UNIVERSITY OF TARTU

Faculty of Social Sciences

School of Economics and Business Administration

Anne-Liis Tamm

# PAST ENTREPRENEURIAL MISBEHAVIOUR AS A PREDICTOR OF FIRM DEFAULT

Master's thesis

Supervisor: Associate Professor Oliver Lukason (PhD)

Tartu 2021

I have written this Master Thesis independently. Any ideas or data taken from other authors or other sources have been fully referenced.

**Abstract**

The aim of this thesis is to evaluate the usefulness of previous entrepreneurial misbehaviour to predict firm default. As for previous entrepreneurial misbehaviour, two non-financial variables are observed, namely annual report delays and tax declaration delays. The study is based on the whole population of VAT obligatory firms from Estonia, including 44 172 observations of defaulted and non-defaulted firms (mainly SMEs). The method used in this study is logistic regression and both of the domains with observed variables showed high prediction capabilities for business failure, namely the accuracy of the best model was over 81% for defaulted and over 87% for non-defaulted firms (reaching 84.7% as the summarized prediction accuracy). The novelty of this thesis is that non-financial variables are not so widely applied in the business failure prediction field and as financial statements may not always be the best indicators for possible default (especially for new firms), novel non-financial variables should be introduced to failure prediction models. Considering the fact that past entrepreneurial behaviour tends to repeat, it is critical to monitor delaying behaviour to potentially foresee future default as early as possible.

Keywords: failure prediction, SMEs, non-financial variables, previous entrepreneurial behaviour, reporting delays

CERCS: S180, S181, S190

## 1. Introduction

Prediction of business failure is an emerging field of research. The area itself is intriguing and useful - if the potential success or failure of firms can be predicted early and with high accuracy, it may help to prevent different risks for investors and banks, and also lower the impact for economical risks for several stakeholders (firm owners, partners, clients, etc) and for the economical environment at large (Alaka et al., 2018). In this area, the "event" that is usually researched and predicted, is called business failure. Business failure prediction has developed into an important research domain within corporate finance in the last 35 years (Pretorius, 2009). The term itself has several definitions - most commonly "business failure" is defined as default, closure, end of operations, termination, distress, performance decline, bankruptcy (Pretorius, 2009) or (financial) distress. Sun et al. (2014) use the term financial distress prediction (FDP) and they define, that FDP is sometimes called financial failure discrimination, bankruptcy prediction, business failure prediction or corporate failure prediction. Based on their definition, FDP predicts whether a company will fall into financial distress based on the current financial data (Sun et al., 2014).

Predicting business failure focuses on creating different prediction models and to be more precise, most commonly bankruptcy prediction models. Although there are numerous types of prediction models, they mainly use financial variables as an input (Balcaen & Ooghe, 2006). The popularity of financial variables was also proven by Prusak (2018), where selected Central and Eastern Europe countries' bankruptcy prediction models were observed. These models can give a good overview of the current financial situation of the existing firm, but most of them don't take non-financial variables into account. There are even fewer (or hardly any) studies and models that measure and use the behaviour of the managers of the firms as a variable to predict the failure of firms.

It becomes even more complex when the size of the firms has to be considered - when predicting the failure of small and medium-sized enterprises (SMEs) some size-specific peculiarities have to be taken into account. For example, financial ratios can decline so quickly (Ciampi et al., 2020) without any previous negative signals and therefore the traditional prediction models may not be suitable for them or at least they are not so accurate for SMEs compared to big companies because of their financial characteristics (Ciampi, 2015). As small firms are vulnerable to (unexpected) external causes which may lead to bankruptcy quickly (Lukason & Hoffman, 2015), it is important to discuss what else can be considered when predicting possible business failure for SMEs.

It is noted that the characteristics of management or the entrepreneur have a great impact on the performance of a company (Ooghe & Prijcker, 2008). There are several ways to measure an entrepreneur`s (in)competence and one way is to take a look at the business history - for example, if a person has been involved in a bankruptcy case (Tobback et al., 2017). As entrepreneurial behaviour tends to repeat (Lattacher & Wdowiak, 2020), it would be useful to be aware of, if and how non-financial variables (such as human behaviour) can be used in prediction models. Lukason and Camacho-Miñano (2019) have proven a strong link between one behavioural event - delayed annual reporting submissions - and a high risk of bankruptcy. The same is supported by Altman et al. (2010) - companies with a high risk of bankruptcy are more likely to delay their financial reporting. In addition, Lukason and Camacho-Miñano (2021) have also explained annual report delays with the same event from the past - managers, who have delayed with their reports in the past, will most likely delay again.

Estonia with its tax system and legal framework is a good economical area for such studies - information about companies, company managers and their previous entrepreneurial behaviour is available and can be used for research purposes and the available information is accurate and not distorted. This thesis aims to evaluate the usefulness of previous entrepreneurial misbehaviour to predict firm default. As for entrepreneurial misbehaviour, two variables from the "legal point of view" (obedience to the law) are observed - annual report delays and tax declaration report delays (VAT reports and income- and social tax reports). Using a logistic regression method, this thesis analyses if any of the variables are suitable and useful to use in failure prediction models and which ones are the most accurate. As a comparison, some more "traditional" financial variables are also added to the thesis.

This thesis has three main parts. The first part of the thesis gives an overview of existing literature - what are the most popular methods and variables used to predict failure of firms and how often non-financial variables are used in those researches. The second part is empirical and gives an overview and explanation of variables, dataset and method that is used. The third part is the results with discussion and the last part is conclusions.

# 2. Review of literature

## 2.1 Failure prediction models

As the term business failure has several definitions, different terms are used in the research field to describe business failure, prediction models. The breakthrough of the overall bankruptcy prediction model was done by Altman in 1968, where his research was focused on the prediction of corporate financial distress and after that, in most cases, authors use bankruptcy as the dicing line for distinguishing the failed and non-failed firms (Altman, 1968). When predicting business failure, several variables are used and based on Prusak (2018) study about bankruptcy prediction, they can be distributed to financial and non-financial variables, where financial variables are calculated by using data from financial statements and non-financial variables are therefore all others.

When observing variables in failure prediction models, Balcaen and Ooghe (2006) concluded in their study that there is no prime methods or predictor variables, and although firm failure prediction is a highly researched area, Veganzones and Severin (2020) researched 106 published articles and they found that the most commonly used metrics are related to measurable financial ratios and market variables. Bellovary et al. (2007) also analysed 165 bankruptcy prediction models published from 1965 to 2007 and concluded that a total of 752 different factors are used in studies and the factors considered in any study ranges from 1 to 57, where the most common variable used in multiple studies were the ratio of Net Income to Total Assets (Return on Assets), and Current Assets to Current Liabilities (Current Ratio). Höglund (2017) also showed that solvency (total debt / total assets) and liquidity (current ratio / total assets) seem to be a really good indicator to predict financial distress (to be precise - tax defaults was observed in Höglund`s work). The number of variables used in models also varies, but Bellovary et al. (2007) stated that a greater number of factors do not guarantee higher model accuracy - some models with only two variables are just as good as models with over 20 factors. The timeframe when a variable is reasonable to use is also important - some research shows that for traditional failure models, financial variables are good for a one-year horizon, but their predictive ability decreases quite fast after that (du Jardin and Severin, 2011; du Jardin, 2015).

Even though financial variables are calculated from a firm`s financial statements, there are several issues with failure prediction models- for example, for models that only use financial variables, asymmetry of data and availability of data in a timely manner is a considerable weakness (Kohv & Lukason, 2021). Furthermore, financial figures may not always show the real view of a company's financial situation - studies show that for failing companies, financial data may be manipulated to hide or postpone financial weaknesses or the emergence of it (Laitinen & Laitinen, 2009; Ciampi, 2018). The idea is also supported by Ajinkya et al. (2005), where the study showed that there is a strong link between financial distress (firms with losses) and information disclosure - meaning that firms with losses will probably show their information compared to others. In cases where a firm isn't doing well, managers may try to obfuscate failures and try to hide it with reporting delays (Lukason and Camacho-Miñano, 2019). Also, Lukason and Laitinen (2019) have shown that even the latest annual reports cannot be trusted 100% when indicating the deterioration of a firm, because they usually become available months after the end of the financial year and on average the last financial statement that depicts the failure of an SME is one year before firms bankruptcy.

Another aspect that has to be considered is the size of firms - most of the active companies are SMEs but studies have shown that for SMEs small changes in financial statements has a huge impact on financial ratios (Ciampi, 2015). Based on Lukason and Andresson (2019) study with SMEs, it can be said that some variables are not even effective for SMEs - some commonly used financial ratios, that are suitable for prediction models for big companies, are inefficient for SMEs (Ciampi, 2015; Ciampi et al., 2021). In addition, as SMEs' bankruptcies often happen because of external causes (Lukason and Hoffman, 2015) and their payment default and recession happens really quick (Ciampi et al., 2020), then financial variables are not always 100% trustworthy or using them in a timely manner may be complicated (financial statements are usually published once a year and when a problem appears between two reports, then the last one may not show any signs of default). Therefore, other variables should be considered when predicting business failure - for example, non-financial variables.

**2.2 Non-financial variables in firm failure prediction models**

There aren't many long-term studies that focus on creating credit risk models especially for SMEs and there are even fewer studies that have included non-financial information as a predictor of failure (Altman et al., 2010). The lack of research about SMEs and non-financial variables is also proven by Prusak (2018). Despite that, non-financial variables as a research field seem to become more and more popular over the last years and new variables are taken into usage to predict the failure of firms. The importance of SME failure prediction was also discussed by Ciampi et al. (2021), where the study proposed several changes to failure prediction models specialized for SMEs and one of the main proposals was enlarging the set of qualitative variables. But, leaving aside the size of firms, the domain of non-financial variables is not entirely unexplored.

For example Back (2005) used different non-financial variables in their research and the following variables were included: Character of management (which measured (1) whether the management is active within firms with recorded payment disturbances and (2) private financial situations of individual members of the management), Prior payment behaviour (which measured (1) number of publicly registered payment disturbances during the year prior to the time when the data was composed, and (2) number of payment delays a firm has had before the analyzed period), Age (used as a measure of reputation), Group membership (a theory that firm belonging to a group of firms doing well might be allowed poor performance since resources could be channelled into the poorly performing firm), Size (measured the natural log of total assets), Efficiency (measured ROI) and Leverage (calculated as total debts divided by total assets). The study proved that it is possible to elucidate financial issues in SME firms using non-financial variables. The study also discovered that the model, which was created using non-financial variables, classified SMEs much better than models with financial ratios. The overall research concluded that the best results were achieved by combining financial and non-financial variables and that the most important non-financial variable was the number of payment delays.

Prediction models, that include non-financial variables, have shown great results also in other researches - usage of variables like reporting and compliance (Altman et al., 2010) and tax arrears (Kohv & Lukason, 2021; Lukason & Andresson, 2019) have proven to give better accuracy compared to models based on only financial variables. In one of the study, made by Altman et al. (2020), 11 different characteristics were used in their models (characteristics like prior defaults, board characteristics; age; type of firm; industry payment default and bankruptcy risk; payment delays and a number of payment delays over 60 days and delays/total assets were used) and with a dependent variable bankrupt/non-bankrupt, they managed to create a model with accuracy around 93%.

The closest study about Non-financial variables (namely annual report delays) was made by Lukason and Camacho-Miñano (2019) - the study was conducted in Estonia and included firms with different industries, ages and sizes. They observed if firms' reporting delays are connected with bankruptcy risk and its financial ratios and they made 3 conclusions: (1) firms with lower annual and accumulated profitability and liquidity were more likely to delay with their annual reports; (2) firm reporting delays and leverage was not connected, and (3) firms with a higher risk of bankruptcy are more likely to delay with their annual reports (Lukason and Camacho-Miñano, 2019). All of these three conclusions can be explained and supported also by Ajinkya et al. (2005) - firms with losses will with lower likelihood present their information.

Another research with Estonian SMEs was done by Lukason and Andresson (2019) where they compared financial ratios and tax arrears in bankruptcy prediction models. They analyzed the entire population of Estonian survived and bankrupt SMEs between 2013 to 2017 and concluded that (1) tax arrears` information, compared to financial ratios, has much higher prediction accuracy when bankruptcy becomes closer, and (2) combining tax arrears and financial ratios into one model is more useful than individual ones (Lukason and Andresson, 2019). One SME-based research was also made by Tobback et al. (2017) which focused on bankruptcy prediction for SMEs using relational data - including two large real-life SME data sets (Belgian and the UK), they observed the potential of relational data for bankruptcy prediction. They observed whether 1) there is a higher probability for bankruptcy if a firm is linked to many bankrupted firms and 2) management`s previous success/failure and competence has an influence on the performance of the company. In general, they showed that complementary predictive power to traditional bankruptcy prediction can be added by linking companies based on their managers/board members. The link between these two aspects (bankruptcy and managers/entrepreneurs) is also previously proved by Ooghe and Prijcker (2008), where they qualitatively observed 4 types of failure processes for 1) unsuccessful startups, 2) ambitious growth companies, 3) dazzled growth companies, and 4) apathetic established companies. Their research showed that although all 4 types of processes had several independent causes for bankruptcy, there is always an interaction between the failure process and the direct or indirect importance of management errors (they observed competence skills, Motivation and Personal characteristics from the management side).

In general, it can be said that there is a great amount of research on business failure modelling, where the most used methodology for models was artificial neural networks and logistic regression, but the existing literature usually does not seem to include non-financial variables even though 1) some of the non-financial variables have actual failure prediction capability (for example Lukason & Andresson, 2019) and 2) models that combine financial and non-financial variables outperform models using only financial ratios (for example, Altman et al., 2010). There is even less (or hardly any) research that puts entrepreneurial behaviour into the context of failure prediction, and therefore it can be said that this domain is quite unexplored (Ciampi, 2018). Based on 1) several studies that show a clear link between increased risk of financial distress and late or non-submission of reports (for example Lukason and Camacho-Miñano, 2019), and 2) the fact that entrepreneurial behaviour tends to repeat (Lattacher & Wdowiak, 2020), adding previous entrepreneurial behaviour variables like annual report delays, and tax declaration report delays gives potential to increase the accuracy of business failure prediction models - individually or in combination with other variables.

## 3. Dataset, variables and method

### 3.1 Dataset

The time period in this study is from 2016 to 2019, which means that the global financial crisis effects were disregarded. The dataset is gathered from two sources - Estonian Tax and Customs Board and Estonian Business Register. The dataset of this study contains 2 types of firms - defaulted and non-defaulted firms - with the following restrictions. First, all firms in the dataset had to be value-added tax (VAT) obligatory at some point in time, namely non-defaulted firms registered in the VAT register at the end of 2019 and defaulted firms that had been in that register at some time in the past. Secondly, the firms in the dataset were legally registered as joint-stock companies or private limited companies. Thirdly, defaulted firms in the dataset were firms, who were deleted from Estonian Business Register while having tax arrears and it was possible to observe their board members' last 3 full years entrepreneurial behaviour from the moment tax debt appeared. It must be noted, that the appearance of permanent tax arrears can take place many years before the respective firm is deleted from the business register. Fourth, as the entrepreneurial data availability for this thesis starts from 2013, only those defaulted firms can be included, in case of which the permanent tax arrears did not start before 2016.

Basel II conception of default is that a firm has an unpaid tax debt for 90 days past due (Wagner 2016). In the context of this thesis, it means that a firm is considered non-defaulted when at the end of 2019 it did not have three consecutive months of tax arrears. All the defaulted firms have at least three months of tax arrears, although the past three years of entrepreneurial behaviour is accounted from the point where three months of permanent tax arrears were observable. The dataset presents the overall population of firms in the context of the Estonian Business Register, including all economic sectors and all companies. In addition, almost exclusively SMEs are considered.

The dataset consisted of 44 172 observations - 1200 defaulted and 42 972 non-defaulted firms. All the observations are unique, meaning that the initial dataset is unbalanced. To eliminate possible anomalies and limiting extreme values, winsorization is used in the dataset - all the values under or over certain limits were replaced with maximum or minimum values. Default is used as the binary dependent variable according to the principles explained earlier, while the defaulted firm is coded with 1 and 0 otherwise. Table 1 below describes the dataset for defaulted and non-defaulted firms, dividing them based on the year when each of the firms in the dataset was officially registered. As this study's observation for defaulted firms examine delays within a 3 full year timeframe back from the year when default appeared, the latest defaulted firms in the dataset were registered in 2017 (for example, when a firm defaulted in 2018, then full years delays from 2017, 2016 and 2015 were under observation). The dataset for non-defaulted firms also includes firms from 2018 and 2019.

Table 1

*Overview of the dataset - number of defaulted and non-defaulted firms*

| **Year registered** | **Defaulted** | **Non-defaulted** |
|---|---|---|
| 1995 - 1999 | 54 | 7592 |
| 2000 - 2004 | 86 | 5718 |
| 2005 - 2009 | 181 | 8475 |
| 2010 - 2014 | 554 | 10 814 |
| 2015 - 2017 | 325 | 7541 |
| 2018 - 2019 | - | 2832 |
| Total | 1200 | 42 972 |

Source: Compiled by the author.

Based on Table 1 above, the majority of defaulted firms fail during their first years of operation - around 27% of the defaulted firms used in this study have defaulted within their first years and 73% have defaulted within 9 years timeframe.

### 3.2 Financial ratios

Financial ratios that are used in this thesis are chosen based on previous literature, meaning that domains, which are most commonly used in failure prediction, are included - solvency, liquidity, profitability, productivity. Although previous studies have shown that for SMEs, financial ratios are not suitable for failure prediction and data loss is remarkable, they are added to this thesis for comparison purposes. The 4 ratios have been selected based on Kohv and Lukason (2021) and Lukason and Andresson (2019) similar studies about failure prediction of Estonian SMEs. The chosen financial ratios are described in Table 2 below.

Table 2

*Financial ratios, abbreviations and formulas*

| **Financial ratio** | **Abbreviation and formula** |
|---|---|
| Financial structure, leverage, solvency | TETA = $\frac{total\ equity}{total\ assets}$ |
| Liquidity | WCTA = $\frac{current\ assets - current\ liabilities}{total\ assets}$ |
| Productivity | ORTA = $\frac{operating\ revenue}{total\ assets}$ |
| Profitability | NITA=$\frac{net\ income}{total\ assets}$ |

Source: Based on Kohv and Lukason (2021) and Lukason and Andresson (2019).

To calculate these ratios, the annual financial statements of each firm were used. To include as many firms as possible in the analysis, the latest available financial statement is used, meaning that:

- for non-defaulted firms, the statements are mostly from the end of 2019 financial statements were searched until 2010 due to the limitation of data availability, but there were hardly any firms with such old financial data and,
- for defaulted firms, the latest available financial statement was used (financial statements were searched until 2010 due to the limitations of data availability - this means that when a defaulted firm`s last financial statement was submitted in 2005, then this firm is not included in the prediction model applying financial ratios).

The limitation of the year 2010 was chosen because data from the previous economic boom can distort the analysis. Logically, firms without any reports were excluded from the data.

### 3.3 Variables about tax reporting and annual reporting delays

As previously described, the variables used in this thesis are from two domains: a) tax reporting delays, and b) annual report delays. Companies in Estonia have to submit their reports based on the characteristics of the report(s) - tax declaration reports are submitted every month and an annual report has to be submitted once a year. In this thesis, both types of report delays are considered equal, meaning there is no prioritization between those two. Regarding the selection of variables, other behavioural variables were not considered in this thesis. The main reason is limitations of data availability - data about other types of violations of regulations is not available and therefore not measurable.

Tax declaration report delays measures entrepreneurs' (i.e board members of companies), who were board members the moment when default appeared, all registered firms monthly tax reports delays (VAT report and income- and social tax (i.e IST) report). It measures delays 3 years back (full years) from the moment when default appeared to a firm used to code the dependent variable. As reports are done every month, it observes the fact that firms, where the entrepreneur was officially a board member at that time, did not submit tax reports on time during that time. For example, when a firm defaulted in 2018, entrepreneurs` firms' delays from 2017, 2016 and 2015 were observed. The delays are measured from three points of time for each firm during the 36 months period (i.e. 3 years) and respective three variables are calculated - a) the fact that one and/or another tax report (i.e IST and/or VAT report) was delayed at least one day (SDEC); b) tax report delay went over the end of the month (LDEC), and c) summarized delay of tax reports (DEC = SDEC+LDEC). SDEC value can be a maximum of 2, as it measures report delays from two different points, as two different tax declarations are considered. As tax report delays are measured from three points of time (i.e. one day after IST date, one day after VAT date, the month-end), it means that the maximum number for delays for each firm in a year can be 36 (12 * 3) and 108 for 3 years. When an entrepreneur has several firms, it measures delays over all firms, meaning that delays are summarized over all firms.

Annual report delays measures entrepreneurs' (board members of companies), who were board members the moment when default appeared, all registered firms annual reports delays. It measures delays 3 full years back from the moment when default appeared to a firm. As annual reports are done once a year, it observes the fact that firms, where the entrepreneur was officially a board member, did not submit their reports on time. The delay is measured with three variables for each firm - a) annual report is delayed up to 365 days (SAR); b) annual report(s) are delayed over 365 days (LAR), and c) summarized delay of annual report(s) (AR = SAR + LAR). The maximum number of delays for each firm can be 3. When an entrepreneur has several firms, it measures delays over all firms, meaning that delays are summarized for all firms. Due to data availability, short-term delays were considered equal when the delay was 2 days, 20 days or 200 days.

As the dataset of this study enables to allocate and observe reporting delays year by year, additional variables are used in this thesis - regarding the maximum yearly value of delays over three year time, SDECMAX, LDECMAC and DECMAX are added for tax declaration report delays observations; and SARMAX, LARMAX and ARMAX is added for annual report delays observations. These variables are calculated based on the reporting delays over all the observed 3 years:

- SDECMAX is the yearly maximum value over three years of SDEC values.
- LDECMAX is the yearly maximum value over three years of LDEC values.

- DECMAX is the yearly maximum value over three years of DEC values.
- SARMAX is the yearly maximum value over three years of SAR values.
- LARMAX is the yearly maximum value over three years of LAR values.
- ARMAX is the yearly maximum value over three years of AR values.

Adding these variables to the study adds another perspective - if and how suitable are maximum yearly values in business failure prediction and how much more/less accurate they are compared to more traditional financial variables or summarized non-financial variables like AR or DEC.

### 3.4 Method

The full dataset was used for non-financial variables but as the dataset was not complete to calculate financial variables for all the observations, then if at least some of the financial data were available, possible financial variables were figured and added (meaning that financial variables were calculated to those observations where it was possible). Table 3 shows the number of defaulted and non-defaulted firms, where a financial variable could be calculated.

Table 3

*Overview of the number of firms to calculate a specific financial variable*

| Financial variable | Defaulted | Non-defaulted |
|---|---|---|
| TETA | 370 | 42 310 |
| ORTA | 363 | 41 964 |
| WCTA | 343 | 42 006 |
| NITA | 368 | 42 278 |

Source: Compiled by the author.

This thesis uses the logistic regression (LR) method, which is one of the most used methods in failure prediction models (Ciampi, 2015). As the aim of this thesis is to show how accurate non-financial variables can be in predicting business failure, using classical logistic regression is also motivated by Altman et al. (2020) study which showed that logistic regression is one of the best methods in bankruptcy prediction.

Taking into account the dataset of this study and the fact that it includes the whole population of firms, the observations are not balanced - meaning that the amount of defaulted firms is 2.72% from the whole dataset. As using LR would cause false classifications for the defaulted group, the two groups were made more equal with weighting - weighting techniques are also commonly used in several previous studies (for example Kohv and Lukason, 2021).

In the analysis, first, the prediction ability of non-financial variables was tested individually to each variable to see how accurate they are for predicting business failure. Next, financial variables were tested individually for the same purpose. After that, the variables were grouped and tested based on three characteristics - a) financial variables, b) non-financial variables of summarized reporting delays (AR, LAR, SAR, DEC, LDEC, SDEC) and c) non-financial variables of maximum yearly values (ARMAX, LARMAX, SARMAX, DECMAX, LDECMAX, SDECMAX). Finally, variables were combined to create the best possible and most accurate model.

## 4. Results and discussion

As there are not many previous studies regarding using non-financial variables in failure prediction, many findings from this thesis are unique and direct comparison with previous studies cannot be performed and deserve further investigation and observation.

Descriptive statistics of all (financial and non-financial) variables individually can be seen from Table 4 and 5 below.

Table 4

*Descriptive statistics of financial variables*

| Variable | Statistic | Defaulted firms | Non-defaulted firms |
|---|---|---|---|
| TETA | Mean | 0.438 | 0.589 |
| | Median | 0.476 | 0.661 |
| WCTA | Mean | 0.235 | 0.353 |
| | Median | 0.254 | 0.369 |
| NITA | Mean | 0.665 | 0.116 |
| | Median | 0.393 | 0.786 |
| ORTA | Mean | 2.647 | 2.309 |
| | Median | 1.80 | 1.61 |

Source: Compiled by the author

Table 5

*Descriptive statistics of firm managers annual reporting delays and tax declaration report delays*

| | | | Defaulted firms | Non-defaulted firms |
|---|---|---|---|---|
| | | Total observations | 1200 | 42 972 |
| **Annual report delays** | AR | Maximum | 9 | 9 |
| | | Mean | 5.11 | 0.97 |
| | | Median | 6.00 | 0.00 |
| | SAR | Maximum | 5 | 5 |
| | | Mean | 1.72 | 0.44 |
| | | Median | 0.00 | 0.00 |

| | | | | |
|---|---|---|---|---|
| | LAR | Maximum | 6 | 6 |
| | | Mean | 3.29 | 0.48 |
| | | Median | 4.00 | 0.00 |
| **Tax declaration report delays** | DEC | Maximum | 77 | 77 |
| | | Mean | 45.45 | 12.61 |
| | | Median | 52.50 | 3.00 |
| | SDEC | Maximum | 77 | 77 |
| | | Mean | 21.05 | 11.00 |
| | | Median | 14.00 | 3.00 |
| | LDEC | Maximum | 16 | 16 |
| | | Mean | 10.20 | 1.55 |
| | | Median | 16.00 | 0.00 |

Source: Compiled by the author.

From Table 4, financial variables statistics, it can be seen that non-defaulted firms have better financial variable values (means and medians) than defaulted firms. This is nothing new and confirms Lukason and Camacho-Miñano (2019) findings, that non-defaulted firms usually show better financial performance.

Based on the descriptive statistics for non-financial variables from Table 5 it can be said that maximum values of all the variables may not be the indication of possible default - the values are the same for both, defaulted and non-defaulted firms. On the other hand, mean and median for non-financial variables show that defaulted firms have noticeable differences when it comes to reporting delays - defaulted firms managers have delayed their needed reports much more frequently compared to non-defaulted firm managers. For example, the variable AR mean is over 5 times higher and the DEC variable is 3.6 times higher for defaulted firms, so it is clearly visible that defaulted firms` managers delay with their reports. This can be explained with Lukason and Camacho-Miñano (2021) finding that financial distress (which may lead to default) and annual reporting delays have a strong link. Individual variable-based prediction accuracies (%) for non-financial and financial variables can be seen in Table 6 where all of the variables were tested separately.

Table 6

*Variables` individual prediction accuracies for defaulted and non-defaulted firms*

| Domain used | Defaulted | Non-defaulted | Summarized |
|---|---|---|---|
| DEC | 65.4 | 83.5 | 74.3 |
| SDEC | 49.6 | 76.4 | 62.8 |
| LDEC | 68.7 | 88.6 | 78.5 |
| AR | 66.3 | 85.3 | 75.6 |
| SAR | 43.9 | 86.6 | 65 |
| LAR | 59.8 | 91.5 | 74.5 |
| DECMAX | 58.2 | 85.8 | 71.8 |
| SDECMAX | 46 | 74.8 | 60.2 |
| LDECMAX | 59.7 | 92.8 | 76 |
| ARMAX | 59.9 | 85.8 | 72.7 |
| SARMAX | 43.9 | 86.6 | 65 |
| LARMAX | 52.3 | 91.5 | 71.7 |
| TETA | 59.7 | 54.6 | 57.3 |
| ORTA | 63.9 | 40 | 52.4 |
| WCTA | 53.5 | 54.5 | 54 |
| NITA | 75 | 17.8 | 47.5 |

Source: Compiled by the author.

Table 6 above shows that prediction accuracy differs over variables and it also differs for defaulted and non-defaulted firms. When comparing individual non-financial and financial variable results for summarized prediction accuracy, it can be said that financial variables are not useful in predicting business failure - observing the variables individual accuracy, none of them did not remarkably exceed the 50% accuracy threshold. Variable TETA (financial structure) showed the best-summarized accuracy from financial variables, but with its 57.3% accuracy, it is still not useful. It again confirms Lukason and Laitinen (2019) findings with European bankrupt firms, that financial reports are not always the best indicators to capture problems in an SME company. Only NITA showed a more accurate prediction capacity for defaulted firms, which may also be natural - low/no profitability may not be sustainable. On the other hand, NITA showed almost no prediction accuracy for non-defaulted firms. Moreover, financial variables show generally lower accuracy for non-defaulted firms.

As from non-financial variables, in general, they show good prediction results and it can be said, that there are no remarkable differences between variables that take into account summarized reporting delays (DEC, AR, LDEC, LAR, etc.) or maximum reporting delays over years (DECMAX, ARMAX, LDECMAX, etc.). Still, variables for summarized reporting delays show better results for defaulted firms which means that when seeking better accuracy, it is probably reasonable to combine the non-financial variables. Generally, the detected phenomenon is also supported by Ajinkya et al. (2005), Altman et al. (2010) and Lukason and Camacho-Minano (2019) - firms with problematic situations show their information less likely. Another explanation is by Lukason and Camacho-Miñano (2021) - financial distress and delayed annual reports have a strong link. In addition, short-term reporting delay variables like SDEC, SAR, SDECMAX and SARMAX show slightly lower prediction ability which could point to larger randomness of the occurrence of short-term delays.

Table 7

*Models combining different variables with respective prediction accuracies.*

| Model nr | Variables used | Defaulted | Non-defaulted | Summarized |
|---|---|---|---|---|
| 1 | DEC and AR | 79.4 | 83.8 | 81.6 |
| 2 | SDEC and SAR | 50.7 | 81.7 | 66 |
| 3 | LDEC and LAR | 81.5 | 87.7 | 84.5 |
| 4 | DEC, AR, LDEC, LAR | 81.9 | 87.5 | 84.7 |
| 5 | SDEC, SAR, LDEC, LAR | 81.3 | 88.3 | 84.7 |
| 6 | DEC, AR, SDEC, SAR | 79.1 | 88.8 | 83.9 |
| 7 | DEC, AR, SDEC, SAR, LDEC, LAR | 80.6 | 88.6 | 84.5 |
| 8 | All maximum-variables ( DECMAX, SDECMAX, LDECMAX, ARMAX, LARMAX, SARMAX) | 74.8 | 90.7 | 82.6 |
| 9 | All non-financial variables (DEC, AR, SDEC, SAR, LDEC, LAR, DECMAX, SDECMAX, LDECMAX, ARMAX, LARMAX, SARMAX) | 81.5 | 88.8 | 85.1 |
| 10 | TETA, ORTA, WCTA, NITA | 56.8 | 65.1 | 61 |

Source: Compiled by the author.

Based on the results presented in Table 7, it can be said that all models except one with non-financial variables give over 80% accuracy in predicting possible business default. Only a model with short term reporting delay measures (Model nr 2, SDEC and SAR) gave only 66% accuracy. It can also be said that variables for long-term reporting delays and models with these variables raise the results - accuracy was over 84% for all of those combinations, where LDEC and LAR was included (Model nr 3, 4, 5 and 7). Therefore, when an entrepreneur is constantly delaying with its firms' reports, it is a good indication that most probably he/she will also default with their current/new firms.

Based on the results, it also seems that variables for short-term reporting delays (SAR and SDEC) lower the summarized accuracy - SDEC and SAR in combination gave only 66% accuracy and models where these variables were used, gave also slightly lower results (except Model nr 5) compared to the best ones, but the difference is only 0.2-0.8 percentage points. On the other hand, SDEC and SAR variables individually and both of them combined did not show good accuracy for defaulted firms - maybe it is the indolence of entrepreneurs, but short-term bypassing does not seem to show possible default.

When looking at the accuracy for only defaulted firms, the best results are given by Model nr 4, where summarized and long-term reporting delay variables were used - with the accuracy of 81.9% it is the best for defaulted firms from tested models. On the other hand, Model nr 3 (long-term delays) and 5 (short-term and long-term delays) follow with only a slight, 0.4-0.6 percentage point shortage. Those models are also good at predicting non-default and the summarized prediction is in the range of 84.5 - 84.7%.

Combining all the non-financial variables (Model nr 9) raises the summarized prediction accuracy to 85.1% but not remarkable rise occurs when compared with previous models - therefore, for example, when reporting delays information availability is limited, long-term reporting delays information should be preferred, because adding short-term delay information does not raise the accuracy noticeably.

The main implication of this thesis is that non-financial variables in the form of previous entrepreneurial misbehaviour show high predictive performance when it comes to the prediction of firm defaults. As there is hardly any prior literature, it is a domain that merits further research and attention. Just like in this study, previous studies show that financial variables may show quite good performance before default - companies default quickly and it may not be predicted from financial reports. Therefore the current thesis confirms previous research (for example Laitinen & Lukason, 2014; Ciampi et al., 2020; and Ciampi et al., 2021) where the importance of non-financial information and variables is raised.

The main practical implication from this study is that previous entrepreneurial behaviour on annual report delays and tax declaration report delays is a good and clear sign of increased default risk. Short-term delays do not show such good results, but long-term reporting delays is a strong indicator that something may be going bad. As entrepreneurial behaviour tends to repeat, reporting correctness in previous firms is a good indicator when it comes to business failure - when a person has defaulted several firms beforehand and he/she starts to delay with his/her current/new firms reports, then the possibility for new defaults is higher and actions may need to be taken.

Practical usage of this information should bring benefits to financial institutions like banks or other creditors, or to investors - seeing and acting timely would potentially minimize possible losses. It should also help in decision making for crediting or investing into new firms - as businesses may fail really quick, then every possible red sign should be gathered and considered. The current study, therefore, confirms Ciampi et al. (2021) proposal for adding qualitative variables to SME default prediction models and proved why non-financial variables are important in this domain.

## 5. Conclusion

This study aimed to find out if and how accurate non-financial aspects like past entrepreneurial behaviour in the form of annual report delays and tax declaration report delays is in the context of business failure prediction. For the analysis, logistic regression was applied to a dataset that consisted of VAT obligatory Estonian defaulted and non-defaulted firms. The dataset originated from the Estonian Tax and Customs Board and Estonian Business Register and consisted of over 44 000 firms.

The results of this study showed that past entrepreneurial behaviour from reporting delays (on-time annual reports and tax declaration reports) provide reliable and high-accuracy non-financial variables in business failure prediction. The study evaluated different aspects of reporting delays and for the theoretical implication, it discovered that previous reporting delays are a clear sign of possible problems. As entrepreneurial behaviour tends to repeat, it can be said that when an entrepreneur has a “bad” record regarding these delays, it may also lead to a default of a new or current firm. The practical implication of this thesis should help to understand the importance of non-financial variables as trustworthy and good indicators when it comes to business failure prediction. As firms are affected by unforeseen external causes, they can default quickly and may not show clear signs of financial problems beforehand, other alternative indicators should be taken into account - especially for new firms, who have no or little financial statements. Therefore, additional indicators should be monitored and entrepreneurs' previous behaviour is a good aspect to be implemented into failure prediction.

Future research could elaborate testing these variables with other methods, proving and confirming the accuracy of these variables even more. As this study was conducted with a dataset from one single country, a larger international dataset should also be used (although there may be issues with data availability from several countries). Last but not least, when this further research is done, the next step would probably be adding these variables into wider usage - for example, into commercial business failure and default prediction models.

**List of references**

1. Ajinkya, B., Bhojraj S., Sengupta, P. (2005). The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research*, *43*, 343–76. https://doi.org/10.1109/72.935101
2. 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
3. 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
4. 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
5. 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
6. Back, P. (2005). Explaining financial difficulties based on previous payment behaviour, management background variables and financial ratios. *European Accounting Review, 14(4)*, 839–868. https://doi.org/10.1080/09638180500141339
7. 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
8. 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
9. 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

10. 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
11. 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
12. Ciampi, F., Giannozzi, A., Marzi, G., Altman, E.I. (2021). Rethinking SME default prediction: a systematic literature review and future perspectives. *Scientometrics 126*, 2141–2188. https://doi.org/10.1007/s11192-020-03856-0
13. Du Jardin, P., Severin, E. (2011). Predicting Corporate Bankruptcy Using a Self-organizing Map: An Empirical Study to Improve the Forecasting Horizon of a Financial Failure Model. *Decision Support Systems, 51*, 701–711. https://doi.org/10.1016/j.dss.2011.04.001
14. Du Jardin, P. (2015). Bankruptcy Prediction Using Terminal Failure Processes. *European Journal of Operational Research, 242*, 286–303. https://doi.org/10.1016/j.ejor.2014.09.059
15. 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
16. Kohv, K., Lukason, O. (2021). What Best Predicts Corporate Bank Loan Defaults? An Analysis of Three Different Variable Domains. *Risks, 9(2)*, 1-19. https://doi.org/10.3390/risks9020029
17. 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
18. 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

19. Lattacher, W., Wdowiak, M. A. (2020). Entrepreneurial learning from failure. A systematic review. *International Journal of Entrepreneurial Behavior & Research, 26 (5)*, 1093-1131. https://doi.org/10.1108/IJEBR-02-2019-0085
20. 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
21. 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
22. Lukason, O., Camacho-Miñano, M. D. M. (2021). What Best Explains Reporting Delays? A SME Population Level Study of Different Factors. *Sustainability, 13(9)*, 4663. https://doi.org/10.3390/su13094663
23. Lukason, O., Hoffman, R.C. (2015). Firm failure causes: a population level study. *Problems and Perspectives in Management, 13(1)*, 45-55. https://businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/6325/PPM_2015_01_Lukason.pdf
24. 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
25. Ooghe, H., Prijcker, S.-D. (2008). Failure processes and causes of company bankruptcy: a typology. *Management Decision, 46*, 223-242. https://doi.org/10.1108/00251740810854131
26. Pretorius, M. (2009). Defining Business decline, failure and turnaround: A content analysis. *The Southern African Journal of Entrepreneurship and Small Business Management, 2(1)*, 1-16. http://dx.doi.org/10.4102/sajesbm.v2i1.15
27. Prusak, B. (2018). Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries. *Int. J. Financial Stud, 6(3)*, 60. https://doi.org/10.3390/ijfs6030060
28. 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

29. Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M., Martens, D. (2017). Bankruptcy prediction for SMEs using relational data. *Decision Support Systems, 102*, 69-81. https://doi.org/10.1016/j.dss.2017.07.004
30. Veganzones, D., Severin, E. (2020). Corporate failure prediction models in the twenty-first century: a review. *European Business Review, 33(2)*, 204-226. https://doi.org/10.1108/EBR-12-2018-0209
31. Wagner, H. (2016). Default definition under Basel. N. Siddiqi (Ed.), Intelligent credit scoring: Building and implementing better credit risk scorecards (2nd ed., 119– 130). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119282396.ch7

## Resümee


ETTEVÕTJA EELNEV VÄÄRKÄITUMINE ISIKUGA SEOTUD ETTEVÕTETE MAKSEHÄIRETE ENNUSTAJANA

Anne-Liis Tamm

Ettevõtete pankroti prognoosimise mudelite uurimine ning arendamine majandusteaduses sai alguse juba 1960ndatel ning mudelite ennustuse täpsus on läbi aegade läinud järjest paremaks. Kuigi peamiselt kasutatakse kättesaadavuse lihtsusest tulenevalt taolistes mudelites finantsilisi näitajaid, on mitte-finantsiliste näitajate mõju uurimine muutumas järjest populaarsemaks - erinevad autorid on uurinud näiteks nii firmade vanuse, juhtkonna kompetentsuse, võlgnevuste kui ka valdkonnast tulenevate aspektide mõju ettevõtete edule/ebaedule ning nende pankroti- ja riskimudelitele. Küll aga ei ole uuritud ettevõtjate eelnevat käitumist just väärkäitumiste aspektist, ehk kas ja kui täpselt on võimalik eelneva väärkäitumise pealt ennustada isikuga seotud teiste ettevõtete maksehäirete tekketõenäosust.

Käesolev magistritöö uurib, kuivõrd täpselt on võimalik ettevõtja eelneva väärkäitumise (regulatsioonide rikkumise) pealt ennustada temaga seotud ettevõtete ebaedu väljendatuna püsiva maksehäire tekkes. Vaatluse all on 2 mitte-finantsilist muutujat - maksudeklaratsioonide (käibemaksudeklaratsioon ning tulu- ja sotsiaalmaksu deklaratsioon) ning aastaaruande esitamisega seotud hilinemised. Töö valimis olid kõik Eesti käibemaksukohuslasest ettevõtted - Eesti Äriregistrist ning Maksu- ja Tolliameti andmekogudest sai kokku koondatud info 44 172 toimiva ning oma tegevuse lõpetanud ettevõtte kohta, sisaldades peamiselt väikese ning keskmise suurusega ettevõtteid. Võimalike seoste ning ennustustäpsuse määramiseks kasutati töös logistilist regressiooni. Vaatlusalused muutujad tõestasid, et nende kasutusväärtus ettevõtete edu/ebaedu ennustavates mudelites on väga kõrge - parim mudel andis üle 81% täpsuse firmade ebaedu ennustamisel ning kokku koondtäpsuse 84.7%. Kuna panganduses kasutatakse ettevõtete erinevates riskimudelites peamiselt finantsilisi näitajaid, mis oma olemuselt aga ei pruugi olla alati parima täpsusega eriti uute ettevõtete korral, siis on uute, just mitte-finantsiliste muutujate kaasamine vastavatesse mudelitesse uus ning arenev valdkond. Lisaks on tõestatud, et ettevõtjate

käitumine tõenäoliselt kordub ning seetõttu on vastavate nüansside jälgimine oluline ning kasulik nii pankadele, ettevõtete koostööpartneritele kui ka investoritele, et võimalikult varakult potentsiaalseid maksehäireid ette näha ning vajadusel reageerida. Kuna valdkonnasiseselt pole ettevõtjate eelnevat käitumist liialt palju uuritud, siis väärib käesolev töö kindlasti edasist tähelepanu ning testimist ka teiste andmekogudega - Eesti kontekstis on ennustustäpsus end tõestanud, kuid laiemaks ja praktilisemaks kasutuseks tuleks testida seda ka teiste riikide andmekogudega.

**Võtmesõnad**: ebaedu ennustamine, väikesed ning keskmise suurusega etteõtted, mitte-finantsilised muutujad, ettevõtja väärkäitumine, deklaratsioonidega hilinemised, aastaaruannetega hilinemised