PREDICTING BANKRUPTCY OF NORDIC TECHNOLOGY COMPANIES WITH FINANCIAL STATEMENTS DATA

Lappeenranta–Lahti University of Technology LUT
Bachelor's programme in Business and Management, Bachelor's thesis
2022
Author: Aleksi Sintonen
Examiner: Associate Professor Maija Hujala

ABSTRACT

Lappeenranta–Lahti University of Technology LUT
LUT School of Business and Management
Business Administration

Aleksi Sintonen

Predicting bankruptcy of Nordic technology companies with financial statements data

Bachelor's thesis

2022

31 pages, 2 figures, 9 tables and 9 appendices

Examiner: Associate Professor Maija Hujala

Keywords: Bankruptcy, financial ratio, financial statements, logistic regression, predictive model, technology company

Evaluating the probability of company's bankruptcy can significantly improve credit risk management among financiers, and also provide tools for companies to perform self-evaluation regarding the risks. The aim of this thesis is to find if using historical financial statements data in logistic regression model is applicable in predicting the bankruptcy of Nordic technology companies.

Former research has shown that financial ratios predict bankruptcy at least satisfactory level and, on many occasions, very precisely. Although modern methods utilizing artificial intelligence usually overperforms conventional statistic methods in prediction, business actors still need simple and easy-to-use methods in evaluating the credit risks.

In the thesis, two predictive models were composed: First model predicts if a company is defaulting or not one year prior and the second model two years prior the possible bankruptcy. The results show that the prediction of bankruptcy of Nordic technology companies is possible at least on satisfactory level based on financial statements data. Especially the size of the company, total liabilities divided by total assets, current liabilities divided by current assets and net income divided by total assets were the ratios which predicted efficiently the bankruptcy.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT
LUT-kauppakorkeakoulu
Kauppatieteet

Aleksi Sintonen

Pohjoismaisten teknologiayritysten konkurssin ennustaminen tilinpäätöstietojen perusteella

Kauppatieteiden kandidaatintyö

2022

31 sivua, 2 kuvaa, 9 taulukkoa ja 9 liitettä

Tarkastaja: Tutkijaopettaja Maija Hujala

Avainsanat: Ennustemalli, konkurssi, logistinen regressio, taloudelliset tunnusluvut, teknologiayritys, tilinpäätös

Yrityksen konkurssin todennäköisyyden arviointi voi parantaa merkittävästi rahoittajien luottoriskien hallintaa sekä tarjota myös yrityksille työkaluja riskien itsearvioinnissa. Tämän tutkielman tavoitteena on selvittää, soveltuuko historiallisten tilinpäätöstietojen käyttö logistisen regressiomallin kanssa pohjoismaisten teknologiayritysten konkurssin ennustamiseen.

Aikaisempi tutkimustausta osoittaa, että taloudelliset tunnusluvut ennustavat konkurssin todennäköisyyttä vähintään tyydyttävällä tasolla, ja useissa tapauksissa erittäin tarkasti. Vaikka nykyaikaiset tekoälyä hyödyntävät menetelmät ylittävätkin yleensä tavanomaiset tilastolliset menetelmät ennustamistarkkuudessa, liike-elämän toimijat tarvitsevat silti edelleen yksinkertaisia ja helppokäyttöisiä menetelmiä luottoriskien arvioinnissa.

Tutkielmassa laadittiin kaksi ennustemallia: Ensimmäinen malli ennustaa yrityksen konkurssia vuotta ennen sen mahdollista ilmenemistä ja toinen malli kaksi vuotta ennen konkurssia. Tulokset osoittavat, että pohjoismaisten teknologiayritysten konkurssin ennustaminen on tilinpäätöstietojen perusteella mahdollista vähintään tyydyttävällä tasolla. Erityisesti yrityksen koko, kokonaisvelat jaettuna kokonaisvaroilla, lyhytaikaiset velat jaettuna lyhytaikaisilla varoilla ja nettotulos jaettuna taseen loppusummalla olivat tunnuslukuja, jotka ennakoivat konkurssia tehokkaasti.

ABBREVIATIONS

AIES	Artificial Intelligence Expert System
AUC	Area Under the Curve
GNP	Gross Domestic Product
LDA	Linear Discriminant Analysis
MDA	Multivariate Discriminant Analysis
OLS	Ordinary Least Squares
ROA	Return on Assets
ROC	Receiver Operating Characteristic Curve
SME	Small and Medium-sized Enterprise
VAL	Overall Validation Accuracy

Table of contents

Appendices

Figures

Tables

1 Introduction

Predicting bankruptcy has been one of the main topics among financial research for a long time (see e.g. Balcaen & Ooghe 2006; Aziz & Dar 2006). When evaluating different financing options, financiers are interested in credit risk management. Here assessing the risks of possible bankruptcy is in crucial role. Also, companies by themselves can better protect their businesses by minimising the risks if such can be anticipated. The aim of this thesis is to find if using historical financial statements data in logistic regression models is applicable in predicting the bankruptcy of Nordic technology companies. Two main research questions are presented:

- *How accurately the bankruptcy of Nordic private technology companies can be predicted based on financial ratios?*
- *Which of financial ratios are predicting best the bankruptcy?*

Logistic regression models have been studied widely in default prediction (e.g. Drehmann & Juselius 2014; Brédart 2015) but not specifically focusing on technology companies. Although differences between industries in predicting the bankruptcy have been studied overall, it seems that any specific peer reviewed study focusing on technology companies has not been done. Mirzaei, Ramakrishnan and Bekri (2015) concluded that the probability of bankruptcy is affected by each industry's characteristics. Tech companies have their distinct capital structure and financial ratio averages if compared to regular manufacturing industry which should be considered when forming prediction models. For example, it is argued that in many cases software-based companies prefer outside equity to debt (Hogan & Hutson 2005). Also, venture capital funds have become more common especially among high-tech companies recently (Asay 2022). As regards the ratio averages, if a company sells only a software as a product, it doesn't have any inventories either which is affecting to common financial ratios such as working capital.

It is essential to review specifically technology companies because they are nowadays producing greatly economic growth in digitalizing world, so it is crucial to identify healthy companies from unhealthy ones to prevent welfare losses caused by bankruptcies and inefficient financial decisions. Evaluating financial health is also important because it affects increasingly liquid derivatives markets such as credit ratings and credit default swaps (Tanaka et al. 2019).

Reviewing specifically Nordic companies is appropriate approach because of lack of sound research on this geographical area, although studies using data from only one Nordic country have been done (i.e. Laitinen & Laitinen 1998; Dakovic, Czado & Berg 2010; Yazdanfar 2011). Nordic well-being society and strong bank-emphasized financing environment, in other words the institutionalized financial sector, form a distinct circumstance for business, so it is reasonable to utilize wider dataset from Nordic countries collectively.

The thesis utilizes Ohlson's O-score method as a base for model building. Logistic regression is used to analyze the relationship between financial statements ratios and the probability of the bankruptcy. Logit model was chosen as a predicting model because of its simplicity and solid evidence of its high performance within several decades in bankruptcy prediction (e.g. Aziz & Dar 2006).

The data used in this thesis consists of Nordic private technology companies. Nordic countries are Denmark, Finland, Iceland, Norway, and Sweden. The review has limited only to private companies because bankruptcies of publicly listed companies happen seldom: this would limit available data and, of course, evaluating the probability of bankruptcy of a listed company is not as relevant topic due to its rarity. The data does not include large companies[1] either because the main part of the data available consists of small and medium sized (SME) companies, so fitting the model is more functional with these delimitations.

Two different models are composed, of which first is classifying companies based on company's financial ratios one year prior the bankruptcy, and second based on ratios two year prior the bankruptcy. Regarding the first model, the data is separated into two groups: first group to build the model and the second to evaluate out-of-sample predicting accuracy. In turn, the separation is not functional for the second model due to the smaller sample size, so the secondary testing is done only for the first model.

[1]Here, a large company is defined according to Statistics Finland's and European Commission's (2003/361) definition: Number of employees ≥ 250 and total assets > 43M€ or turnover > 50M€. SMEs were considered to be all other except large companies.

The thesis is organized as follows: The second chapter reviews shortly the definition of bankruptcy and then looks over the most used methods in predicting the bankruptcy. After this, the logit models, specifically Ohlson's model, are reviewed more precisely because logit model and Ohlson's predictors are the methods used in the thesis. At the end of the chapter, hypotheses are introduced. On the third chapter the data collection and methodology of logit models are introduced, including mathematical presentation of logistic regression, logit model, and the relationship between these terms. Then, the fourth chapter comprises the actual model development and testing the selected dataset according to research questions. Lastly, the fifth chapter gathers the important conclusions and proposes the relevant applications and further research needs.

2 Predicting the bankruptcy

First the definition of bankruptcy in Nordic countries is reviewed. Then different methods of predicting the bankruptcy are introduced briefly to clarify how logit models are positioned in the field of prediction in overall. Lastly, the logit models in prediction are reviewed more precisely, mainly focusing on the advantages and issues, due to their importance in developing thesis' models.

2.1 Definition of bankruptcy

Bankruptcy as a term can be approached from different directions. In this thesis the data to be processed consists of companies from several countries, so to be sure that the data from different countries will not differ in processing, the definition should be as strict as possible. Hence, a legalistic approach to the definition of bankruptcy is well-defined and precise.

In Finland bankruptcy process is regulated by Bankruptcy Act (120/2004). According to Bankruptcy Act's Chapter 1 Section 2 "bankruptcy is a form of insolvency proceedings covering all the liabilities of the debtor, where the assets of the debtor are used in payment of the claims in bankruptcy". Insolvency is a prerequisite for bankruptcy; according to Chapter 2 Section 1 "insolvency means that the debtor is otherwise than temporarily unable to repay his or her debts as they fall due".

The definition of bankruptcy and assumption considering the insolvency is almost identical also in other Nordic countries (see Andersson & Hansen-Nord 2021; EMCC 2021; Sætermo 2022; Sigurðsson 2014). EU legislation has been in the background of this harmonization, specifically European Council's regulation (1346/2000), which has directed countries' legislation to form solid and uniform laws regarding bankruptcy procedures. Hence, the differences in countries' legislation are not a challenge in data collection, and more exact comparison is not needed.

2.2 History of predicting the bankruptcy

Over the decades, many different methods in studying the prediction of bankruptcy have been used. One way to classify different approaches is to separate three categories of prediction models: statistical models, artificially intelligent expert system models (AIES) and theoretical models, from which first two are focusing on symptoms of failure and the last on qualitative causes of failure (Aziz & Dar 2006). In practical terms, all these three categories are still dependent on statistical basis. The focus of this thesis is on statistical empiricism, so next, different statistical methods are discussed. Because of recent great performance of AIES methods, they are also briefly introduced to enable the comparison between classical statistical and modern AIES methods. Theoretical models are not in the focal point in this thesis.

As regards the statistical methods, one of the most notable bankruptcy prediction methods has been Altman’s Z-score (Altman 1968), which has been widely used and developed during the last 50 years. It is based on multivariate discriminant analysis (MDA) and has achieved 90–95 % in-sample accuracy in correctly predicting the bankruptcy in Altman’s studies. It has also been accurate with several secondary samples, but naturally out-of-sample accuracy has been lower (Altman 1968). MDA approach has later been updated and tested with different datasets and better predictive ability has been achieved (i.e. Altman, Haldeman & Narayanan 1977; Altman 1993). In subsequent studies the accuracy of MDA prediction models has varied somewhat. For example, in an extensive literature review of Aziz and Dar (2006), the MDA method achieved 85 % average overall predictive accuracy with the sample of 25 studies.

Alongside MDA analysis, Ohlson’s (1980) logit model approach rose in an important role in bankruptcy prediction. Logit models are very suitable for classification problems because of the pivotal objective to evaluate dichotomous dependent variable. The MDA method has some statistical requirements such as a requirement of normally distributed predictors and the similarity of variance-covariance matrices in both groups, whereas logit model doesn’t require these (Ohlson 1980). Logistic regression is considered to be superior compared to MDA method in predicting the bankruptcy because of these simpler statistical assumptions and the possibility to use non-linear parameters (Ohlson 1980; Altman & Sabato 2007).

Similar method to logit model used in predicting the bankruptcy is probit model which differs from logit model only by using the cumulative normal distribution whereas logit uses cumulative logistic function (Boritz & Kennedy 1995). Differences between probit and logit in accuracy are comparatively low: average overall predictive accuracy of logit models has been reported to be 87 % and 89 % with probit models, although in this particular evaluation only two probit studies were included whereas 19 logit models were included (Aziz & Dar 2006).

In the 21st century, modern analysis methods have been tested widely in bankruptcy prediction. For example, hazard models have been argued to be more appropriate in prediction than single-period models (logit and MDA models) because hazard models produce consistent estimates whereas static models produce biased and inconsistent estimates (Shumway 2001). This is since hazard model pays more attention to the passage of time: the explanatory variable in hazard model is the company's time which it has stayed in healthy group, and the model predicts bankruptcy as a function of the most recent financial statements and the age of a company (Almaskati et al. 2021). The model allows to adjust for period at risk, so it can detect more precisely deterioration in company's health during the time it is leading to bankruptcy (Shumway 2001). Hazard models may generate improved out-of-sample results by utilizing more data than static single-period models (Almaskati et al. 2021).

Since comprehensive studies with wider perspective, such as utilizing several different methods at the same time and using more numerous companies as a dataset, have made. One of the largest and most comprehensive one is made by Tanaka et al. (2019), who used over 66 000 companies in model building. The study utilized random forests as an evaluating method and proposes a method of analyzing the vulnerability of industrial matters.

Models which utilize artificial intelligence expert systems, have performed marginally better than statistical and theoretical models (Aziz & Dar 2006). AIES models can learn and improve problem-solving performance with cumulative experience. One of these methods which is used frequently in bankruptcy prediction is random forests. Random forests method is relatively simple why it is more suitable for pragmatic use than are very accurate machine learning methods such as deep learning, which requires laborious hyperparameter tuning to train the data (Tanaka et al. 2019). Random forests can be used to evaluate each variable's impact in the model separately, and recently it has outperformed models based on logistic regression, LDA and neural networks (Barboza, Kimura & Altman 2017).

In overall, the use of MDA and logit models have dominated the research: they have consistently high predictive accuracy and, which is quite important in bankruptcy prediction, they have achieved low error rates in both type I and II errors with a large number of studies in evaluation (Aziz & Dar 2006).

2.3 Logit models in bankruptcy prediction

One of the most influential papers considering bankruptcy prediction with logit model is Ohlson's (1980) *Financial Ratios and the Probabilistic Prediction of Bankruptcy.* The models were built with nine independent variables composed of companies' financial ratios, including ratios from profitability, solvency, and liquidity:

1. SIZE = log (total assets / GNP price level index)
2. TLTA = Total liabilities / total assets
3. WCTA = Working capital / total assets
4. CLCA = Current liabilities / current assets
5. OENEG = 1, if total liabilities exceeds total assets, 0 otherwise
6. NITA = Net income / total assets
7. FUTL = Funds provided by operations / total liabilities
8. INTWO = 1, if net income was negative for the last two years, 0 otherwise
9. CHIN = (NI_t – NI_{t-1}) / (|NIt| + |NI_{t-1}|), where NI_t is net income for the most recent period. The denominator acts as a level indicator → The variable is measuring change in net income.

These variables are considered more detailed in the third chapter. As a dataset, Ohlson used 105 bankruptcy and 2058 non-bankruptcy companies. It is important to notice that most of other studies referenced in this thesis, the dataset has been balanced between bankruptcy and non-bankruptcy companies.

Ohlson (1980) used three models in his research in assessing the probability of defaulting. In the first model the bankruptcy was predicted within one year (96.12 % of companies correctly predicted), in the second model within two years if the firm did not fail within the next year (95.55 %), and in the third model within one or two years (92.84

%). Statistically significant factors in affecting the probability of failure within one year were the size of the company, measures of the financial structure, measures of performance and measures of current liquidity. The probability of bankruptcy was higher with smaller companies and companies with higher debt to equity -ratio, lower net income -ratio and lower working capital -ratio.

One significant point which Ohlson (1980) considered was the timing issue. Unlike the previous studies (i.e. Altman 1968; Altman, Haldeman & Narayanan 1977), his model took it into account whether the bankruptcy occurred prior to or after the date of financial statements release. Another important conclusion was that the models were relatively simple to use in practical applications. Ohlson argued that the model's classification faces some restrictions from the fact that no pairing was performed for dataset. He also reasoned that the dichotomic explanatory variable (bankrupt or non-bankrupt company) gives a very coarse evaluation for decision-making. One significant deficiency was that Ohlson didn't perform out-of-sample testing. The model does not either utilize any market transaction data which could restrict the predicting accuracy. On the other hand, market transaction data is not a relevant matter as regards private companies in this thesis.

In the 21st century logistic regression models with financial ratios have still performed well in bankruptcy prediction. For example, Cultrera and Brédart (2016) tested methods with Belgian small and medium -sized companies by using logit model with five following explanatory variables: current ratio, return on operating assets before depreciation, global degree of financial independence, proportion of gross value added allocated to tax expenses and cash flow divided by total debt. The control variables were company's size, its age, its activity field (industry), and its region. The results indicated that chosen ratios were better explanatory variables for the non-bankruptcy companies. The bankruptcy was more likely for companies with lower liquidity, profitability, debt structure and added value ratios. Also, smaller and younger companies were more likely to go bankrupt. The model correctly classified more than 80 per cent of the companies and pairing the healthy sample with bankrupt sample could have increased the prediction rate even further.

Ohlson's O-score has been tested also with the data from listed Thailand companies via the straight application of scoring formula and seeing if there is a difference between bankruptcy and non-bankruptcy firms in their scoring. The results indicated that differences between groups were significant, so the companies with a higher O-score were more likely to go bankrupt (Lawrence, Pongsatat & Lawrence 2015).

Altogether, logit models in predicting the bankruptcy have been tested and developed further on 2000s, and still nowadays logit models are usable, mainly because of their simplicity and thus usability to a wide range of different actors. The prediction accuracy of the models utilizing logistic regression has been repeatedly reported to achieve 80 to 90 per cent of correctly classified companies, even with out-of-sample testing (e.g. Alaminos, Del Castillo & Fernández 2016). As for the financial statements data, it is universally available which makes predicting the bankruptcy based on financial ratios so easy (Tanaka et al. 2019).

Although modern AIES predicting methods usually outperform logit models with in-sample results, logit models have frequently outperformed them in out-of-sample accuracy (Fantazzini & Figini 2009). On the other hand, Almaskati et al. (2021) recently found that out-of-sample accuracy was better with AIES models, so there is still no unambiguous viewpoint to the issue. Out-of-sample accuracy is quite important aspect while the practical implementation of prediction models happens with secondary data.

Utilizing logistic regression analysis in predictive purposes has also its weaknesses, and one is caused by purely statistical characteristics of the method. Blum (1974) considered that if an objective and well-sustained theory of symptoms is not included in review, the expectation of a sustained correlation between predictive variables and failure prediction is not meaningful. Hence, statistical models are answering to the question when the bankruptcy will happen but not why. Although, systematic literature review indicates that statistical techniques marginally maintain their dominance in the overall prediction accuracy (OPA) of 68.19 per cent[2] compared to 67.76 and 63.14 per cent for AIES and theoretical models (Appiah, Chizema, & Arthur 2014). Consequently, logistic regression has a successful track record from the several decades in predictive purposes, while AIES models have challenged it with excellent results in the 21st century.

[2] The accuracy rates of prediction models are varying quite a lot between different studies, so it is not reasonable to stare just the absolute percentage rates, but make comprehensive comparison studies where different performance measures are evaluated. Although OPA percentages are relatively low in the particular study, the comparison inside systematic review is however a reliable perspective to evaluate the performance.

2.4 Hypotheses

The accuracy of the prediction models used in this study is assumed to be over 90 % in classifying companies correctly into healthy and unhealthy ones, based on previous studies' accurate results and extensive research background of logit models. Another point supporting presumably high accuracy is that the accuracy has been found to improve if the review is focused on a specific industry (Tanaka et al. 2019).

The effect of different financial ratios on the probability of bankruptcy is presumed to be similar to Ohlson's original study (1980) when dealing with technology companies. The presumptions are shown in Table 1.

Table 1. Presumed effects of different variables on the probability of bankruptcy.

Variable	**Definition**	**Effect**
TLTA	*Total liabilities / total assets*	Positive
CLCA	*Current liabilities / current assets*	Positive
INTWO	*Negative net income during two last years**	Positive
SIZE[3]	*log ((total assets (y-1) + total assets (y-2)) / 2)*	Negative
WCTA	*Working capital / total assets*	Negative
NITA	*Net income / total assets*	Negative
CFTL[4]	*Cash flows from operations / total liabilities*	Negative
CHIN	*Change in net income*	Negative
OENEG	*Total liabilities exceeding total assets**	Indeterminate

** Dichotomous variable, having a value of 0 or 1*

The following four ratios are presumed to predict best the bankruptcy: size of the company (SIZE), total liabilities divided by total assets (TLTA), net income divided by total assets (NITA) and current liabilities divided by current assets (CLCA). Three of the first mentioned ratios performed well also in Ohlson's (1980) case, but he also mentioned that while the size of the company was the most significant predictor in the

[3] Different from the Ohlson's predictive variables, the size of the company is evaluated with the mean of the last and the second last year before bankruptcy. The common logarithm -conversion is also made to the mean to smoothen the large deviation of the company size. Ohlson used price level index to standardize the variation in time, but here it is not necessary because data is collected only from two years of time frame. Now, using the mean of the total assets is also smoothening the large yearly deviations.

[4] Cash flows from operations is used here as almost corresponding value for funds provided by operations because cash flow was more commonly used in the selected accounting data. The same choice was made by Lawrence, Pongsatat & Lawrence (2015) when defining predictive variables in bankruptcy prediction.

model, the dataset was heavily biased towards large companies which should be noticed when assessing the size. Therefore, the size of the company in this study's dataset is under the highlighted focus when evaluating the modeling results of the size on defaulting probabilities because the dataset is mainly consisted of SME's. As for last mentioned CLCA, it has been taken into consideration due to the forementioned point that technology companies do not have similar inventories like on manufacturing industries. Hence, CLCA might work as more appropriate indicator of current liquidity than WCTA.

In case of high classification accuracy, the thesis could provide more evidence that logit models and financial ratios are still appropriate method to predict bankruptcy also when concentrating in technology companies. Although AIES models are improving and providing great results, financiers still need straightforward and easy-to-use methods in financing decision-making, while AIES models usually require more specification and teaching before they work properly.

3 Data and methodology

3.1 Data collection

Data used in the thesis was collected from Amadeus -database (Bureau van Dijk 2022). Collection was delimited to Nordic companies due to similar legislation regarding bankruptcy so different definitions of bankruptcy were not a topic to be considered more specifically. The study is processing specifically technology companies due to their important role in the current economy, so delimitation was conducted to cover only information technology industry described below. The classification system used was NACE (*The Statistical Classification of Economic Activities in the European Community*). Also, company's legal form was delimited to cover only private limited companies because of the rarity of bankruptcies of listed companies.

In Amadeus -database the oldest bankruptcies among tech companies occur in 2016, but 2016–2018 had data only from a very small number of tech companies. Thus, data collection could have been done by selecting companies which have defaulted in 2019–2021. However, the COVID pandemic is assumed to affect bankruptcy policies extraordinarily. For example, in Finland the amendment in Bankruptcy Act (721/2020) was valid in 2020 and 2021 which increased a threshold for a company to default. Nonetheless, the challenges of reviewed companies bankrupted in 2020 were markedly visible already in the statements from 2018 and 2019, so it was justified to take also the companies bankrupted in 2020 into the dataset albeit the amendments in legislation. Hence, the data collection focused to companies defaulted in 2019–2020. From every company selected, the financial statements data regarding profitability, solvency and liquidity was fetched from one and two years[5] prior the bankruptcy.

It is not unambiguous what kind of companies the term "technology" includes in industry classification because different technology-related companies are positioned under several industry sub-classes. One term used to mean the most innovative, intangible commodities producing companies which use a great amount of money into research and development is "a high-tech company" (Oakey 2010, 174). These high-tech

[5] Also, CHIN needed values of net income from three years prior the bankruptcy to get calculations for model 2.

companies would be interesting subject of research, but to acquire a coherent dataset is challenging, because there is no clear industry classification for these kinds of companies. In turn, concentrating to companies which are positioned under the class “information technology” is more coherent approach. International and widely used NACE classification has following classes which are representing information technology: *Computer programming, consultancy and related activities* (class 62) and *Information service activities* (class 63). These classes include mainly activities in programming, computer facilities management, data processing and web services. In this thesis the main goal is to review companies working with information technology and regarding the structure with low working capital and mainly software and services as products. Hence, it was eventually punctual to acquire the dataset for the review from NACE classes 62 and 63.

First, with the chosen criteria, 574 bankrupted companies were chosen in the dataset. Then, the comparison group of healthy companies was collected by exporting all the companies with the same criteria than the bankruptcy group, except the selection of a company’s status, which was now active companies. Because the selection covered over 30 000 companies, and the objective was to acquire balanced dataset, the random sample of 574 healthy companies was then taken from the data in Stata. Accordingly, the total amount of companies in the dataset was 1148.

This was a good starting point for the analysis. Next, the dataset was cleaned from observations missing focal values. Also, the most extreme outlier observations were deleted because they could have misled the modelling and form incorrect interferences. The interquartile range was used to create fences around the data and then defining the outliers which were outside the fences. In situations where many values were outside the fence in a quite dense cluster, the values were left in the dataset.

After cleaning the data, the final dataset consisted of 746 companies for the first model predicting the bankruptcy one year prior and 492 companies for the second model predicting the bankruptcy two years prior the bankruptcy. The reason for limited data of bankrupted companies is obvious: The main part of data consists of small and medium -sized companies which means that they have no similar reporting standards as large companies, not to mention listed companies which have very strict obligations considering the financial statements and their reporting. The statistics of the companies selected in the dataset are introduced in tables 2 and 3.

Table 2. Companies selected for the model 1.

	Bankruptcy		Non-bankruptcy	
Variable	*Mean*	*Std. dev.*	*Mean*	*Std. dev.*
SIZE (log((total assets(y-1) + total assets (y-2)) / 2))	1.5637	.7205	2.1623	.6849
TLTA (total liabilities / total assets)	1.0438	1.0449	.4394	.3781
WCTA (working capital / total assets)	.1018	.3105	.1446	.2169
CLCA (current liabilities / current assets)	1.4248	1.6578	.5564	.7768
NITA (net income / total assets)	-.3042	.8401	.0940	.2806
CFTL (cash flows from operations / total liabilities)	-.0128	.6913	.1707	.7246
INTWO (negative net income during two last years)	.2842	.4516	.1609	.3679
OENEG (total liabilities exceeding total assets)	.3378	.4736	.0456	.2088
CHIN (change in net income)	-.1362	.8299	.0343	.6232
n	373		373	

Differences in the ratio means between bankruptcy and non-bankruptcy companies can be seen in every ratio, and some are very clear. Basically, every difference is down the line with hypotheses, namely the healthy companies have better means in all respective ratios. The most prominent findings are that the size of bankruptcy companies is smaller, total liabilities divided by total assets and current liabilities divided by current assets are both larger, and net income divided by total assets is smaller (even negative) than non-bankruptcy companies. In turn, the differences are minor with the following variables: working capital divided by total assets, cash flows divided by total liabilities, change in net income, and dichotomous variable depicting negative net income during two last years.

Table 3. Companies selected for the model 2.

	Bankruptcy		Non-bankruptcy	
Variable	*Mean*	*Std. dev.*	*Mean*	*Std. dev.*
SIZE (log((total assets(y-1) + total assets (y-2)) / 2))	1.7215	.6343	2.1902	.7106
TLTA (total liabilities / total assets)	.7451	.4619	.4386	.3257
WCTA (working capital / total assets)	.1466	.3115	.1458	.2126
CLCA (current liabilities / current assets)	.8912	.6876	.5383	.5603
NITA (net income / total assets)	-.0642	.4776	.1197	.2906
CFTL (cash flows from operations / total liabilities)	.0339	.4430	.2191	.5857
INTWO (negative net income during two last years)	.1667	.3734	.1463	.3542
OENEG (total liabilities exceeding total assets)	.2033	.4032	.0244	.1546
CHIN (change in net income)	.0028	.8675	.0156	.6223
n	246		246	

As regards the second dataset, the means of non-bankruptcy companies are similar to the first dataset, but the means of bankruptcy companies differ somewhat. However, the directions of differences between bankruptcy and non-bankruptcy companies are identical to the first dataset. It is also prominent that the differences are only marginal with WCTA, CHIN and INTWO.

3.2 Research methods and models

The effect of company's specific financial ratios on the probability of bankruptcy are tested by logistic regression analysis, utilizing logit model. Logistic method was chosen since it is frequently used and tested in bankruptcy prediction research as mentioned in the chapter 2.2 and 2.3. Logit model is fundamentally suitable for using in classification problem, and it eliminates the weaknesses of MDA method's assumptions. There is always a need for simple methods to be used by practical actors who can easily acquire the financial statements data, and these logistic models do not require long-lasting testing and teaching such as deep learning AIES models.

The logit model is complementing regression models in a case where the dependent variable is not continuous but a categorial, qualitative variable. The dependent variable in logistic regression analysis can be dichotomous (with two categories) or, polytomous nominal or polytomous ordinal (Menard 2010, 2). In this thesis, the dependent variable is dichotomous, answering to the question whether a company is classified as a bankrupt or a healthy one. While the logit model is having its own characteristics, it has much common with the linear regression model: they both are designed for the analysis of experimental data as a tool to estimate relationships between independent and dependent variables (Cramer 1991, 5-8).

Altogether, the logit model is modelling the odds of being in one category of a dependent variable as a function of independent variables (Menard 2010, 24). If the objective is to evaluate probabilities of dependent variable in belonging to either class 1 or 0, and the specification of probability needs to be limited from the range 0 to 1, we need a sigmoid curve which flattens from both ends. The logistic function is such a function and can be noted as

$$P(X) = \frac{\exp(\alpha + \beta X)}{1+\exp(\alpha + \beta X)} = 1 + \exp\,(\alpha + \beta X)^{-1} \tag{1}$$

Now, when specifically estimating probabilities, the probability function is defined as

$$P(Z) = \exp Z \,/\, (1 + exp\ Z) \tag{2}$$

$$Q(Z) = 1 - P(Z) = P(-Z) \tag{3}$$

Finally, the inverse transformation of P(Z) is defined as a log odds ratio R(Z)

$$R(Z) = \log \frac{P(Z)}{1-P(Z)} \tag{4}$$

The logistic function follows the sigmoid curve (figure 1), and its midpoint is P(0) = 0.5. The inversed transformation, the logit curve, is shown in figure 2.

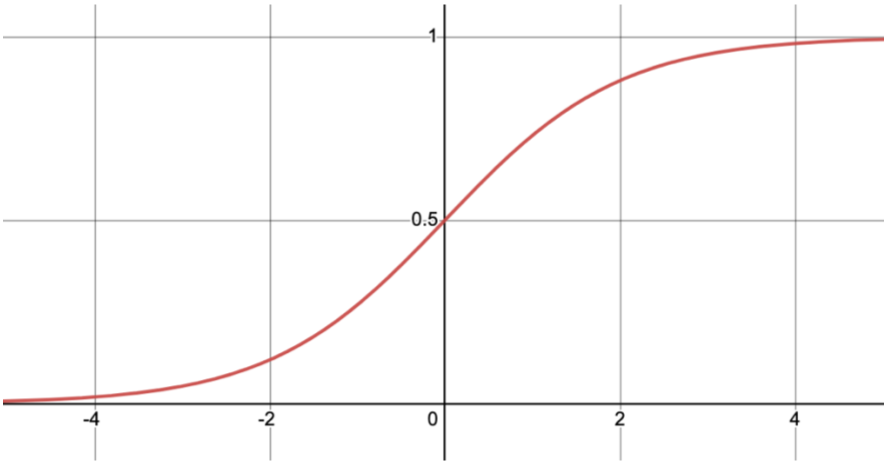


Figure 1. The logistic curve P(Z).

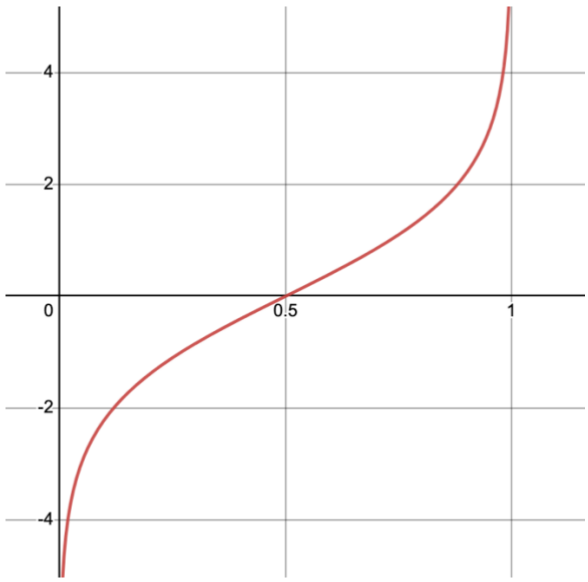


Figure 2. The logit curve R(Z).

Methodologically, some are making divergence between logit analysis as a method for dealing with the categorial variables and logistic regression analysis dealing with the continuous variables, but this is only factitious separation: from more comprehensive perspective, logistic regression is mixed from logit model and OLS regression model and is dealing equally with both continuous and dichotomous variables (Menard 2010, 40). The logit transformation, log odds ratio R(Z), has many necessary properties of a linear regression model: it is linear in its parameters, may range from $-\infty$ to ∞ and may be continuous (Hosmer & Lemeshow 1989, 7). Hence, both functions were considered and presented mathematically.

3.3 Explanatory variables

Models are developed following the same explanatory ratios as in Ohlson's original O-score -model. The dependent variable in such a model is classified, dichotomous variable: whether a company is healthy or defaulting on the upcoming accounting periods. Two different models were composed, of which first is classifying companies based on company's financial ratios one year prior the bankruptcy, and second based on ratios two year prior the bankruptcy.

The predictive independent variables in this thesis are based on the same predictors as in Ohlson's (1980) logistic model explained in the chapter 2.3. The modifications were made to SIZE (size of the company) and CFTL (cash flows divided by total liabilities; originally FUTL and now CFTL) as explained in the chapter 2.4.

3.3.1 Profitability variables

NITA (net income divided by total assets) is representing profitability via company's relative net income; how well the company is using its assets to generate profits from its operations. NITA is practically the same ratio as return on assets (ROA) which is commonly used in comparing the profitability of different financial alternatives. The higher profits the company is able to generate, the better it can maintain healthy growth. CHIN (change in net income) is another variable assessing the profitability, but it is doing that by indicating the intensity of change in net income.

The third variable measuring profitability is INTWO (negative net income during two last years): if the company has two consecutive years with negative income, the variable is emphasizing the possible challenges in company's operations. In turn, many high-tech companies are operating with strongly negative net income for several years until they possibly get acquired by a larger company. However, the dataset in this thesis is not mainly consisted of these kinds of high-tech companies with a fast exit-strategy but more of regular software and programming -based firms.

Lastly, the fourth variable measuring the profitability is CFTL (cash flows divided by total liabilities) which is based on cash flows rather than net income. It is also differing from other variables above regarding the denominator, which is setting the relation to liabilities rather than assets. Hence, it determines how long it would take to repay all the company's debt if it uses all the cash flows to repayments. If the cash flows are low compared to total debt, the repayments can weaken the financial position of the company.

3.3.2 Solvency variables

TLTA (total liabilities divided by total assets) measures the total debt leverage of the company. The higher the ratio is, the stronger debt leverage company is using. This debt leverage is in the core of optimal capital structure research, where the objective has been, for a long time, to find the balance between the utilities of tax shield obtained from debt financing and increased bankruptcy costs as an outcome from increased use of debt (Bradley, Jarrell & Kim 1984). As a predictive variable it is clearly presumed that increased use of debt leverage is increasing the probability of bankruptcy. OENEG (total liabilities exceeding total assets) is on the other hand an amendment for a discontinuity of TLTA. Ohlson (1980) argued that a company, which has a negative book value, is a special case when assessing the probability of failure. Hence, this variable is needed in addition to TLTA.

3.3.3 Liquidity variables

An obvious way to represent company's short-term ability to cover its liabilities is to compare its current liabilities and current assets, which CLCA (current liabilities divided by current assets) is doing. If CLCA is over 1, the company has larger current liabilities than it has current assets, in which case the short-term solvency might be deteriorated.

WCTA (working capital divided by total assets) is also representing liquidity, but with a slightly different approach than CLCA. Working capital is meaning the difference between current assets and current liabilities, so it is positive when current assets are higher than liabilities. In WCTA the difference is divided by total assets, when it is possible to evaluate how large part of the assets are usable in covering short-term obligations. If the ratio is high, the company has a good cover to bear any unexpected obligations, but on the other hand, working capital can also be too high because of lost investment opportunities.

4 Development and analysis of prediction models

As a tool in analysis, Stata was used to perform the analysis of dataset, modelling, and analysis of models. Before the modelling of variables with logistic regression analysis, the independent variables were reviewed more specifically. First, the correlations between variables were checked (Table 4).

Table 4. Correlation matrix of independent variables in model 1.

	SIZE	TLTA	WCTA	CLCA	NITA	CFTL	INTWO	OENEG	CHIN
SIZE	1.0000								
TLTA	-0.1677	1.0000							
WCTA	0.0203	-0.0478	1.0000						
CLCA	-0.0630	0.6075**	-0.1486	1.0000					
NITA	0.1844	-0.4465*	0.0727	-0.2550	1.0000				
CFTL	0.0443	-0.1038	0.0784	-0.0979	0.2805	1.0000			
INTWO	-0.0092	0.2654	-0.1068	0.2207	-0.3184	-0.2208	1.0000		
OENEG	-0.1791	0.7047**	-0.0668	0.4446*	-0.3889	-0.1193	0.2635	1.0000	
CHIN	-0.0031	-0.1024	-0.0011	-0.0959	0.4733*	0.1878	-0.0243	-0.1738	1.0000

** Moderate correlation (0.4–0.6), ** Strong correlation (> 0.6)*

TLTA is strongly correlating with CLCA which is presumptive while they are both dealing with a ratio of liabilities and assets, one with the total and another with short term values. TLTA (and CLCA) is also strongly (moderately) correlating with OENEG which is measuring dichotomously if total liabilities are exceeding total assets or not. OENEG forms a challenge due the strong correlation with TLTA and the exact same financial statements values the ratios are using, why it could have been left out of the model to be able to focus on the base effect of TLTA. However, the models below show that the inclusion of both variables is still giving quite relevant results and the tests without OENEG did not really change the outcome.

It is noticeable that NITA is moderately correlating with TLTA and CHIN. The correlation between profitability variables is also presumptive while they are using similar values from financial statements. The negative correlation between variables considering profitability and solvency, here NITA and TLTA, is also presumptive because smaller companies have better opportunities to make greater relative net income. For example, Switzer (2012) has demonstrated that small-cap companies perform generally better than large companies over long time horizon. Weak correlations (0.2–0.4) are not considered here specifically. Conclusively, the assumption in logistic regression analysis of avoiding multicollinearity is not wholly, but almost fulfilled. As for the model 2, the correlations were very closely the same between the variables in model 2, so the same premises can be used in model 2.

4.1 Fitting of model 1

Model 1 was composed with the ratios calculated from financial statements one year prior the bankruptcy. The dataset of 373 bankrupted companies and 373 healthy companies were divided into two groups: 70 % of the data (524) was used as a training data to compose the model, and then, 30 % (222) was used to test out-of-sample accuracy of the model.

In Table 5 are represented the number of observations, log likelihood value, likelihood ratio (LR) chi-square test value, p-value of the model, pseudo-R-squared value, and estimates, z-values, p-values and confidence intervals (for odds ratios) for every predictor.

Table 5. Coefficients and z-values of predictors in model 1.

Model 1						
Obs	524					
Log likelihood	-260.9202					
LR chi2(9)	204.57					
Chi2	p < 0.0001					
Pseudo R2	0.2817					
	*Odds ratio**	*Estimate*	*z*	*p > z*	*95 % conf. Interval*	
SIZE	.3000	-1.2040	-7.38	p<0.001	.2179	.4130
TLTA	4.8009	1.5688	3.78	p<0.001	2.1290	10.8261
WCTA	.8269	-.1900	-0.46	0.645	.3681	1.8577
CLCA	1.3402	.2928	2.18	0.029	1.0230	1.7439
NITA	.5215	-.6510	-2.03	0.043	.2778	.9790
CFTL	.7671	-.2651	-1.46	0.145	.5371	1.0957
INTWO	.6066	-.4998	-1.65	0.100	.3345	1.1010
OENEG	1.1967	.1796	0.38	0.703	.4761	3.0076
CHIN	1.0084	.0084	0.05	0.961	.7229	1.4066
const	2.8697	1.0542	3.06	0.002	1.4597	5.6417

** Odds ratio =* $e^{Estimate}$

The model was statistically significant with the p-value < 0.0001 and regarding to pseudo-R-squared value of 0.2817 the model is explaining the probability of bankruptcy somewhat. Statistically significant predictors were SIZE, NITA, TLTA and CLCA.

In addition to coefficients for each variable, in table 5 are also introduced odds ratios which are easier to understand and interpret. The change in odds can be interpreted in percentage from the baseline of odds ratio of 1 (change-% = odds ratio - 1). If the odds ratio is lower than 1, the probability of bankruptcy is decreasing and if the odds ratio is greater that 1, the probability of bankruptcy is increasing.

According to results, holding other independent variables constant, the odds of the company going bankrupt decreased by 70 % for each additional unit the logarithm of company's size is grown, and decreased by 48 % for each additional unit the net income divided by total assets is grown. The odds of the company going bankrupt increased by 380 %[6] for each additional unit the total liabilities divided by total assets is grown and increased 34 % for each additional unit the current liabilities divided by current assets is grown.

The direction of the effect was similar with these significant predictors as the hypotheses were set. In terms of a more general presentation, the forementioned results are following:

- When the size of the company is increasing, the probability of bankruptcy is decreasing
- When net income compared to total assets are increasing, the probability of bankruptcy is decreasing
- When debt leverage is increasing, the probability of bankruptcy is increasing
- When current liabilities compared to current assets are increasing, i.e. the liquidity of the company is deteriorating, the probability of bankruptcy is increasing

Although the model was fitting according to the chi-square probability, the goodness-of-fit was also tested with Hosmer Lemeshow -test. Usually Hosmer Lemeshow -test is based on the percentile-type grouping, where g = 10 groups (deciles of risk), if not stated

[6] The change of one unit in proportionated predictive variables such as TLTA, requires a tremendous change in accounting parameters, so the large changes in odds, consequently, are reasonable.

otherwise (Hosmer & Lemeshow 1989, 142). In the model 1, the chi-square value was 29.57 and p-value was 0.003, so with the risk level α = 0.05 Hosmer Lemeshow -test resulted that the model is not fitting. This makes the evaluation of the model's fit somewhat contradictory. Probably, the main reason behind the conflicting results of these two tests is too small dataset used to build the model.

Reviewing model's classification accuracy is in the most important role in the evaluation of the model in the particular context of classification problem. It can be seen from the ROC-curve (Appendix 1), that area under the curve (AUC) was 0.8508. This indicates that the model has good prediction power, but AUC is not directly telling the accuracy of the model. When reviewing model's classification tables, with the default cutoff value of 0.50, overall classification accuracy was 77.86 % (Appendix 3). Because classification is about trade-off between false positive and true positive rate, the cutoff value is used to adjust the predictive performance. With the sensitivity-specificity -graph (Appendix 2) the optimal cutoff-point was found to be in 0.42. Here, the model classified correctly 78.82 % of the companies. The results with optimal cutoff value are showed in table 6.

Table 6. Classification table of model 1 with the cutoff of 0.42.

	True		
Classified	B	NB	Total
+	**216**	**66**	**282**
-	**45**	**197**	**242**
Total	**261**	**263**	**524**
Classified + if predicted Pr(B) >= .42			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**82.76 %**
Specificity		Pr (- |NB)	**74.90 %**
Positive predictive value		Pr (B | +)	**76.60 %**
Negative predictive value		Pr (NB| -)	**81.40 %**
False + rate for true NB		Pr (+ |NB)	**25.10 %**
False - rate for true B		Pr (- | B)	**17.24 %**
False + rate for classified +		Pr (NB| +)	**23.40 %**
False - rate for classified -		Pr (B | -)	**18.60 %**
Correctly classified			**78.82 %**

In addition, one perspective in classification is to try to minimize type-2 errors, i.e. to minimize the classification of bankruptcy companies into healthy ones. In the table above this value is noted as “false - rate for true D”. For example, one possible threshold can be to lower the type-2 error under 10 % of classifications. In this example, the cutoff needed to be 0.32 (Appendix 4), in which case the overall classification accuracy was 75.19 %. The same point can be seen via sensitivity: now the sensitivity of the model rose to 90.06 %, when only 1 from 10 unhealthy companies are incorrectly classified into healthy ones. This matter is interesting especially from the aspect of financier’s credit risk management. When type-2 errors are minimized, the risks are also decreased. Of course, at the same time, the false positive hits are increasing, i.e. the healthy companies are classified as unhealthy ones, in which case the financier is missing possibly profitable financing options. Hence, adjusting of the model is a trade-off game.

Along with the accuracy, it is also possible to assess the quality and performance of the prediction model with F1-score. F1-score is defined as a harmonic mean of precision (*positive predictive value*) and recall (*sensitivity*) (Huang et al. 2015). F1-score is not as easy to interpret as accuracy while it is a blend of two indicators, but it takes into account imbalances in classes and if the model is classifying serious number of false negatives. In the thesis’ dataset the classes are balanced, but still, it is reasoned to interpret the performance with the F1-score to pay more attention to false negatives also. As an equation it is expressed as

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

F1-score of the model 1 with the cutoff of 0.42 is 79.56 %. It is very close to the accuracy (78.82 %), so the performance appears to be satisfactory also with the F1-score.

Lastly, the model 1 was tested with the secondary sample. The sample was 30 % (222) of collected companies which were left outside from model fitting process. The odds for companies were calculated with the estimates fitted above, and prediction performance was tested with different cutoff values by iteration. The best overall accuracy was achieved with the cutoff value of 0.39. Secondly, the model was iterated to reach sensitivity of 90 %, which was achieved with the cutoff value of 0.28 (Appendix 5). The results with the cutoff value of 0.39 are presented in the table 7.

Table 7. Classification table of out-of-sample testing for model 1.

Classified	True B	True NB	Total
+	**90**	**26**	**116**
-	**22**	**84**	**106**
Total	**112**	**110**	**222**
Classified + if predicted Pr(B) >= .39			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**80.36 %**
Specificity		Pr (- |NB)	**76.36 %**
Positive predictive value		Pr (B | +)	**77.59 %**
Negative predictive value		Pr (NB| -)	**79.25 %**
False + rate for true NB		Pr (+ |NB)	**23.64 %**
False - rate for true B		Pr (- | B)	**19.64 %**
False + rate for classified +		Pr (NB| +)	**22.41 %**
False - rate for classified -		Pr (B | -)	**20.75 %**
Correctly classified			**78.38 %**

4.2 Fitting of model 2

Model 2 is predicting the probability of bankruptcy prior two years, so it was composed with the ratios calculated from financial statements two year prior the bankruptcy. Other specifications were the same with the model 1, except the dataset was not divided into subsets of training data and out-of-sample testing data. The number of observations available was lower than for model 1, so the number would have been too low if divided into subsets. Hence, out-of-sample testing was not possible for model 2. This is still not a major problem because the focus is on the model 1, for which the out-of-sample testing was able to be made. The statistics of model 2 are presented in the table 8.

Table 8. Coefficients and z-values of predictors in model 2.

Model 2						
Obs	492					
Log likelihood	-262.3095					
LR chi2(9)	157.44					
Chi2	p < 0.0001					
Pseudo R2	0.2308					
	Odds ratio	*Estimate*	*z*	*p > z*	*95 % conf. Interval*	
SIZE	.2891	-1.2410	-6.99	p<0.001	.2042	.4094
TLTA	5.0192	1.6133	3.29	0.001	1.9182	13.1333
WCTA	1.5650	.4479	1.04	0.299	.6716	3.6471
CLCA	1.6950	.5277	2.32	0.020	1.0858	2.6460
NITA	.3009	-1.2009	-2.59	0.010	.12139	.7460
CFTL	.6511	-.4292	-1.74	0.082	.4011	1.057
INTWO	.4893	-.7148	-2.09	0.036	.2507	.9550
OENEG	1.5064	.4097	0.71	0.476	.4884	4.6463
CHIN	1.5283	.4242	2.45	0.014	1.0881	2.1467
const	3.5031	1.2536	3.44	0.001	1.7143	7.1581

The model was statistically significant with the p-value < 0.0001 and regarding to pseudo-R-squared value of 0.2308 the model is explaining the probability of bankruptcy somewhat. Statistically significant predictors were SIZE, NITA, TLTA and CLCA like in the model 1, and in addition also INTWO and CHIN were significant.

According to the results, the effect of SIZE and TLTA on probability of bankruptcy was almost the same as in the model 1. The effect of NITA was parallel but somewhat larger (decrease of 70 % with each additional unit) and the effect of CLCA was also parallel but somewhat larger (increase of 70 % with each additional unit) than in model 1. Hence, the direction of the effect was similar with these significant predictors as the hypotheses were set, except INTWO and CHIN. In terms of a more general presentation, the forementioned results, which were distinct from the model 1, are following:

- When the change in net income is increasing, the probability of bankruptcy is increasing
- When the net income has been negative for two years, the probability of bankruptcy is decreasing

Lastly mentioned effect of INTWO was peculiar and it fights against common sense. It seems that the multicollinearity might be affecting here. However, the correlations between INTWO and other variables were not severe, and when tested to drop out NITA and CHIN which were also derivated from net income, the effect of INTWO was still negative. Hence, the limitations of the size of the dataset are good to keep in mind. The same peculiarity was with CHIN, because positive increase in net income should be a positive indicator. The reason behind this observation might be the large range in variation of net income, since the dataset consists mainly of small companies. Here, the large changes in income might indicate challenges. What was surprising, the Hosmer Lemeshow -test gave way better results than with the model 1. In the model 2, the chi-square value was 6.96 and p-value was 0.5408, so with the risk level α = 0.05 Hosmer Lemeshow -test resulted that the model is fitting.

The classification accuracy of model 2 turned out to be only slightly weaker than model 1. From the ROC-curve (Appendix 6) AUC was 0.8058 which indicates that the model has quite good prediction power. When reviewing model's classification tables, with the default cutoff value of 0.50, overall classification accuracy was 73.78 % (Appendix 8). With the sensitivity-specificity -graph (Appendix 6) the optimal cutoff-point was found to be in 0.55. Here, the model classified correctly 75.00 % of the companies. The results with optimal cutoff value are showed in table 9. As for the F1-score, it was 73.66 % with the particular cutoff value.

Table 9. Classification table of model 2 with the cutoff of 0.55.

	True		
Classified	B	NB	Total
+	**172**	**49**	**221**
-	**74**	**197**	**271**
Total	**246**	**246**	**492**
Classified + if predicted Pr(B) >= .55			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**69.92 %**
Specificity		Pr (- |NB)	**80.08 %**
Positive predictive value		Pr (B | +)	**77.83 %**
Negative predictive value		Pr (NB| -)	**72.69 %**
False + rate for true NB		Pr (+ |NB)	**19.92 %**
False - rate for true B		Pr (- | B)	**30.08 %**
False + rate for classified +		Pr (NB| +)	**22.17 %**
False - rate for classified -		Pr (B | -)	**27.31 %**
Correctly classified			**75.00 %**

Minimizing type-2 errors was tested in the same way as with the model 1. Threshold was set to 10 % of false negatives for true bankruptcies, in which case the the cutoff needed to be 0.28 (Appendix 9). The sensitivity was then 90.56 %. The overall classification accuracy was then 66.26 %, which was slightly too low because the false positives, i.e. the healthy companies classified as unhealthy ones, increased to the total of 58.13 % which is basically too high.

4.3 Testing of alternative predictors

Tanaka (2019) argued that net assets turnover (net sales / average total assets) seems to be particularly significant predictor for information and communication industry. Hence, net assets turnover was tested in thesis' modelling, but after testing it as an independent variable in the model, it turned out that it didn't affect the model's predictive power. Although variable was statistically significant (p-value 0.002), its coefficient was only 0.9998 and the percent of correctly classified companies didn't change. Also, it had not any effect to goodness-of-fit, as for pseudo R2 or Hosmer-Lemeshow test did not gave different results.

Another alternative predictor tested from global ratios was credit days period. Same results were occurred with credit days period as predictive variable than with net assets turnover, but it wasn't even statistically significant. Due to the inefficiency of these alternative predictors, the final models were composed with only original Ohlson's predictors, with the minor adjustments to original predictors described above.

5 Discussion and conclusions

In this thesis, two predictive models were composed: model 1 which predicts if a company is defaulting or not one year prior and model 2 two years prior the possible bankruptcy. Model 1 was statistically significant, and it was somewhat explaining the probability of bankruptcy, but the results should be evaluated carefully because goodness-of-fit was not unambiguous. Model 2 was statistically significant, and it was somewhat explaining the probability of bankruptcy. The goodness-of-fit was better than in model 1 but testing with secondary sample was not possible for model 2 because of the lack of adequate data.

The first research question was considering how accurately the bankruptcy of Nordic private technology companies can be predicted based on financial ratios. The empirical tests resulted that the prediction of bankruptcy is possible at least on satisfactory level. The second research question was considering which of the financial ratios are predicting best the bankruptcy. The best ratios were the size of the company, total liabilities divided by total assets, net income divided by total assets and current liabilities divided by current assets. Hence, the hypothesis of best predictors was fulfilled. It was also satisfactory that the best ratios are presenting widely every sub-section of company's financial performance: profitability, solvency, and liquidity. Total liabilities divided by total assets is particularly effective predictor of the failure among technology companies, so it should be observed especially for small businesses in credit risk management.

The hypotheses considering the effects of different financial ratios were realized in accordance with the expectations regarding statistically significant predictors, except the change in net income and net income being negative for two years in the second model. However, the hypothesis considering the prediction accuracy was not exactly achieved because the prediction accuracy one year prior the bankruptcy remained on the level of 80 %. Anyway, the positive finding was that the accuracy with secondary sample (78.38 %) was almost as good as with in-sample data. Previous studies in recent years have achieved similar out-of-sample prediction accuracies with financial statements data and logistic regression (79,23 % for Cultrera & Brédart, 2016 and 90.11

% for Alaminos, Del Castillo & Fernández, 2016), so the results of this thesis are tolerably in line with the recent studies.

The most obvious challenge in the modelling was relatively small dataset used to develop the models. Although the absolute size of the dataset for logistic regression was satisfactory in the thesis, according to Appiah, Chizema and Arthur (2014) the bankruptcy prediction models need per se larger samples to achieve robust results. Now, the data acquired from Amadeus database was temporally limited because the collection targeted bankrupted companies whose information were partly deleted from the database. Hence, the results should be interpreted carefully. If one has later demands for similar data from wider time frame, Orbis, which is also maintained by Bureau van Dijk, is the database which could provide more comprehensive data than Amadeus for bankrupted companies.

The models did not utilize market data or macroeconomic variables which could have improved the prediction performance. For example, stock pricing and ratios associated with it could be one way to improve bankruptcy prediction models. However, this thesis considered private small and medium -sized companies whose stocks are not priced on open stock markets. In addition, it is reasoned that stock pricing implicates largely financial statements data, so the usage of financial statements data is well formed basis for prediction as regarding bankruptcies. As for macroeconomic factors such as inflation and interest rates, they could be important predictors to include into predictive models. Further, the corporate governance structure and its usage in predictors, such as board independence and director experience measures, has been argued to work efficiently in bankruptcy prediction (Aziz & Dar 2006; Almaskati et al. 2021). Corporate governance could act in an important role also with technology companies. As regards the type of the predictive model, the same approach of reviewing technology companies should also be tested with AIES models, such as regression trees and neural networks, since Almaskati et al. (2021) have been lately found them to be more precise than classical statistical methods, even with secondary datasets.

In overall, the prediction of bankruptcy with financial statements data is restricted by the fact that annual reports are not available publicly at the end of the fiscal year which can restrict the application of predictive model if evaluating the probability prior one year. Therefore, it was appropriate to also develop a model which utilizes financial statements from two years prior the bankruptcy. Further, the improvement of models to precisely detect the increased bankruptcy risk even earlier than one or two years could be very valuable for those who assess credit risks. Alaminos, Del Castillo and Fernández (2016) achieved the prediction accuracy of 87.22 % three years prior the bankruptcy with logistic regression and European companies as a secondary dataset, so it seems entirely possible to reach as precise results at an earlier point of time.

Reliability of the results of this thesis was weakened by relatively small dataset used in model development, so similar estimations should be tested with larger datasets. Goodness-of-fit of model 1 was indistinct, but with model 2 the fit was clearly good. In turn, the usage of out-of-sample testing for model 1, i.e. secondary data testing, improved amply the reliability. The lack of using secondary samples is criticized by Aziz and Dar (2006), as many models in journal articles have not been tested with secondary samples.

Ohlson (1980) mentioned that the O-score -model was not directly made for non-manufacturing companies, why the use of his original predictors needed to be considered cautiously. Eventually, the results of this thesis indicate that the major ratios described above still work quite well in failure prediction, also for small and medium -sized technology companies.

The thesis provides simple models for practical actors, such as financiers and company managers, to evaluate the probability of bankruptcy one and two years prior the possible occurrence. The use of models utilizing logistic regression is a suitable way to evaluate the credit risks because the user can adjust the cutoff values to match organization's defined risk levels like demonstrated in the fourth chapter. Especially, the approach of this thesis of reviewing technology companies is a relevant perspective to develop further, while technology industry is in a constantly growing state.

References

Alaminos, D., Del Castillo, A. & Fernández, M. Á. (2016) A Global Model for Bankruptcy Prediction. *PLoS ONE*, 11, 11. DOI: 10.1371/journal.pone.0166693.

Almaskati, N., Bird, R., Yeung, D. & Lu, Y. (2021) A horse race of models and estimation methods for predicting bankruptcy*. Advances in Accounting*, 52. DOI: 10.1016/j.adiac.2021.100513.

Altman, E. I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23, 4, 589–609.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977) ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1, 1, 29-54.

Altman, E. I. (1993) Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy, 2nd Ed. New York: John Wiley and Sons.

Altman, E.I. & Sabato, G. (2007) Modelling credit risk for SMEs: evidence from the US market. *Abacus*, 43, 3, 332-357.

Andersson, K., G. & Hansen-Nord, A. (2021) Corporate insolvency & restructuring report 2021: Denmark. IFLR. Published on 30 September 2021. [Accessed 28 May 2022]. Available at: https://www.iflr.com/article/b1trzmw00v4vxm/corporate-insolvency-amp-restructuring-report-2021-denmark

Appiah, K. O., Chizema, A. & Arthur, J. (2014) Predicting corporate failure: a systematic literature review of methodological issues*. International Journal of Law and Management,* 57, 5, 461-485.

Asay, M. (2022) Why tech companies have venture capital funds. [Online article]. [Accessed 19 June 2022]. Available at: https://www.infoworld.com/article/3651502/why-tech-companies-have-venture-capital-funds.html

Aziz, M. & Dar, H. (2006) Predicting corporate bankruptcy: Where we stand? *Corporate Governance,* 6, 1, 18–33.

Balcaen, S. & Ooghe, H. (2006) 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review,* 38, 1, 63–93.

Barboza, F., Kimura, H. & Altman, E. (2017) Machine learning models and bankruptcy prediction. *Expert Systems With Applications*. 83, 405-417.

Blum, M. P. (1974) Failing company discriminant analysis. *Journal of Accounting Research*, 12, 1, 1–25.

Boritz, J. E, & Kennedy, D. B. (1995) Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications*, 9, 4, 503–512.

Bradley, M., Jarrell, G. A., & Kim, E. H. (1984) On the Existence of an Optimal Capital Structure: Theory and Evidence. *The Journal of Finance*, 39, 3, 857-878.

Brédart, X (2015) Bankruptcy Prediction Model: The Case of the United States. *International Journal of Economics and Finance*, 6, 1–7.

Cramer, J. S. (1991) The logit model: An introduction for economists. London: Edward Arnold.

Cultrera, L., & Brédart, X. (2016) Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance,* 15, 1, 101–119.

Dakovic, R., Czado, C. & Berg, D. (2010) Bankruptcy prediction in Norway: a comparison study. *Applied Economics Letters*, 17, 17, 1739-1746.

Drehmann, M., & Juselius, M. (2014) Evaluating Early Warning Indicators of Banking Crisis: Satisfying Policy Requirements. *International Journal of Forecasting*, 30, 3, 759–80.

EMCC (2021) Sweden: Rescue procedures in insolvency. [Accessed 28 May 2022]. Available at: https://www.eurofound.europa.eu/observatories/emcc/erm/legislation/sweden-rescue-procedures-in-insolvency

Fantazzini, D., Figini, S. (2009) Random Survival Forests Models for SME Credit Risk Measurement. *Methodology and Computing in Applied Probability*, 11, 29–45.

Hogan, T. & Hutson, E. (2005) Capital structure in new technology-based firms: Evidence from the Irish software sector. *Global Finance Journal*, 15, 3, 369-387.

Hosmer, D. W. & Lemeshow, S. (1989) Applied Logistic Regression. New York: Wiley.

Huang, H., Xu, H., Wang, X. & Silamu, W. (2015) Maximum F1-Score Discriminative Training Criterion for Automatic Mispronunciation Detection. *IEEE/ACM transactions on audio, speech, and language processing*. 23, 4, 787–797

Laitinen, E. K., & Laitinen, T. (1998) Misclassification in bankruptcy prediction in Finland: human information processing approach. *Accounting, Auditing & Accountability Journal*, 11, 2, 216-244.

Lawrence, J. R., Pongsatat, S. & Lawrence, H. (2015) The Use of Ohlson's O-Score for Bankruptcy Prediction in Thailand. *The Journal of Applied Business Research*, 31, 6, 2069–2078.

Menard, S. (2010) Logistic Regression: From Introductory to Advanced Concepts and Applications. London: SAGE.

Mirzaei, M., Ramakrishnan, S. & Bekri, M (2015) Corporate Default Prediction with Industry Effects: Evidence from Emerging Markets. *International journal of economics and financial issues*, 6, 3S, 161-169.

Oakey, R. P (2010) New Technology-Based Firms in the New Millennium. Vol. 8. Amsterdam: Pergamon.

Ohlson, J. A. (1980) Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18, 109–131.

Shumway, T. (2001) Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business,* 74, 1, 101-124.

Sætermo, H (2022) The Norwegian bankruptcy proceedings in a nutshell. Published on 3 February 2022. [Accessed 28 May 2022]. Available at: https://businesslaw.no/articles/corporate-recovery-insolvency/the-norwegian-bankruptcy-proceedings-in-a-nutshell/

Sigurðsson, O. A. (2014) Restructuring and insolvency in Iceland: overview. Thomson Reuters Practical Law. [Accessed 28 May 2022]. Available at: https://uk.practicallaw.thomsonreuters.com/0-591-2866?transitionType=Default&contextData=(sc.Default)&firstPage=true

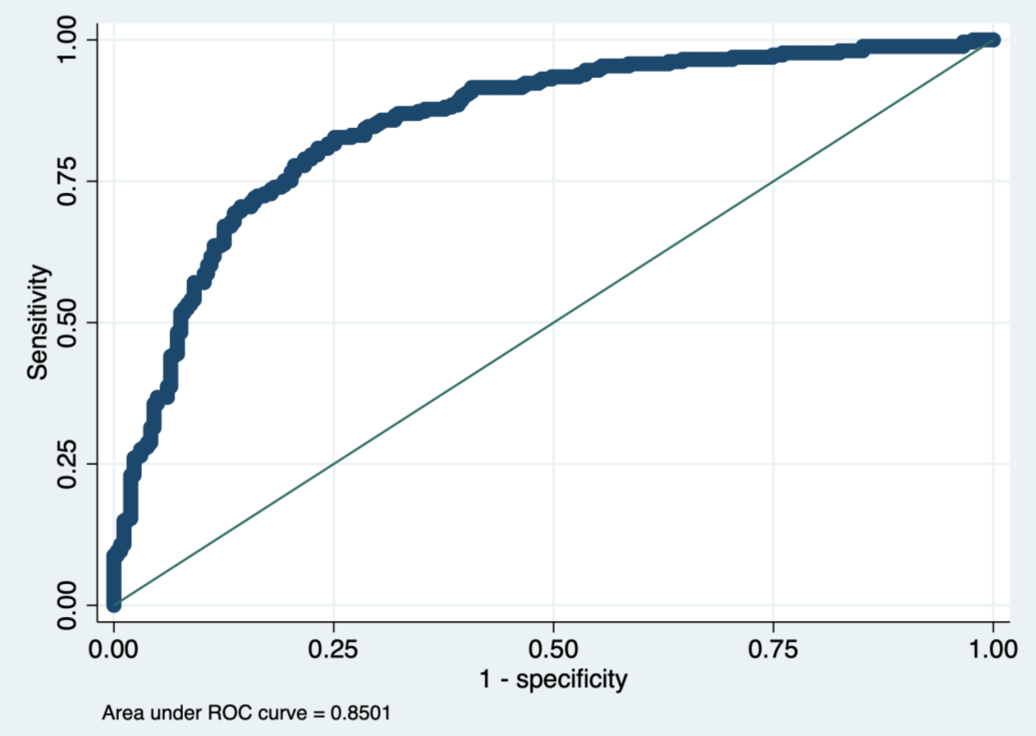
Switzer, L. N. (2012) The Relative Performance of Small Cap Firms and Default Risk across the Business Cycle: International Evidence. *International journal of business*. 17, 4, 379-396.

Tanaka, K., Higashide, T, Kinkyo, T. & Hamori, S. (2019) Analyzing industry-level vulnerability by predicting financial bankruptcy. *Economic inquiry,* 57, 4, 2017–2034.

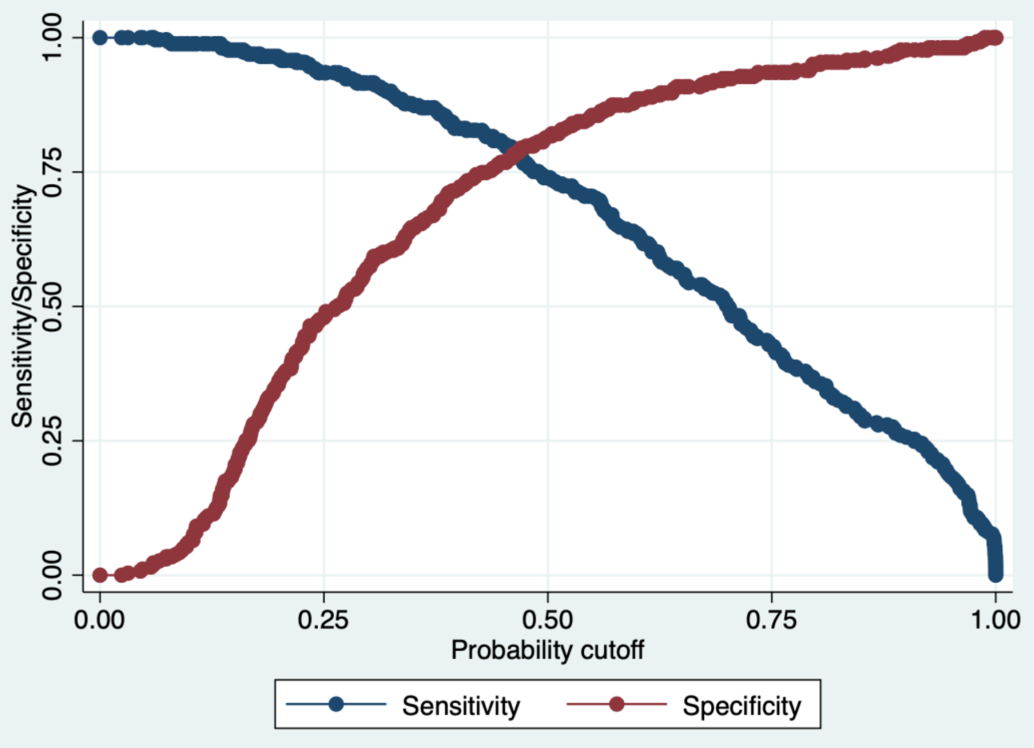
Yazdanfar, D. (2011) Predicting bankruptcy among SMEs: evidence from Swedish firm-level data. *International Journal of Entrepreneurship and Small Business,* 14

Appendices

Appendix 1. ROC-curve of model 1.



Appendix 2. Sensitivity/specificity -graph of model 1.



Appendix 3. Classification table of model 1 with the default cutoff of 0.50.

Classified	True B	True NB	Total
+	**193**	**48**	**241**
-	**68**	**215**	**283**
Total	**261**	**263**	**524**
Classified + if predicted Pr(B) >= .50			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**73.95 %**
Specificity		Pr (- |NB)	**81.75 %**
Positive predictive value		Pr (B | +)	**80.08 %**
Negative predictive value		Pr (NB| -)	**75.97 %**
False + rate for true NB		Pr (+ |NB)	**18.25 %**
False - rate for true B		Pr (- | B)	**26.05 %**
False + rate for classified +		Pr (NB| +)	**19.92 %**
False - rate for classified -		Pr (B | -)	**24.03 %**
Correctly classified			**77.86 %**

Appendix 4. Classification table of model 1 with the cutoff of 0.32.

Classified	True B	True NB	Total
+	**235**	**104**	**339**
-	**26**	**159**	**185**
Total	**261**	**263**	**524**
Classified + if predicted Pr(B) >= .32			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**90.04 %**
Specificity		Pr (- |NB)	**60.46 %**
Positive predictive value		Pr (B | +)	**69.32 %**
Negative predictive value		Pr (NB| -)	**85.95 %**
False + rate for true NB		Pr (+ |NB)	**39.54 %**
False - rate for true B		Pr (- | B)	**9.96 %**
False + rate for classified +		Pr (NB| +)	**30.68 %**
False - rate for classified -		Pr (B | -)	**14.05 %**
Correctly classified			**75.19 %**

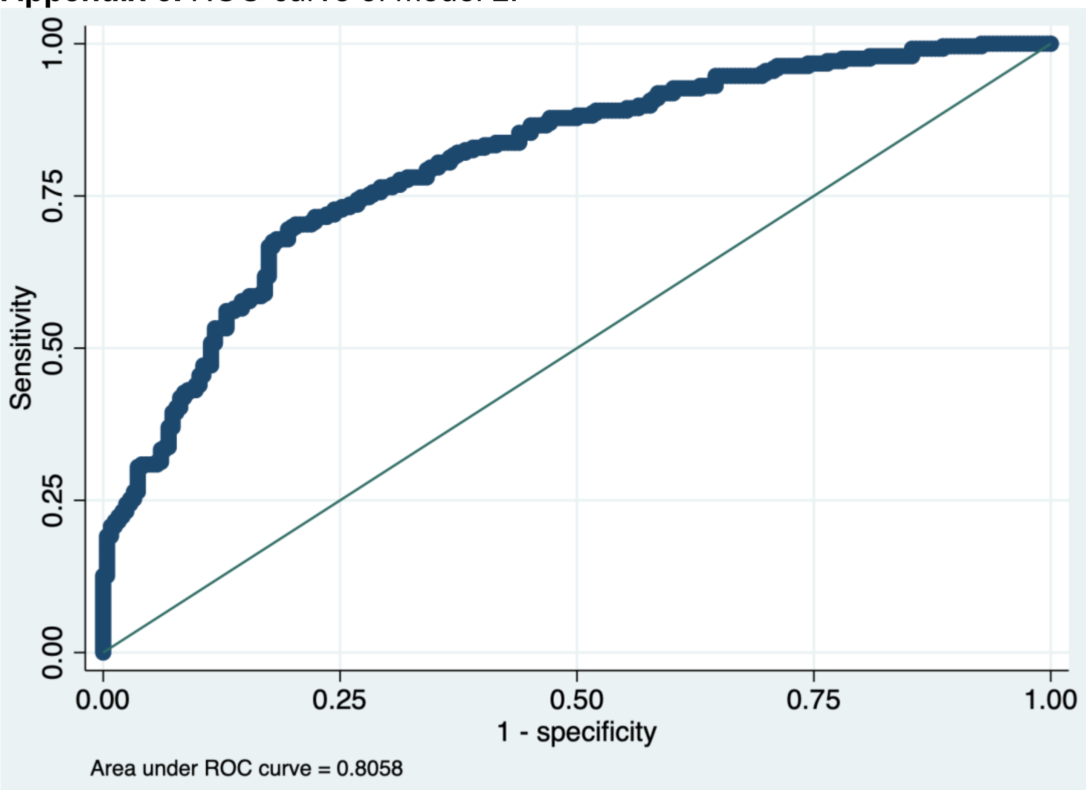
Appendix 5. Classification table of out-of-sample testing of model 1 with the cutoff of 0.28.

Classified	True B	True NB	Total
+	**102**	**55**	**157**
-	**10**	**55**	**65**
Total	**112**	**110**	**222**

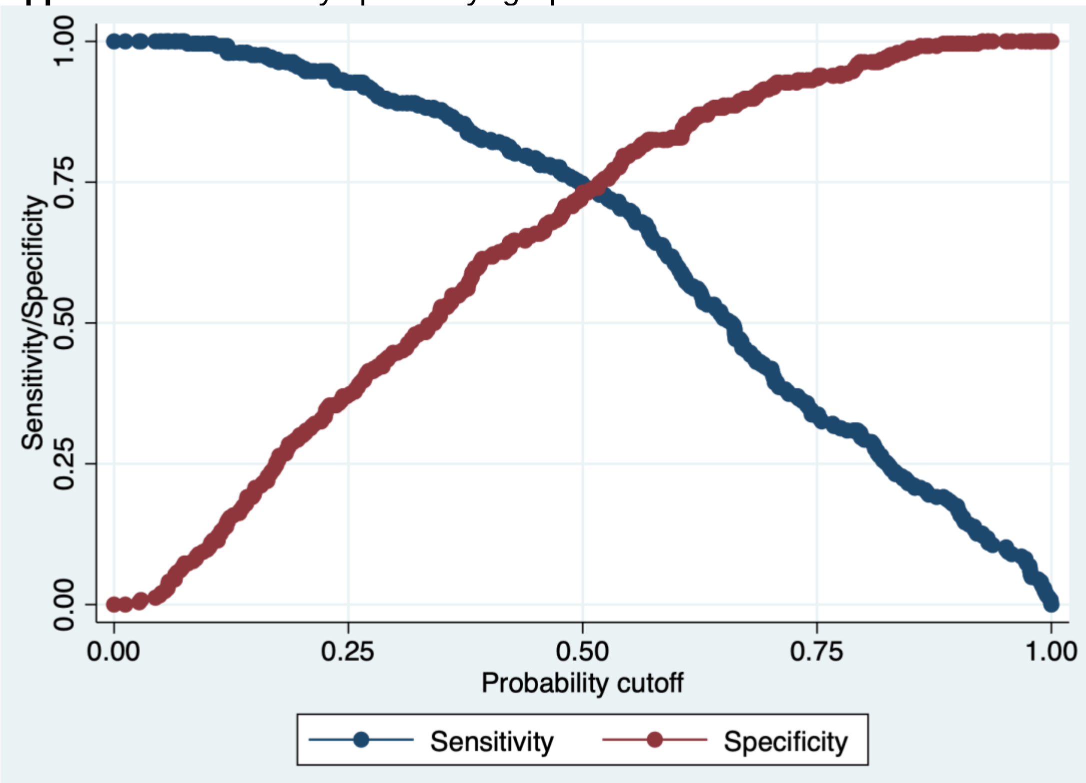
Classified + if predicted Pr(B) >= .28
True B defined as bankruptcy != 0

Sensitivity	Pr (+ | B)	**91.07 %**
Specificity	Pr (- |NB)	**50.00 %**
Positive predictive value	Pr (B | +)	**64.97 %**
Negative predictive value	Pr (NB| -)	**84.62 %**
False + rate for true NB	Pr (+ |NB)	**50.00 %**
False - rate for true B	Pr (- | B)	**8.93 %**
False + rate for classified +	Pr (NB| +)	**35.03 %**
False - rate for classified -	Pr (B | -)	**15.38 %**
Correctly classified		**70.72 %**

Appendix 6. ROC-curve of model 2.



Appendix 7. Sensitivity/specificity -graph of model 2.



Appendix 8. Classification table of model 2 with the default cutoff of 0.50.

	True		
Classified	B	NB	Total
+	**183**	**66**	**249**
-	**63**	**180**	**243**
Total	**246**	**246**	**492**
Classified + if predicted Pr(B) >= .50			
True B defined as bankruptcy != 0			
Sensitivity		Pr (+ | B)	**74.39 %**
Specificity		Pr (- |NB)	**73.17 %**
Positive predictive value		Pr (B | +)	**73.49 %**
Negative predictive value		Pr (NB| -)	**74.07 %**
False + rate for true NB		Pr (+ |NB)	**26.83 %**
False - rate for true B		Pr (- | B)	**25.61 %**
False + rate for classified +		Pr (NB| +)	**26.51 %**
False - rate for classified -		Pr (B | -)	**25.93 %**
Correctly classified			**73.78 %**

Appendix 9. Classification table of model 2 with the cutoff of 0.28.

Classified	True B	True NB	Total
+	**223**	**143**	**366**
-	**23**	**103**	**126**
Total	**246**	**246**	**492**

Classified + if predicted Pr(B) >= .28
True B defined as bankruptcy != 0

Sensitivity	Pr (+ | B)	**90.65 %**
Specificity	Pr (- |NB)	**41.87 %**
Positive predictive value	Pr (B | +)	**60.93 %**
Negative predictive value	Pr (NB| -)	**81.75 %**
False + rate for true NB	Pr (+ |NB)	**58.13 %**
False - rate for true B	Pr (- | B)	**9.35 %**
False + rate for classified +	Pr (NB| +)	**39.07 %**
False - rate for classified -	Pr (B | -)	**18.25 %**
Correctly classified		**66.26 %**