–832. https://doi.org/10.3846/16111699.2013.791635
36. Lukason, O. (2013). Firm bankruptcies and violations of law: An analysis of different offences. *(Dis)Honesty in Management*, 10, 127–146. https://doi.org/10.1108/S1877-6361(2013)0000010010
37. Lukason, O., Andresson, A. (2019). Tax Arrears Versus Financial Ratios in Bankruptcy Prediction. *Journal of Risk and Financial Management*, 12(4), 1-13. https://doi.org/10.3390/jrfm12040187
38. Lukason, O., Camacho-Miñano, M. D. M. (2019). Bankruptcy Risk, Its Financial Determinants and Reporting Delays: Do Managers Have Anything to Hide? *Risks*, 7(3), 1-15. https://doi.org/10.3390/risks7030077
39. Lukason, O., Laitinen, E. K. (2016). Failure processes of old manufacturing firms in different European countries. *Investment Management and Financial Innovations*, 13(2), 310–321. https://doi.org/10.21511/imfi.13(2-2).2016.06
40. Lukason, O., Laitinen, E. K. (2019). Firm failure processes and components of failure risk: An analysis of European bankrupt firms. *Journal of Business Research*, 98, 380–390. https://doi.org/10.1016/j.jbusres.2018.06.025
41. Lukason, O., Urbanik, A. (2013). Why Reorganization of Firms Fails: Evidence From Estonia. *Estonian Discussions on Economic Policy*, 21(1), 73-84. https://doi.org/10.15157/tpep.v21i1.1065
42. McCann, F., McIndoe-Calder, T. (2015). Firm size, credit scoring accuracy and banks' production of soft information. *Applied Economics*, 47(33), 3594–3611. https://doi.org/10.1080/00036846.2015.1019034
43. Merwin, C. L. (1942). *Financing small corporations in five manufacturing industries*, 1926–1936. https://www.nber.org/books/merw42-1
44. State Gazette. (2020). *Commercial Code*. https://www.riigiteataja.ee/en/eli/522062017003/consolide
45. Statistics Estonia. (2020). GDP chain-linked volume change (seasonally and working day adjusted). https://www.stat.ee/68589
46. Sun, J., Li, H., Huang, Q. H., He, K. Y. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, 57, 41–56. https://doi.org/10.1016/j.knosys.2013.12.006
47. Süsi, V., Lukason, O. (2019). Corporate governance and failure risk: evidence from

Estonian SME population. *Management Research Review*, 42(6), 703–720. https://doi.org/10.1108/MRR-03-2018-0105

48. Valgenberg, G. (2020). Forecasting corporate permanent insolvency. Master's thesis. *School of Economics and Business Administration, University of Tartu,* 1-29. https://dspace.ut.ee/handle/10062/67044
49. Weitzel, W., Jonsson, E. (1989). Decline in Organizations : A Literature Integration and Extension. *Administrative Science Quarterly*, 34(1), 91-109. https://doi.org/10.2307/2392987
50. Wu, W. W. (2010). Beyond business failure prediction. *Expert Systems with Applications*, 37(3), 2371–2376. https://doi.org/10.1016/j.eswa.2009.07.056
51. Xu, M., Zhang, C. (2009). Bankruptcy prediction: The case of Japanese listed companies. *Review of Accounting Studies*, 14(4), 534–558. https://doi.org/10.1007/s11142-008-9080-5

Appendix 1. Overview of MLP method's independent variable importance

Independent variable importance		
	Importance	Normalized Importance
TMAX	0.070	51.8%
TCOUNT	0.136	100.0%
TMEDIAN	0.103	75.6%
CCLA	0.078	57.1%
NWCA	0.065	47.6%
NIA	0.081	59.8%
NIOR	0.077	56.7%
DA	0.070	51.7%
ORA	0.062	46.0%
FREOR	0.073	53.7%
FREA	0.133	98.2%
RDD	0.051	37.8%

Source: compiled by the author.

Appendix 2. Correlation matrix of used independent variables

Variable	TMAX	TCOUNT	TMEDIAN	CCLA	NWCA	NIA	NIOR	DA	ORA	FREOR	FREA	RDD
TMAX	1	.774**	.572**	-.143**	-.115**	-.061**	-.048**	.125**	0.003	.024**	.021*	.147**
TCOUNT	.774**	1	.903**	-.130**	-.095**	-.059**	-.048**	.117**	0.006	.023**	.030**	.152**
TMEDIAN	.572**	.903**	1	-.112**	-.079**	-.050**	-.037**	.102**	0.005	0.017	.027**	.136**
CCLA	-.143**	-.130**	-.112**	1	.660**	.345**	.168**	-.701**	-.156**	-.045**	-.049**	-.024**
NWCA	-.115**	-.095**	-.079**	.660**	1	.347**	.103**	-.622**	.109**	-.039**	-.038**	-0.016
NIA	-.061**	-.059**	-.050**	.345**	.347**	1	.422**	-.354**	.059**	-.110**	-.200**	0.011
NIOR	-.048**	-.048**	-.037**	.168**	.103**	.422**	1	-.205**	-.151**	-.677**	-.354**	0.012
DA	.125**	.117**	.102**	-.701**	-.622**	-.354**	-.205**	1	.158**	.109**	.093**	.021*
ORA	0.003	0.006	0.005	-.156**	.109**	.059**	-.151**	.158**	1	.066**	.045**	-.036**
FREOR	.024**	.023**	0.017	-.045**	-.039**	-.110**	-.677**	.109**	.066**	1	.504**	-0.004
FREA	.021*	.030**	.027**	-.049**	-.038**	-.200**	-.354**	.093**	.045**	.504**	1	0.014
RDD	.147**	.152**	.136**	-.024**	-0.016	0.011	0.012	.021*	-.036**	-0.004	0.014	1

Source: compiled by the author; Note: *. Correlation is significant at the 0.05 level (2-tailed). **. Correlation is significant at the 0.01 level (2-tailed).

Non-exclusive licence to reproduce thesis and make thesis public

I, Keijo Kohv

herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

PREDICTION OF BANK LOAN DEFAULTS WITH FINANCIAL RATIOS, TAX ARREARS AND ANNUAL REPORT DELAYS

supervised by Oliver Lukason.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Keijo Kohv
10.08.2020