CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2018 **Michael Dibon**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Michael Dibon

Accounting-based credit scoring models – The Altman Z-score

Bachelor thesis

Prague 2018

Author: Michael Dibon
Supervisor: Mgr. Aleš Čornanič

Academic Year: 2017/2018

Abstract

This Bachelor thesis is focused on accounting-based credit scoring models, predominantly on Altman (1968) Z-score. We examine the relevance of the Z-score model on European publicly traded companies over the period 2012- 2017. Moreover, we analyze whether it is important to calibrate original models as well as we test the performance of models given different misclassification costs. Our results suggest that Altman original Z-score model is still, after 50 years of existence, relevant in the European after-crisis environment. Further, we found evidence that re-estimation of the model is unnecessary and could even cause harm to model performance. Finally, the performance of models seems to be stable given not equal misclassification costs, as the more accurate models from ROC analysis reported better results in an economic test.

Bibliographic note

DIBON, Michael. *Accounting-based credit scoring models – The Altman Z-score.* Praha, 2018. 45 p. Bachelor thesis. Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. Supervisor Mgr. Aleš Čornanič

Keywords

Z-score, accounting-based models, credit score, Altman, financial ratios, bankruptcy, ROC, Europe

Abstrakt

Tato bakalářská práce je zaměřena na kredit skóringové modely založené na účetnictví, převážně na Altmanovo Z-skóre z Altman (1968). Práce zkoumá relevanci Altmanova Z-score modelu na Evropských veřejně obchodovaných společnostech v průběhu let 2012 – 2017. Dále, analyzujeme, zda je důležité originální modely kalibrovat a také testujeme jejich predikativní výkon vzhledem k různým nákladům vyplývajících ze špatné klasifikace. Naše výsledky naznačují, že Altmanovo Z-score je stále, po 50 letech existence, relevantní v Evropském pokrizovém prostředí. Také jsme nalezli evidenci, která potvrzuje, že kalibrování originálních modelů není nutné a dokonce může způsobit oslabení výkonnosti. V neposlední řadě jsme zjistili, že výkonost modelů s ohledem na různé misklasifikační náklady vypadá stabilně, jelikož nejlepší modely v ROC analýze ukázaly lepší výsledky v ekonomickém testu.

Bibliografická evidence

DIBON, Michael. Credit scoringové modely využívající finanční ukazatele - Altmanovo Z-score. Praha, 2018. 45 str. Bakalářská práce. Univerzita Karlova v Praze, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce Mgr. Aleš Čornanič

Klíčová slova

Z-score, modely založené na účetnictví, kreditové skóre, Altman, finanční poměrové ukazatele, bankrot, ROC, Evropa

Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

2. The author hereby declares that all the sources and literature used have been properly cited.

3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague, May 5, 2018 **Michael Dibon**

Acknowledgments

I am grateful to my supervisor, Mgr. Aleš Čornanič for all his insightful comments, encouragement and fruitful discussions. Further, I would like to thank to my dear friend Dominik Filip for providing me with access into ORBIS database. Finally, I would like to thank to my family and girlfriend Sára for support and encouragement.

Bachelor's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author's name and surname: Michael Dibon

E-mail: michael.dibon@gmail.com

Phone: +420 721 572 444

Supervisor's name: Mgr. Aleš Čornanič

Supervisor's email: ales.cornanic@gmail.com

Proposed Topic:

Accounting-based credit scoring models – The Altman Z-score

Preliminary scope of work:

Research question and motivation

Altman's Z-score model is one of the most well-known financial ratio based models. It was the first multivariate bankruptcy prediction model. After the Z-score development, the financial distress research spread worldwide among researchers. For example, Ohlson (1980) proposed a logit model, Taffler (1984) developed a Z-Score model for the United Kingdom, and Zmijewski (1984) used a probit approach. Furthermore, Z-scores gained wide acceptance among auditors, management accountants and financial analysts.

Many papers have been written about the Z-score model, which proved its creditability over different markets. The recent work about usefulness of accounting-based models is ALTMAN, Edward I., et al. (2016). The study analyzed Z-score model for predominantly private firms from 31 European and 3 non-European countries using different modifications of the original model between the years 2002 - 2010.

The majority of the academic researches were conducted on private companies. Therefore, the aim of this thesis is to focus on European public companies and original Z-score model performance in recent years. Besides validity testing, I would like to also compare Altman's Z-score model, which was designed originally for publicly held manufacturing companies, with other accounting-based models. In addition, I would like to further examine the Z-score formula.

Hypotheses:

H1: Altman's Z-score model is still relevant in the current environment.

H2: The original Z-score model is being outperformed by other accounting-based models.

H3: The relationship between Z-score and rating from credit rating agencies is positively significant.

H4: Z-score shows stable results across different industries.

The goal of the first hypothesis is to test the original Z-score model and its performance in recent years. Since the thesis is focused on the publicly listed companies, the original Z-score, which includes a variable market value of equity, could be tested.

In the comparing section, I would like to compare Altman's original Z-score model, which was designed originally for publicly held manufacturing companies, with other, more recently developed accounting-based models, such as Altman's Z''-score model, which is designed to be applicable for non-manufacturing companies, Altman's Z'-score model - developed as implementation for private companies or the Taffler Z-score, which is UK-based.

The study by ALTMAN, Edward I.; RIJKEN, Herbert A. (2004) examines the stability of credit rating based on modified Altman's Z-score model. The paper does not research a degree of the relationship between the rating and Z-score. Therefore, in the third hypothesis, I would like to examine the relationship between Z-score and long-term credit rating issued by renowned credit rating agencies (CRAs). The possible positive relationship could be useful in financial analysis, as some analysts use rating from CRAs as a proxy for obtaining the cost of debt.

In the last hypothesis, I am going to compare Z-scores across different industries, in order to find potential credit characteristics. Do some industries have, for example, a higher Z-score due to its growth or capital intensity?

Contribution

This thesis should reveal a degree of relevance of the Z-score model in the current economic environment compared to different accounting-based models. Testing only publicly listed companies will also help broaden and update aggregated knowledge about Z-score model.

Moreover, the relationship between Z-score and credit rating from renowned credit rating agencies as well as possible advantages which may arise from industry specifics, will contribute to the existing literature and could help better understand the accounting-based models, predominantly Altman's Z-score model, in terms of practical use.

Methodology

The thesis is going to be based on financial ratios, Multiple Discriminant Analysis and regression analysis. Discriminant functions estimated by Altman and Taffler will be used. For comparing credit rating with Z-score I propose quantile analysis. In order to find out possible effects in different industries, I would like to use mean value for comparison.

Data

The examined data will consist of European publicly listed companies. The primary source of financial data and credit ratings will be the Thomson Reuters and EMIS databases. Financial ratios will be calculated from financial statements Information about bankruptcy will be obtained from either the mentioned databases or from the BancruptcyData platform.

Outline

1. Introduction
2. Review of existing literature
3. Hypotheses development
4. Data and Methodology
5. The Model
6. Results Discussion
7. Conclusion

List of academic literature:

Bibliography

ALTMAN, Edward I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 1968, 23.4: 589-609.

ALTMAN, Edward I., et al. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z−Score Model. *Journal of International Financial Management & Accounting*, 2016.

ALTMAN, Edward I.; RIJKEN, Herbert A. How rating agencies achieve rating stability. *Journal of Banking & Finance*, 2004, 28.11: 2679-2714.

ALTMAN, Edward I., et al. Predicting financial distress of companies: revisiting the Z-score and ZETA models. *Stern School of Business, New York University*, 2000, 9-12.

ZMIJEWSKI, Mark E. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 1984, 59-82.

BEAVER, William H. Financial ratios as predictors of failure. *Journal of accounting research*, 1966, 71-111.

Author **Supervisor**

Contents

List of Tables

List of Figures

1. Introduction

Being able to identify corporate insolvency in advance and predict financial distress is crucial for creditors, debtors as well as for shareholders and stakeholders. It is vital for both corporation and individuals to mitigate the risk of possible financial losses. Moreover, the bankruptcy destroys large economic values as loss created due to default is estimated by Agarwal and Taffler (2008) to be 45% of total investment. Therefore, the importance of having reliable predictive tool for lending and investment decisions appears to be crucial.

It seems instinctive to look at financial statements to see a better picture of company's health. In fact, an easy-to-apply method for evaluating the health of a company is to use accounting-based models. This method utilizes information from financial statements by implementing multiple financial ratios into a model to assign company a score which quantifies the degree of corporate risk.

Accounting-based models synthetize financial ratios into an intuitive single index, which should provide a higher value than only sum of its parts. Furthermore, the advantage of accounting-based models is in their practicality. These models provide users with an objective tool for risk measurement. Objectivity is a very powerful characteristic because users do not have to subjectively decide whether the company is solvent or not. It seems to be unfortunate to subjectively judge if company's leverage is unbearable or if the liquidity is sufficient. Agarwal and Taffler (2007) confirm that using own judgement to quantify the corporate risk is a difficult task.

There are also other methods how to measure a corporate risk, for instance in recent literature popular market-based models, which even tend to produce more accurate results. However, these models are impractical as the needed market-based data are difficult to obtain and sometimes even unobservable (e.g., volatility). Accounting-based models succeed in accessibility as accounting statements are not difficult to obtain. Another advantage of accounting-based models is in its applicability usually to both private and public companies. Moreover, according to Altman et al. (2017), accounting-based models are still a main or supporting tool for financial distress or bankruptcy prediction used worldwide in both practice and research.

In 1968 Edward I. Altman developed first multivariate bankruptcy prediction model – Altman (1968) Z-score, which is one of the most known bankruptcy predicting models. Financial textbooks are still writing about Z-score (e.g., Vernimmen et al. (2014), Padmalatha (2011) or Guerard and Schwartz (2007)), the model is being taught in financial universities across the globe and is also implemented in financial platforms such as Bloomberg.

Nevertheless, is the fame of Altman's model justified? Does the original 50-years old Z-score model deserve an interest from both academics and practitioners? The aim of this thesis is, therefore, to shed light into the relevance of Altman Z-score model. To do so, we choose to conduct the research in the current environment of European companies, as to our knowledge, such a study has not been written.

Moreover, as it is presumed by more than 50 years of research in the field of bankruptcy prediction, many famous accounting-based models credit scoring models were developed many decades ago. Mensah (1984) suggested that accounting-based models should be re-estimated from time to time so that models can incorporate changes in the economic environment. Thus, another objective of the thesis is to find whether updated formulas are more effective than its original version.

Furthermore, as the current literature is primarily focusing on the overall model accuracy and neglecting the different costs, which are caused by model misclassifications. One may ask, whether the cost, which arises from acceptance or investment into a bad debt is equal to the denial of loan or investment into a good debt. Probably not, as Altman et al. (1977) described, the latter costs appear to be predominantly costs of missed opportunity, whereas the loss created due to default is immense. Hence, the thesis will also be interested in the performance of models given different misclassification costs.

Overall the thesis should contribute to the insolvency and credit scoring literature as well as to financial practitioners by examining and evaluating Altman Z-score and other significant accounting-based models in Europe. The thesis should provide up to date results as the analysis is going to be conducted over the span of years 2012 – 2017.

This thesis will be structured in the following manner. In chapter 2, we will describe the development in credit analysis based on accounting models. Then we build and reason main hypotheses of the thesis. Chapter 4 will address selected models, and chapter 5 will discuss and describe data. After that, in chapter 6, the methodology of the thesis will be explained. Finally, Chapter 7 and 8 will provide results of the thesis as well as concluding remarks.

2. Literature review

The research in the field of accounting-based credit scoring models originated in the 1930s when academics started to observe differences in financial ratios between bankrupted and non-bankrupted companies. The literature then followed by using simple ratios in bankruptcy prediction. Further, in the late 60s, first multivariate models were developed such as Altman (1968) Z-score model. Onwards, many models and techniques for accounting-based models were proposed. In recent years, literature started to be interested in market-based models, which are theoretically sound, but data demanding. This chapter, is going to summarize the existing literature and capture the overall development in the field of accounting-based credit scoring models.

2.1. Early literature – univariate approach

Due to the economic downturn in the 1930s, which created a financially distressed environment, the academic research on solvency was initiated. Researchers started to search for a link between financial statements and bankruptcy. This early literature employed simple ratio analysis in order to find significant connections and possible trends in the field of corporate bankruptcy.

The first major mention about credit analysis was dated in 1930 when the Bureau of Business Research (1930) analyzed 24 financial ratios on the sample of 29 failing industrial companies in order to determine the major common characteristics, which can be observed in financial distress. The study recognized a few significant ratios, where the most important indicator of company's weakness was found to be Working Capital to Total Assets (WC/TA).

Then Fitzpatrick (1932) conducted research using financial ratios on both failed and non-failed companies. The study was created on a sample of 16 bankrupted and 16 non-bankrupted. Fitzpatrick found a significant difference in ratios and their trends between groups of failed and non-failed companies. Furthermore, the author stressed the importance of leverage ratios Net Worth/Debt and Net Profit/Net Worth.[1] On the other hand, author pointed out, that importance of Quick and Current ratio[2] was in case of

[1] Net worth is a term for book value or shareholders' equity.

[2] Quick ratio = (current assets – inventories) / current liabilities; Current Ratio = current assets / current liabilities.

companies with long-term liabilities low and even though the popularity of these ratios, should be looked at with a degree of caution.

The study of Smith and Winakor (1935) was following on the paper of Bureau of Business Research using a far larger sample of 183 failed companies. The publication confirmed the importance of the WC/TA ratio, which was according the study surprisingly better than Current ratio and Cash to Total Assets.

Another significant study was conducted by Merwin (1942), who used much larger sample compared to previous studies. This paper, which focused on manufacturing corporations with assets up to $250,000, showed the importance of three ratios, from which the most important ratio was again WC/TA. Moreover, the study observed the beginning of unfavorable trend already 4-5 years before bankruptcy. Besides, the paper justified the importance of dependable credit analysis – when in the random sample of 939 firms between years 1926 – 1936, Merwin recorded that almost 60% of corporations had failed.[3]

Walter (1957) presented an extension to the financial ratio analysis in solvency determination. He argued that the importance of working capital – availability of current assets and dismissal of current liabilities, is undeniable, however incomplete. Author compared company to “reservoir” where working capital is being filled by cash inflow and drained by cash outflow, stressing the importance of prevailing cash inflow that along with cash reserves should cover cash outflows with sufficient margin. This was one of the first explanations of distress in literature. Hence author suggested extending the classical ratio analysis with variables such as net cash flow and also with sales trend.

At this time researchers were focused on investigating significances of financial measures, they did not try to use financial ratios as a predictive tool for credit scoring models. Nevertheless, studies conducted between the years of 1930 – 1960 created an initial framework, which was needed for models with predictive abilities.

2.1.1. Foundation of a predictive model

In 1966, the fundamental work in credit scoring models was written by professor William H. Beaver in his paper “Financial Ratios as Predictors of Failure”. Beaver (1966) which has been inspired by early literature, was the first study which utilized the potential

---

[3] 558 of 939 companies stopped making reports for income tax purposes.

of financial ratios and developed a prediction model for credit scoring/bankruptcy prediction.

Beaver similarly to previous studies such as Fitzpatrick (1932), Smith and Winakor (1935) or Merwin (1942) compared means of financial ratios. In his case, 30 ratios, which were in the literature already recognized as important, were picked for further analysis.

Beaver used a sample of 79 failed, and 79 non-failed companies across various industries between years 1954 - 1964. Failure is according to Beaver defined as *"the inability of a firm to pay its financial obligations as they mature. Operationally, a firm is said to have failed when any of following events have occurred: bankruptcy, bond default, an overdrawn bank account, or nonpayment of preferred stock dividend."* Beaver (1966, p. 71). Beaver also selected non-bankrupted companies, which were then paired with bankrupted ones based on firm size and similar SIC (Standard Industrial Code). The author believed that firm size is an important link between ratios and bankruptcy because a small firm has a higher chance of bankruptcy than a large company, therefore should be indirectly included as well as industry specifics, which had been already shown by many earlier studies.

Despite the paired sample, the classical paired analysis was disregard in favor of the whole sample analysis. Beaver argued that comparing ratios with only the matched pairs has a serious drawback. If company A is compared with company B and A is more solvent than B, a question about solvency of company B emerges. Even though the paired analysis was not used the study still suggested paired sample design as a convenient method, which mitigates the biasing influence of industry and asset size factors.

By analyzing the means for failed and non-failed group for each year up to 5 years before failure, Beaver was able to find a significant difference between failed and non-failed firms. He found a significant difference (as long as five years ahead) for 6 ratios - cash-flow to total debt, net income to total assets, total debt to total assets, working capital to total assets, current ratio and no credit interval. Beaver's best performing ratio was cash-flow to total debt, which one-year prior bankruptcy provided 87% accuracy. However, the prediction power was decreasing with increasing years prior bankruptcy. That is also the reason why the study was suggesting employing multiple-ratio model, which may have better predictive ability than a single ratio.

Those were the beginnings in the field of accounting-based models. The literature at that time was focused solely on univariate methods. The most influencing work of this era was undoubtedly the Beaver's research, which originally employed bankruptcy prediction – it was a founding paper that helped shape the academic research onwards.

2.2. Multivariate models based on discriminant analysis

After the breakthrough work of William H. Beaver in the field of bankruptcy prediction, many studies were encouraged to proceed with research. Inspired by Beaver, Edward I. Altman (1968) published a paper, in which he created one of the first multivariate models.

2.2.1. Altman (1968) Z-score

Altman in his paper used Multivariate Discriminant Analysis (MDA), which was developed by Fisher (1938) and was used at that time predominantly in biology. Discriminant analysis is statistical technique, which is based on reverse analysis of variance (ANOVA). However, ANOVA uses a categorical independent variable, whereas discriminant analysis uses category as a dependent variable. MDA is used to classify and make predictions in cases where the dependent variable appears in qualitative form – in Altman's case, it is bankrupted and non-bankrupted companies.

Estimation sample consisted of 66 carefully selected companies, where bankrupted group counted 33 American manufacturing companies with data gathered between 1946 - 1965. Altman defined bankruptcy as a situation when a company filed a bankruptcy petition under Chapter X. Then a matching sample of 33 non-bankrupted companies based on firm size and industry was established. Altman did not pick large companies, because of a very low probability of chapter X at that time.

Then MDA was used to provide function, which was in the form of a linear combination of financial ratios, which was able to determinate differences between failed and non-failed companies. Altman was observing 22 ratios, which were recognized as significant indicators from previous literature in the field of financial distress. Altman then tried to select the best set of ratios using MDA repeatedly looking for the best discriminator. After that, the final function was established. Altman found a "Z-score" model, which utilized 5 balance sheet indicators, 2 income statement indicators, and 1 market data input in the form of 5 ratios (factors) - Working capital/Total assets, Retained Earnings/Total assets, Earnings before interest and taxes/Total assets, Market value equity/Book value of total debt and Sales/Total assets.

After the model was established Altman had to determine the interval, where the company is going to be considered as safe from default and the zone, where the company is predicted to fail. In order to correctly classify the cut-off points, Altman used a confusion matrix, where he recorded hits and misses of his model. Altman was focused on finding an ideal cut-off point to establish the most accurate model for predicting bankruptcy. Altman simply observed Z-scores of failed and non-failed companies and classified them based on a value which correctly classified the largest number of firms

By using point 2.99, Altman reached 95% accuracy one-year prior bankruptcy on the initial estimation sample. Altman then, similarly to Merwin (1942) or Beaver (1966) tested accuracy as far as 5 years before bankruptcy. Z-score model accuracy for correct classification from 1 to 5 years prior the bankruptcy declaration was 95%, 72%, 48%, 29%, 36% respectively. Since the accuracy dropped significantly with time, Altman concluded that Z-score has low predictive power for prediction more than 2 years into the future.

Altman then tested his model on a second sample, which consisted of 25 bankrupted and 66 non-bankrupted companies. This validation is based on Frank et al. (1965), who recommended using split or "hold-out" sample - two samples where first is being used for determination of multivariate discriminant model and second one serves for validation. The model's one-year predictive ability when tested on a validation sample was 96% for classifying bankrupted firms and 79% for non-bankrupted when model correctly classified 52 out of 66 non-failed companies. In the validation sample, the model suffered from higher misclassification of non-failed companies – type II error. Altman noted that the majority of false positively determined companies were positioned within the interval of 1.81 – 2.99 which Altman called "zone of ignorance" or commonly known a grey zone. This zone is considered as a zone where companies may suffer from financial distress but are still alive.

Atman in his paper showed the significance of multivariate model and sparked the further interest in the field of credit scoring models. The MDA technique has become one of the major statistical techniques in bankruptcy prediction. Moreover, Altman's model has become widely known and is still being mentioned in current literature as well as in financial textbooks.[4]

---

[4] For example, see Vernimmen et al. (2014), Padmalatha (2011) or Guerard and Schwartz (2007).

2.2.2. Inspired by Altman's Z-score

After Altman's success in multivariate analysis, Deakin (1972) analyzed 14 ratios, which were noted in Beaver (1966). Deakin saw better potential in the predictive ability of Altman's method, than reasonably precise Beaver's univariate approach. Following MDA methodology of Altman, Deakin plugged into discriminant function all 14 ratios. All ratios used previously by Beaver were very popular in literature and in practical application. Deakin stated that each of all 14 ratios was significant for all 5 years prior bankruptcy at the 5% significance level and hence added value to the model. Above that, Deakin did not use the cut-off point method as reported in Altman (1968). Author instead, in order to determine the cut-off point, assigned probability for each company based on Z-test and chi-square distribution. Overall, the study confirmed that multivariate model has better discriminatory power than using single ratios.

Edminster (1972) published a paper on small business failure prediction following the Altman (1968), although his Z-score model was built using binary coding system for MDA. Edminster also stressed the difficulty to predict bankruptcy for small companies with only one year of financial statements – as Altman did. Thus, the three-year statement discriminant function was used. In this model, implemented ratios were being inputted into Z-score model in transformed binary form. Thus, every ratio had its own cut-off point which transformed given value into 0 or 1. Overall Z-score was then calculated as a linear combination of binary inputs. Edminster showed the predictive ability of MDA even for small businesses. Although the study, to some degree, had to adjust the data because ratios of small companies suffered from a wide distribution of values. Edminster also concluded that multivariate analysis yields better predictability power as no single ratio tested was able to be as precise as a group of ratios.

After more than 10 years Altman et al. (1977) updated original Z-score function and prepared new model called ZETA (the model was built for commercial purposes in private sector, hence the exact formula was not disclosed). In the study authors discussed main reasons that justified why a new model should be built or updated. They concluded that companies were getting bigger than they used to be, therefore older models could have been estimated on inaccurate samples. Authors also included broader selection of companies including retail industry – which is according the study vulnerable to failure and therefore important for a research. Another reason suggested for a model update was development in accounting standards over the years.

Altman et al. (1977) built a 7-factor model (Zeta), which was overall more stable and accurate for bankruptcy prediction than the original Z-score model. Zeta model was also able to determine groups with a very low overlap zone. However, the original Altman's Z-score showed trustworthy accuracy, which was surprisingly very similar to updated Z-score.

Zeta model took an interesting approach to evaluating cut off points. The principal question was whether the wrong determination of failed and non-failed company bears the same costs. It was a first model, which did not treat misclassification costs as equal. The study instead assessed different specification to costs of type I error (i.e., C_I – cost from acceptation of loan to a bankrupted company) and costs of type II error (i.e., C_{II} – cost from rejection of loan to a non-bankrupted company). Authors, after both financial statement analysis and questionnaire results from banks, decided that representative result for C_I costs would be 70% of the loan. C_{II} costs were estimated to be 2% since the bank can in the worst-case scenario invest into theoretically riskless assets such as government securities of the same maturity as the loan to mitigate the costs of missed opportunity. Hence the C_I costs were estimated to be 35 times higher than costs C_{II} and the cut-off point was then adequately calculated.

The paper tackled the problem of different costs previously only theoretically addressed. Altman et al. concluded that it is important to clearly characterize the error cost for each decision-making process, in order to optimize the usefulness of the model.

2.2.3. Drawbacks of discriminant analysis

The Altman et al. (1977) also acknowledged weaknesses of the MDA analysis as some studies in the 70s mentioned.[5] In theory, MDA, which is classifying companies into two groups, should meet two assumptions.

Firstly, factors have to follow multivariate normal distribution – which is in terms of financial ratios not very common. If this assumption is not met, then the tests of significance and results are prone to be biased. Unfortunately, as extracting true joint distribution is an unfeasible mission, a majority of researchers accepted the fact that they would get only an approximation of true values.

[5] For example, see Eisenbeis (1978).

Secondly, both groups must have identical variance-covariance matrices. Then if each group reports different means the MDA (as noted in Altman – linear MDA) predicts optimally bankrupted vs. non-bankrupted firms. Contrary, if the groups do not have identical variance-covariance matrices, then the better technique should be quadratic MDA (QMDA). However, Altman et al. (1977) showed that even though the QMDA should be theoretically more precise, the classical linear MDA yielded better results.

Drawbacks of Hold-out samples are also addressed. For instance, Scott (1978) argues that hold-out sample should not be used for a small sample because it leads to poor prediction. Wood (2012) adds an intuitive commentary that reducing the small estimation sample in order to get unique validation sample seems illogical.

Another shortcoming could be the grey zone, which limits the practical application of Z-scores. Z-score is also criticized for its statistically driven foundation, where the financial theory is being sidelined. However, the input variables Altman has been using are tested and evaluated in the financial literature as significant, so the argument that models lack financial theory seems to be in many cases unjustified.

Overall, the MDA analysis is one of the most used techniques in the literature, especially in the 70's and 80's and achieved convincing predictive power, which resulted in practical use of Z-scores.[6]

2.3. Probabilistic approaches

After the Z-score era, where the majority of literature was focused on multivariate discriminant analysis, new approaches were popularized. One of the most popular approaches was based on conditional probability – predominantly logistic regression (logit). As Bellovary et al. (2007) reveal, logistic regression together with discriminant analysis were the most used methods in the 80s. On the same note, in the 90s the most used method was again logit model alongside with Neutral network method (NN).[7]

[6] For example, see Eidleman (1995).

[7] This paper will not discuss the artificially intelligent systems, even though they were popular in 90s. The main reason is the inconvenience in practical use as the NN models do not provide explicit formulas. Additionally, the NN method seems to be outperformed by market-based models Jackson and Wood (2013).

2.3.1. Ohlson – logit

Ohlson (1980) was the first academic research who utilize the logistic regression approach in the field of credit scoring models. Ohlson argued that MDA has theoretical drawbacks as mentioned in previous section (section 2.2.3). On the other hand, Ohlson agreed that the violation of theoretical background could be irrelevant. Nevertheless, the Ohlson added more practical reasons why to choose a logistic regression. Ohlson argued that MDA model is relatively difficult to interpret, because of its nature which is to be an ordinal discriminatory device, whereas a logit model can be interpreted as a probability. Another unfortunate characteristic linked to MDA stated by Ohlson is to have a matched sample of failed and non-failed companies. According to Ohlson's paper, it is not obvious what is gained or lost by paired matching. Paper suggests to rather include variables which are used in a matching process such as asset size or industry, into a model than using a matched sample.

Indeed, Ohlson did not use a matched sample. Instead, he chose to collect as large sample as possible – gathering 2163 companies. As a consequence, Ohlson used a sample, in which 5% of the sample represented failed and 95% non-failed companies.

The Ohlson's model – "O-score" was built from 9 ratios and had very high prediction power, using 0.5 as a potential cut-off point. However, the accuracy could have been partially biased due to author's decision to not use hold-out sample in order to have a larger sample for better model estimation. As the 0.5 cut-off point was just an intuitive decision, Ohlson conducted another interesting investigation. Author's approach was to use a very large number of cut-off points in order to find a cut-off point with a minimal sum of errors (error I + error II). The cut-off point was settled to be 0.038 (or 3.8%) and misclassified 17.4 % of non-bankrupted and 12.4% of bankrupted companies.

2.4. Market based models vs. accounting based

Black and Scholes (1973), Merton (1973) and Merton (1974) developed an option pricing theory, which led to creation of new market-based framework in the field of credit scoring. These market-based models are using contingent claims approach to assign credit score to given company.

In the beginning of 21st century, after more than 30-year dominance of accounting-based models, great number of studies started to focus on contingent claims approach as a predicative tool.[8] The expansion of interest which market-based models experienced was due to two main reasons. Firstly, credit agency Moody's published details about its credit rating methodologies, which were based on contingent claims approach (Moody's KMV model). Secondly, the motivation for academic research was supported with characteristics that market-based models theoretically possess. The concept was considered to be sound as efficiency of stock market should be greater than of accounting statements as market variables should theoretically reflect future cash flows and are not biased by different accounting policies.

As market-based and semi-market-based models became in academic research mainstream, studies started comparing the traditional accounting-based versus market-based approaches. Hillegeist et al. (2004) conducted study on a very robust sample between the years 1980 - 2000 concluded that the predictive power of bankruptcy of used market-based model is superior over the accounting-based models (tested on Altman (1968) Z-score and Ohlson (1980) O-score). The confirmation of this superiority can be seen in Tudela and Young (2003) for UK. Reisz and Perlich (2007) also showed that market-based models performed better in longer horizon prior bankruptcy. Although, Altman's Z-score (1968) was determined as very accurate in short-term prediction (1-year prior bankruptcy) similarly to market-based KMV type model and Down and out barrier option (DOC) model.

Recent European study of Wood (2012) tested 16 models on validation sample of UK public companies between the years 2006 – 2009. All four market-based models have beaten all univariate ratio-based models and mildly outperformed Altman's (1968) Z-score, Taffler (1983) Z-score and Ohlson (1980) O-score. To author's surprise, original accounting-based models performed very well in comparison with updated models and same models estimated using Neutral Network. However, the market-based models confirmed its superiority over accounting-based models. Furthermore, O-score and Z-score models, which were tailored on US markets outperformed the UK based Taffler (1983) Z-score.

Nevertheless, UK focused study of Agarwal and Taffler (2008) found that Taffler's Z-score produced the best results, slightly outperforming both European call

[8] For example, see Falkenstein et al. (2000), Bharath and Shumway (2004), Hillegeist et al. (2004).

option (EC) models from Hillegeist et al. (2004) and näive version from Bharath and Shumway (2004). When different misclassification costs were applied, and economic test was used to test the 2 best models, then accounting-based model outperformed a näive version of the EC model. Agarwal and Taffler in their study concluded that *"...despite extensive criticism of traditional accounting-ratio based credit risk assessment approaches, and the theoretically appealing contingent claims framework, in practice such conventional approaches are robust and not dominated empirically by KMV- type option-based models."* Agarwal and Taffler (2008, p. 1550)

Overall, recent studies, found a solid degree of evidence that accounting-based models have been outperformed by market-based models. However, in some studies, accounting-based models performed well in short term distress prediction. Further, accounting-based models are easy to apply as financial statements availability is abundant due to many financial platforms. Also accounting-based models advantage is in wider application – applicable for both private and public companies.

3. Hypotheses and motivation

The following chapter provides the big picture motivation for the thesis as well as the development of individual testable hypotheses.

As the review of literature implies, recent academic research has been aiming predominantly towards market-based models. Nevertheless, accounting-based models are still important. Firstly, literature does still support an accounting approach, despite the evidence that market-based models seem to be more accurate and stable. Moreover, new studies solely focused on accounting models are still being conducted – for example, Altman et al. (2017) a recent review of Altman (1983) Z’’-score model.

Secondly, and more importantly, the undeniable advantage of models based on financial ratios is in real-life applications, as models are based on relatively easily accessible data – financial ratios, which are then synthesized into an intuitive single index. Hence, these models serve as a practical communication device that utilized a whole set of ratios into a combined information about an overall economic condition of a company rather than limited one at a time information provided by conventional financial ratio analysis. This is a powerful characteristic, which can be easily utilized by managers and credit or equity investors.

Country-wise, more research in Europe seems to be desired as a vast majority of studies in the field of credit-scoring have been conducted on US companies. The reason seems to be very practical - bankruptcy data are much more easily accessible for companies incorporated in the US, because of uniform laws on the subject of bankruptcies throughout the United States and easily applicable bankruptcy codes such as Chapter 7, Chapter 11 and others. In Europe, the most fruitful researches were conducted in the UK, where the widely known Taffler (1983) Z-score was created. However, in other European countries, the research has been conducted in much smaller scale. In some countries, the research of original Altman (1968) Z-score is limited due to lack of publicly traded companies – for example in the Czech Republic.

Driver for empirical model comparison is simple. The developed models should be compared on the same dataset with the same validation methods to ensure true comparison of predictive abilities. That being said, comparison of accounting-based models in Europe as a whole after the financial crisis appears to be an interesting topic for an examination as such a study has not been to our knowledge produced.

3.1. Hypotheses development

Altman Z-score (1968) is undoubtedly one of the most known credit scoring models in the finance. As the model is widely-known across the word and is being lectured by many financially oriented universities and directly implemented in financial platforms such as Bloomberg. Hence, the fundamental thesis question is, whether the 50-years old model is still relevant given its publicity. Moreover, is Altman Z-score even suitable for European countries? Agarwal and Taffler (2007) argue that applying the "very accessible" Altman (1968) US model in different environments besides the US is potentially dangerous. Therefore, the objective of the first hypothesis is to evaluate the relevance of the model in the European after-crisis environment.

H_1: *Altman (1968) Z-score model is relevant in the current economic environment for European publicly listed companies.*

Moreover, is the original formula less accurate than a model with updated coefficients? Does generally matter to have calibrated models? Or the calibration is not so important. One can argue that model update should include the change over time in accounting standards and the degree and way of financial statement manipulation. Moreover, the technical development has been enormous in last decades and changed the way how companies operate. The supply chain management, working capital and just in time delivery and many more trends, which occurred in company's management. Mensah (1984) suggested that accounting-based models should be re-estimated from time to time so that models can incorporate changes in the economic environment. However, some studies showed the opposite, for example, Wu et al. (2010) showed that his updated logit Z-score model performed very poorly compared to other models tested. Furthermore, do models, which were developed on US companies, have a reliable predicting power for European companies?

H_2: *Models with calibrated coefficients have better predictive abilities than original formulas.*

Another important characteristic in the bankruptcy prediction is the evaluation of cut-off points. It seems that literature is predominantly focusing on minimizing the overall misclassification. Models accuracy is usually tested using methods, which treats misclassification costs equally. One may ask, whether the cost, which arises from acceptance of in future defaulting loan is equal to the denial of a loan, which is going to be fully repaid. According to Altman et al. (1977), these costs are not the same. Hence, the third hypothesis is going to address the issue of misclassification costs.

H3: *The better performing models are going to, after controlling for misclassification costs, perform better.*

4. Model selection

To ensure the relevant evaluation of Z-score and proper comparison, this thesis is going to test and compare the main model – Altman (1968) Z-score with other accounting-based models – Altman (1983) Z''-score model, Ohlson (1980) O-score model and UK-based Taffler (1983) Z-score model. Updated versions of models are going to be used for validation as well. Thus, two formulas of models are going to be presented, first version is going to be the original formula and the second version is going to be the model with updated coefficients. Updated versions of models are going to be taken from recent literature since the own re-estimation of models does not seem reasonable. The accounting-based models are based on statistical techniques, which need to be estimated and validated on separate data sets. Therefore, the hold-out sample was in the literature usually used. Since the aim of the thesis is to compare and validate models in a current economic environment, the process of an update is not desirable, as the data set would significantly shrink (bankruptcy data are very difficult to obtain) due to estimation sample requirement. Although there are some techniques, which can be employed such as Lachenbruch jack-knife, rolling window method or random sample cross-validation the main issue is that these techniques do not deliver a true ex-ante determination (Wood, 2012).

The re-estimated versions are found in such way, that models are re-estimated on newer data than original formula and possibly on European companies.

As a control group, two univariate models are going to be used as a main benchmark for multivariate models. The first ratio is going to be one of the most important ratios among literature working capital divided by total assets (WC/TA) and the second is going to be the Beaver's favorite ratio cash flow divided by total debt ("Beaver") defined by Beaver (1966) as net income plus depreciation and amortization divided by total liabilities.

4.1. Z-score

The main model is the 5-factor Altman (1968) Z-score, which was estimated using a matched sample of 66 carefully selected US companies during the period 1946 – 1965. The form of the model is as follows:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5 \quad (1)$$

Where X_1 = *Working capital/Total assets*

X_2 = *Retained Earnings/Total assets*

X_3 = *Earnings before interest and taxes/Total assets*

X_4 = *Market value equity/Book value of total debt*

X_5 = *Sales/Total assets*

Z = *Overall Index*

Updated version is going to be taken from Wood (2012) – version is estimated on 3003 UK companies between the years 2000 - 2005 (Wood in the same way as Ohlson did not use matched sample, instead maximized the sample size – 1.3% companies were bankrupted). The model was selected because the estimation is conducted on UK companies, which is a desirable characteristic since the thesis is interested in European companies.

Furthermore, another model, a revised version of the Z-score model is going to be used. It is US-based model from Altman (1983) known as Z''-score. Z''-score model was developed for both private and public manufacturing and non-manufacturing companies. The motivation to create the model was in a wider application. Hence, the market value of equity was changed to book value of equity and the sales to total assets ratio (S/TA) was excluded, because of a potential industry effect (e.g., retail vs. manufacturing). So, the model should be better applicable to for more industries regardless the private or public status of companies. The updated version of Z''-score model is taken from Altman et al. (2017). The model is developed on predominantly European both private and public companies between the years 2002 – 2010 on massive sample of 2.6 mil companies (1.4% bankrupted).[9]

[9] The updated model is named "Model 1" in the Altman et al. (2017).

Table 1: Original and updated coefficients for Z-score and Z''-score

Altman models	*WC/TA*	*RE/TA*	*EBIT/TA*	*VE/TL*	*S/TA*	*Const.*
Original [a]	1.20	1.40	3.30	0.60	0.999	-
Wood (2012)	2.67***	-0.001	0.423***	0	0.38**	-
Altman (1983)[a]	6.56	3.26	6.72	1.05[b]	-	3.25
Altman et al. (2017)[a]	0.561	0.724	1.791	0.021[b]	-	0.042

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided test). WC/TA is working capital divided by total assets; RE/TA is retained earnings divided by total assets; EBIT/TA is earnings before interest and taxes divided by total assets; VE/TL is market value of equity divided by total liabilities; S/TA is sales divided by total assets.

[a] In the study, t-statistics are not reported.

[b] Altman (1983), Altman et al. (2017) models are using book value of equity instead of market value since Z''-score model can be applied either to private or public companies.

Interestingly, coefficients are very different for Z-score models. Wood's updated Z-score assigns very low coefficients to variables RE/TA and VE/TL. For Z''-score the difference in coefficients is not as dramatic if we take into account the difference in years and countries.

4.2. O-score

The next model is O-score from Ohlson (1980), the model has been chosen due to its popularity in the literature[10]. The O-score is 9-factor linear model (plus constant) and was created on the sample of 105 bankrupted and 2058 non-bankrupted US companies. Besides financials, the model is utilizing lags on net income and also log size of a company corrected for inflation as a predictive tool.

As updated version, the Hillegeist et al. (2004) is going to be utilized predominantly due to its very robust training sample. The model is assessed on a sample of more than 15,000 US companies between the years 1980 - 2000 (5% of the sample consisted of bankrupted companies). Coefficients of both, original and updated O-score are recorded in Table 2.

[10] According to Google Scholar, study Ohlson (1980) has been recently (since 2014) cited more than 1900 times.

Table 2: Original and updated coefficients for O-score

Ohlson (1980) model	*Const.*	*SIZE*	*TL/TA*	*WC/TA*	*CL/CA* ...
Original	-1.32	-0.407***	6.03***	-1.43**	0.0757
Hillegeist et al. (2004)	-5.91***	0.04***	0.08***	0.01***	-0.01

...	*NI/TA*	*FU/TL*	*INTWO*	*OENEG*	*CHIN*
Original	-2.37**	-1.83***	0.285	-1.72***	-0.521***
Hillegeist et al. (2004)	1.20**	0.18***	0.01***	1.59***	-1.10***

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided test). SIZE is the log(total assets/GNP price level index); TL/TA is total liabilities divided by total assets; WC/TA is working capital divided by total assets; CL/CA is current liabilities divided by current assets; NI/TA is net income divided by total assets; FU/TL is funds provided by operation (pre-tax income plus depreciation and amortization) divided by total liabilities; INTWO is one if net income was negative over last 2 years is negative, zero otherwise; OENEG is one if owners' equity is negative (TL>TA), zero otherwise; CHIN is changed in net income calculated as $(NI_t - NI_{t-1})/(|NI_t| - |NI_{t-1}|)$.

Interestingly, the coefficients of the updated O-score model show a higher statistical significance compared to its original version. The updated model and also shows a big difference in coefficients compared to the original version. The sign is different for 5 from 10 coefficients, potentially indicating significant changes in economic environment.

4.3. T-score

The last model is possibly the most known accounting-based model for non-US companies – Taffler (1983) Z-score.[11] It is a 4-factor linear model estimated though discriminant analysis. Model coefficients were not disclosed until in Agarwal, Taffler (2007) study which was evaluating the T-score model. The model was developed on a small sample of UK companies, which were, similarly to Z-score, carefully selected. Unfortunately, the updated version of Taffler was not available as the formula was not until recently disclosed. Moreover, Taffler argues that formula should not be re-estimated, but rather redeveloped if a model shows signs of poor performance, hence he encourages

[11] For purpose of this paper the Taffler Z-score is named "T-score", to avoid misinterpretation with Z-score from E. I. Altman.

to use the original formula. The exact formula of the model and description of variables is shown in Table 3 below.

Table 3: Original coefficients for T-score

Taffler (1983) model	*PBT/CL*	*CA/TL*	*CL/TA*	*NCI*	*Const.*
Original coef.	12.18	2.50	10.68	0.029	3.20

Notes: PBT/CL is profit before tax divided by current liabilities; CA/TL is current assets divided by total liabilities; CL/TA is current liabilities divided by total assets; NCI is no-credit interval calculated as (quick assets – current liabilities) divided by daily operating expenses proxied by (sales – PBT – depreciation and amortization)/365.

5. DATA

In this section process of data collection is described, then a definition of a bankrupt and non-bankrupt company is defined, and finally, description of dataset and financial variables is conducted.

5.1. Data collection

To get the desired dataset for a research two steps need to be realized. First and foremost, the list of bankrupted companies, as well as non-bankrupted companies has to be obtained. Second, the accounting and other data for given companies have to be collected.

As the main source of data for the thesis, we use database ORBIS of Bureau Van Dijk (BvD) from which all financial, market and macroeconomic data are obtained. As an additional source of information, fact sheets from European stock exchanges are used to enlarge the sample of failed companies.[12]

Several requirements are set for the data sampling. Firstly, data are obtained only for publicly listed companies since the Z-score formula cannot be applied on private firms. Secondly, data are collected for companies from countries of European Union (28 countries including the UK) as these countries should be more unified in economic conditions that countries outside EU, because of the unified regulation framework and accounting standards (IFRS). Consequently, to compare firms from different counties, we obtained all financial data in Euros. Thirdly, we omit companies from financial industry due to different structure of accounting statements. Lastly, the time span was set to be for the years 2012 – 2017 as the trade-off between sample size and the stability of economic conditions, accounting standards, and overall present-day relevance seems to be reasonable.

Furthermore, since a cross-section data could suffer from a selection bias for alive companies, the decision was made to not use a set containing only one observation for

---

[12] From fact sheets of the biggest stock exchanges in Europe, i.e., French, Polish, Spanish, Swedish and Italian, list of 167 failed companies was built, because the reason of delisting was described as due to a bankruptcy. Unfortunately, the reason of delisting is not stated for many exchanges such as German Xetra. The list was then imported to ORBIS and the filters were applied (non-financial industry, geography, available financials).

each alive company. For this reason, each non-bankrupted company has on average 4.4 observations in the final sample. For the bankrupted companies the year of last available financial statements was set to be the year of interest – as it is set to be in majority of the literature.[13] Thus, each bankrupted company is a unique observation.

Finally, to remove the impact of extreme values, all explanatory variables have been winsorized at the 1st and 99th percentiles for both failed and non-failed sub-samples.

5.1.1. Definition of Bankrupted/Non-Bankrupted companies

In the literature the definition of bankruptcy (failure) is defined variously from strict definition of only taken failure as bankruptcy declared by a court to less rigorous definitions such as Beaver's overdraw of bank account or loan renegotiations/defaults, rescue rights issues or forced disposals from Bauer and Agarwal (2014).

The rigorous definition although seems as inconvenient as the size of the bankrupted sub-set from sample is crucial. Therefore, for purpose of this thesis, the bankruptcy (failure) is defined as follows: (i) when a company defaults on the payment of its debt, (ii) when the insolvency proceeding is initiated or (iii) when a company goes bankrupt.

On the other side, non-bankrupted company is defined as a company, which has active status in the ORBIS database and is as of 31.12.2017 publicly listed in a stock exchange.

5.2. Data description

The collected data amounted in a list of 705 bankrupted companies, however there were many missing variables at least for one model.[14] The importance to have the same homogenous dataset for all models is crucial for a proper model comparison. Hence, the final dataset for bankrupted companies consists of 136 companies. For the Non-bankrupted companies, the dataset shrank due to incomplete variables from 6,574 companies to 4,444. Overall, as it is shown in Table 5, the final dataset is composed of 19,460 non-bankrupted and 136 bankrupted observations, with a 0.7% bankruptcy rate.

[13] For example, see Hillegeist et al. (2004), Agarwal and Taffler (2008), Bauer and Agarwal (2014).

[14] There was an effort to find the missing variables in other databases such as Thomson Reuters Eikon, however the issue of missing data preserved. Therefore, to ensure consistency, input data for variables were taken only solely from ORBIS database.

Table 4: Sample description

Year	*NB*	*B*	*% B*	*Total*
2012	3115	17	0.5%	3132
2013	3359	18	0.5%	3377
2014	3915	13	0.3%	3928
2015	3972	21	0.5%	3993
2016	4020	62	1.5%	4082
2017	1079	5	0.5%	1084
Total	19,460	136	0.7%	19,596

Notes: NB means non-bankrupted, B means bankrupted companies.

Concerning the country distribution, every country from European Union is represented in the sample. The largest shares in the sample belong to the United Kingdom, France, Germany and Poland with the share of 22%, 12%, 12% and 12% respectively (table of geographic distribution is enclosed in Appendix as Table 15). Industry-wise, the biggest sectors, according to BvD sorting, are (i) other services; (ii) machinery, equipment, furniture, recycling; (iii) chemicals, rubber, plastics, non-metallic products; and (iv) wholesale & retail trade with the share of 63% (table of industry distribution can be seen in Appendix as Table 16).

5.2.1. Description of variables

Tables 5 - 8 present descriptive summary statistics for each explanatory variable of Ohlson, Altman, Taffler and benchmark ratio of Beaver used in this thesis over the 2012 - 2017 period. Variables are divided into four categories – profitability, liquidity, leverage and size.[15] Descriptive statistics are an important tool, which helps to understand the importance of variables. Nevertheless, they have to be observed with a degree of caution, due to high deviations in the bankruptcy sub-sample. Hence, the min, max and median value can help bring a clarity.

The profitability ratios measure the ability of a company to generate sufficient profit ("inflow to reservoir") to remain financially healthy and also measure a capability of management in dealing with competitive conditions. It can be seen that bankrupt companies have as expected much lower profitability compared to assets than a non-

[15] The classification is following Wu et al. (2010).

bankrupted sub-set. On the other hand, an insignificant difference is shown in the asset turnover ratio measured as Sales divided by total assets. The small difference can be impacted by the share of generally higher asset turnover industry, such as retail and wholesale, in each sub-sample (8.4% for non-bankrupted vs. 14% for bankrupted).

Table 5: Profitability ratios

		Mean	*SD*	*Max*	*Median*	*Min*	*Who*	*Sig.*
EBIT/TA	B	-0.39	1.08	1.00	-0.09	-5.92	Z	***
	NB	0.005	0.19	0.33	0.04	-1.08		
S/TA	B	1.10	1.84	12.11	0.61	0.00	Z	-
	NB	0.90	0.74	4.04	0.76	0.00		
NI/TA	B	-0.99	3.41	1.08	-0.14	-22.84	O	***
	NB	-0.02	0.20	0.30	0.03	-1.21		
CHIN	B	-0.12	0.66	1.00	-0.08	-1.00	O	**
	NB	0.02	0.54	1.00	0.03	-1.00		

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided Welch's t-test for the means). NB is non-bankrupted, B is bankrupted. EBIT/TA is earnings before interest and taxes divided by total assets; S/TA is sales divided by total assets; NI/TA is net income divided by total assets; CHIN is change in net income calculated as $(NI_t - NI_{t-1}) / (|NI_t| - |NI_{t-1}|)$.

In Table 6, Liquidity ratios, which measure the ability of a company to fulfill its short-term obligations, are depicted. One of the most important ratio WC/TA is, as in the literature, significantly lower for the bankrupted companies. CL/CA supports the WC/TA and shows that failing companies have liquidity issues compared to the rest. Moreover, cash flow based FUTL and Beaver's CF/TL shows negative cash inflow to bankrupted companies and its inability to deal with company's liabilities compared to positive cash inflow in non-bankrupted sub-sample. Ohlson's variable INTWO – a binary ratio that assigns 1 if net income was negative over last two years was negative, zero otherwise, shows that 54% of failed companies had a negative income compared to only 21% of healthy companies.

Table 6: Liquidity ratios

		Mean	*SD*	*Max*	*Median*	*Min*	*Who*	*Sig.*
WC/TA	B	-4.03	20.42	0,77	-0.12	-146	Z, O	**
	NB	0.15	0.27	0,87	0.13	-0.78		
CL/CA	B	12.65	45.96	310	1.55	0.15	O	***
	NB	0.95	1.20	9.36	0.68	0.03		
FUTL	B	-0.22	0.71	1.58	-0.09	-4.37	O	***
	NB	0.05	0.94	2.49	0.13	-5.83		
INTWO	B	0.54	0.50	1,00	1.00	0.00	O	***
	NB	0.21	0.41	1,00	0.00	0.00		
PBT/CL	B	-0.64	1.62	1,95	-0.23	-10.44	T	***
	NB	-0.03	1.53	5,43	0.12	-8.79		
CA/TL	B	0.63	0.90	6,48	0.45	0.00	T	***
	NB	1.54	2.60	19,71	0.88	0.03		
NCI	B	-543	1302	547	-96.86	-6568	T	***
	NB	48.79	532	3146	9.19	-2375		
CF/TL	B	-0.22	0.71	1,42	-0.09	-4.37	B	***
	NB	0.03	0.89	2,42	0.11	-5.60		

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided Welch's t-test for the means). NB is non-bankrupted, B is bankrupted. CL/CA is current liabilities divided by current assets; FU/TL is funds provided by operation (pre-tax income plus depreciation and amortization) divided by total liabilities; INTWO is one if net income was negative over last 2 years is negative, zero otherwise; PBT/CL is profit before tax divided by current liabilities; CA/TL is current assets divided by total liabilities; NCI is no-credit interval calculated as (quick assets – current liabilities) divided by daily operating expenses proxied by (sales – PBT – depreciation and amortization)/365; CF/TL is cash flow divided by total debt calculated as net income plus depreciation and amortization divided by total liabilities.

Another category - leverage variables are ratios, which measure the overall obligations of the firm. From both book or market value of equity to total debt as well as from retained earnings to total assets can be seen that leverage is an important factor for a bankruptcy – bankrupting companies are clearly more indebted. Bankrupting companies are also more likely to report total liabilities in excess of the total assets (OENEG), or else the equity is negative for 42% of bankrupting companies compared to only 4% non-bankrupting. Ratio TL/TA supports the OENEG results.

Table 7: Leverage ratios

		Mean	*SD*	*Max*	*Median*	*Min*	*Who*	*Sig.*
VE/TL	B	1.88	6.22	41.62	0.25	0.00	Z	***
	NB	5.08	13.28	102	1.40	0.02		
BVE/TL	B	0.96	4.32	33.47	0.11	-0.99	Z"	***
	NB	2.73	7.41	59.90	0.93	-0.37		
RE/TA	B	-14.03	75.46	0.57	-0.62	-575	Z	**
	NB	-0.24	1.41	0.80	0.06	-9.91		
TL/TA	B	5.70	25.47	175	0.90	0.03	O	**
	NB	0.52	0.27	1.58	0.52	0.02		
OENEG	B	0.42	0.50	1.00	0.00	0.00	O	***
	NB	0.04	0.19	1.00	0.00	0.00		
CL/TA	B	4.49	20.48	147	0.56	0.02	T	**
	NB	0.31	0.21	1.17	0.27	0.01		

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided Welch's t-test for the means). NB is non-bankrupted, B is bankrupted. VE/TL is market value of equity divided by total liabilities, BVE/TL is book value of equity divided by total liabilities; RE/TA is retained earnings divided by total assets; TL/TA is total liabilities divided by total assets; OENEG is one if owners' equity is negative (TL>TA), zero otherwise; CL/TA is current liabilities divided by total assets.

At last, size of the company calculated as of inflation corrected assets support the literature observation that on average the probability of failure for smaller companies is higher than for larger corporations.

Table 8: Size ratio

		Mean	*SD*	*Max*	*Median*	*Min*	*Who*	*Sig.*
SIZE	B	5.37	2.28	10.09	5.44	-0.81	O	***
	NB	7.22	2.52	13.23	7.08	1.37		

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided Welch's t-test for the means). NB is non-bankrupted, B is bankrupted. SIZE is the log(total assets/GNP price level index).

Finally, the overall difference between the bankrupted and non-bankrupted companies is shown in Table 9. O-score models and updated Z''-score model are built to be increasing in the probability of default, on the other hand, Z-sores and T-score are decreasing. Thus, mean values are for bankrupted companies higher for O-score models and lower for Z-score and T-score models. All models have significantly different means for sub-samples at least at 5% significant level, showing a sign of discriminatory abilities.

Table 9: Description of models

		Mean	*SD*	*Max*	*Median*	*min*	*Sig.*
Z-score	B	-23.55	128	24.17	-0.46	-990	**
	NB	3.80	7.99	65.72	2.27	-18.26	
Z-score (upd.)	B	-10.49	54.54	4.17	-0.21	-391	**
	NB	0.74	0.79	4.01	0.72	-2.53	
T-score	B	-66.48	227	53.16	-15.36	-1639	***
	NB	4.82	24.62	210	3.47	-124	
Z"-score	B	-70.57	369	37.84	-0.67	-2852	**
	NB	6.35	9.92	74.69	5.79	-41.81	
Z"-score (upd.)	B	-13.06	65.44	1.47	-0.78	-504	**
	NB	0.02	1.29	2.75	0.27	-9.51	
O-score	B	0.79	0.29	1.00	0.94	0.00	***
	NB	0.32	0.32	1.00	0.18	0.00	
O-score (upd.)	B	0.01	0.017	0.124	0.005	0.00	***
	NB	0.005	0.004	0.065	0.004	0.00	

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively (two-sided Welch's t-test for the means). NB is non-bankrupted, B is bankrupted. O-scores are transformed into probabilities using logistic cumulative distribution function.

6. Methodology

In this chapter, the methodology for model evaluation is going to be described. To compare the models and evaluate them two approaches are utilized: (i) the receiver operating characteristics (ROC) curve is used to evaluate predictive power, and additionally, (ii) test of economic value is used to address misclassification costs.

6.1. The ROC curve

The predicative power of credit-scoring models is relying on minimizing of type I and type II errors. In the literature researchers usually set an arbitrary cut-off point, which was used to test a validity of a model. For instance, Altman (1968) used a confusion matrix to find a best possible cut-off point and to count the accuracy of the model.

Table 10: Confusion matrix

		Actual situation		
		Bankrupted	Non-Bankrupted	Sum
Prediction	Bankrupted	True positive prediction (Sensitivity)	False positive prediction (1-Specificity) Type II error	TP+FP
Prediction	Non-bankrupted	False negative prediction (1-Sensitivity) Type I error	True negative prediction (Specificity)	FN+FP
	Sum	TP+FN	FP+TN	TP+FP+TN+FN

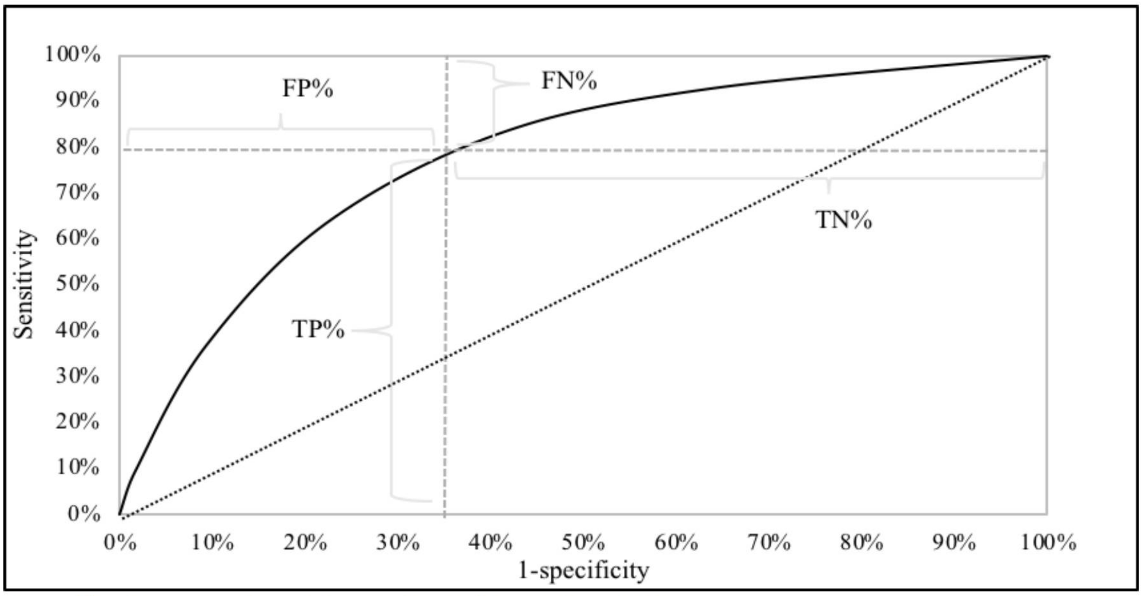
Notes: True positive (TP) is a correct classification of a default; True negative (TN) is a correct classification of non-default; False positive (FP) is a misclassification of default when a company actually survives; False negative is a is a misclassification of non-default when a company actually fails.

However, the problem with the confusion matrix is that the model is evaluated given a one cut-off point. Sobehart et al. (2000) argue that these techniques are often inappropriate in the domain of credit models and suggests more complex methods, such as receiver characteristic operator (ROC).[16]

ROC method has been widely used in the medicine and nowadays is a common method to represent the discriminatory power of a credit scoring model. It deals with the problem of the one single cut-off point by evaluating the performance over the whole range of cut-off points.

ROC depict true positive (TP), false positive (FP), false negative (FN), true negative (TN) predictions for all cut-off points (thresholds) in the simple plot. Usually, the ROC plots sensitivity, which is true positive rate (TP/(TP+FN)), called sensitivity on the y-axis. On the other hand, the x-axis is called 1-specificity and it is false positive rate calculated as (FP/(TN+FP)). To plot ROC, companies are first ordered by model score, from riskiest to safest. Then for each of possible cut-off point, the true positive rate and false positive rate is calculated and then constructed into a curve. The ROC curve is depicted in Figure 1, where 45° dotted line represents a random model and the black line represents a model, which has some predictive power. The performance of a model is higher when the nearest point from ROC is closer to upper left corner as the area under ROC curve (AUC) is increasing and overall errors are decreasing. Thus, the size of the area under a curve is crucial indicator of model's quality.

Figure 1: ROC curve – Theoretical example



Notes: The Figure is adopted from Stein, R. M. (2005, p. 1216).

[16] ROC has variety of names - Cumulative Accuracy Profiles, lift-curves, dubbed-curves, power curves.

As Bauer and Agarwal comment: "*the area under the ROC curve (AUC) is the decisive indicator of a model's predictive ability*" Bauer and Agarwal (2014, p. 436). Advantage of AUC is in a convenient interpretation. "*It is equivalent to the probability that a randomly chosen defaulting loan will be ranked worse than a randomly chosen non-defaulting loan by a specific model*" Stein (2002, p. 83).

The AUC is going to be estimated using a non-parametric method - the trapezoidal rule as:

$$\widehat{AUC} = \sum_{i=1}^{n} \widehat{AUC}_i = \sum_{i=0}^{n} \frac{(f(x_{i-1}) + f(x_i)}{2} \Delta xi \tag{2}$$

Where x_i is false positive rate for a given i^{th} cut-off point, where $i = 1, \dots, n$ and function f is a plotted ROC curve, so $f(x_i)$ is a true positive rate for i^{th} cut-off point.

This method is according to Hanley and McNeil (1982) valid since they showed that is equivalent to Mann-Whitney statistics (Wilcoxon statistics), which is, as again demonstrated by Hanley and McNeil (1982), an unbiased estimator of true area under curve. The study further demonstrated that the standard error of the area under ROC curve is calculated from the variance of the Wilcoxon statistic as follows:

$$SE\left(\widehat{AUC}\right) = \sqrt{\frac{\widehat{AUC}\left(1 - \widehat{AUC}\right) + (n_F - 1)\left(Q_1 - \widehat{AUC}^2\right) + (n_{NF} - 1)\left(Q_2 - \widehat{AUC}^2\right)}{n_F n_{NF}}} \tag{3}$$

Where: n_F is a number of failed firms in the sample, n_{NF} is a number of non-failed firms in the sample, Q_1 is the probability that two randomly selected firms will be both classified as having a higher chance of failure than one randomly selected non-failed firm, calculated as $Q_1 = \frac{\widehat{AUC}}{2 - \widehat{AUC}}$ and Q_2 is the probability that one randomly selected failed firm will be classified as having higher chance of failure than two randomly selected non-failed firms – calculated as $Q_2 = \frac{2\widehat{AUC}^2}{1 + \widehat{AUC}}$.

In order to compare the AUC of two models the normally distributed critical ratio Z from Hanley and McNeil (1983) is going to be used.

$$Z = \frac{\widehat{AUC_1} - \widehat{AUC_2}}{\sqrt{(SE\,(\widehat{AUC_1}))^2 + (SE\,(\widehat{AUC_2}))^2 - 2rSE(\widehat{AUC_1})SE(\widehat{AUC_2})}} \tag{4}$$

Where r is estimated correlation between AUC_1 and AUC_2. To obtain r, Hanley and McNeil (1983) proposed a following procedure – firstly calculate the correlation between the models on a bankrupted and non-bankrupted sub-sample ($\widehat{r_b}\,, \widehat{r_{nb}}$). Then calculate the $\hat{r}$ which is an average value of $\widehat{r_b}\,, \widehat{r_{nb}}$ and $\overline{AUC}$, which is an average value of $\widehat{AUC_1}$ and $\widehat{AUC_2}$. Finally, correlation r is determined from a statistical table published in Hanley and McNeil (1983) using the values $\hat{r}$ and $\overline{AUC}$. The correlation matrix of r is attached in the appendix as Table 17.

Moreover, as comparing the single AUC numbers seems insufficient since they are not very meaningful from a statistical perspective, Engelmann et al. (2003) propose the accuracy ratio (AR) and showed that it is simply a linear transformation of AUC, calculated as:

$$AR = 2(AUC\ - 0.5) \tag{5}$$

Accuracy ratio shows how close the model is to a perfect model, where 0 is a random model and 1 is a perfect model.[17]

To build a ROC and compute AUC, R software is going to be used, particularly, R package pROC by Xavier et al. (2011) is going to be employed.

6.2. Test of economic value

For a performance test, when the misclassification costs are being different we are going to use a test of economic value. The motivation is simple, although, ROC is a widely used method in the bankruptcy validation, it assumes that the misclassification costs are considered to be equal. As discussed in the literature review, misclassification costs are not considered to be the same. Hence, to tackle the issue of misclassification

[17] Accuracy ratio is basically a Gini coefficient.

costs, studies such as Agarwal and Taffler (2008), Bauer and Agarwal (2014) or Wood (2012) proposed a test of economic value.

The test of economic value is going to simulate a simplified artificial economy – credit market, where banks are going to compete, in a loan market, for customers. In the test, each previously presented credit scoring model will represent a bank, which will be loaning money to companies based on the score of assigned model. Companies will always choose a loan with the most favorable interest. Whenever more banks offer the same interest, the loan will be split equally among these banks (basically the banks will offer a syndicated loan with the equal share).

The size of the loan market is assumed to be €100 billion and to keep the analysis tractable all loans are going to have the same size. Also, every loan is for one year, so a company can borrow multiple times. Hence, each company from 19596 observations will be seeking €5.1 million loan.[18] For the tractability reason, all companies will be assigned with the same loss given default (LGD). Altman et al. (1977) suggested that costs of type I error would be 70%. However, more recent European study Agarwal and Taffler (2008) propose that in line with Basel II requirements the LGD constant should be equal to 45%. Same value of constant was also used in the study of Bauer and Agarwal (2014), under Basel III requirements. Thus, this thesis will also assume LGD to be 45% of the loan value. On the other hand, Altman (1977) determined the cost of type II to be 2%. Nevertheless, in the test of Economic value type I error will not be directly determined but will indirectly affect a bank profitability.

The pricing framework for models will follow Wood (2012). Each model ranks companies from the lowest probability of default to highest and then separate companies into 20 groups. The general rule is that the riskier loan, the higher premium. The 5% of firms with the safest companies will be ascribed a premium of 0.3%, and on the other hand, the 5% of the riskiest companies will not be offered loan at all. The overall credit spread matrix proposed by Wood is determined as in Table 11 below.

[18] Since each model will refuse to loan money to the bottom 5%, the overall market does not have to sum up to 100 €billion.

Table 11: Credit spread for models

Loan quality	*Group*	*Premium*
Highest quality	1	0.30%
	2	0.55%
	3	0.80%
	4	1.05%
	5	1.30%
Good quality	6	1.55%
	7	1.80%
	8	2.05%
	9	2.30%
	10	2.55%
Medium quality	11	2.80%
	12	3.05%
	13	3.30%
	14	3.55%
	15	3.80%
Poor quality	16	4.05%
	17	4.30%
	18	4.55%
	19	4.80%
No Loan	20	N/A

Notes: The Table is adopted from Wood (2012, p. 277).

To assess the economic value of using different models for loan pricing, the return on assets (ROA) is going to be an indicator of model quality. ROA will be calculated from net profit generated by a bank divided by total assets loaned.

7. Results

The following sections provide a detailed overview of results for every model tested. Moreover, hypotheses are going to be addressed and discussed. Also, results are going to be compared with existing literature.

7.1. Results of ROC analysis

The ROC results are presented in Table 12, where the predictive power is interpreted by AUC and AR. Moreover, test of significance is presented in Table 13 and graphical interpretation of ROCs can be seen in Figure 2.

Seven out of nine tested models are multivariate models estimated via MDA or logit method, other 2 are univariate models serving predominantly as a benchmark. In the Table 13 can be seen that 8 out of 9 models proved to be better than a random model at the 1% level and only single model – the updated version of O-score at 5% level, meaning that accounting models report substantial ability to detect bankruptcy. In addition, all 8 models, which report significance at 1% level have extremely high Z scores (over 10) implying unquestionable predictive powers.

Focusing on the predictive power, AUC of models range from 56% to 85%. Interestingly, both extremes are results from Ohlson model. In our sample 6 from 7 multivariate models have beaten univariate models, where univariate models cover on average 76% of the plot area, whereas multivariate models 77.3%.

The best model – original Ohlson formula (O-score) showed an accuracy of 0.70, followed by updated Z''-score, Z''-score and Taffler's original T-score showed accuracy ratio of 0.63. The fifth most powerful model is Z-score model with solid accuracy ratio of 0.60 (AUC=80.2%). Moreover, Altman's original Z-score formula outperformed both its updated version from Wood (2012) with AR of 0.51 and univariate benchmarks. Surprisingly, univariate models performed relatively well with accuracy ratio being 0.51 and 0.53 for WC/TA and Beaver's ratio respectively. Lastly, the worst performing model is updated Ohlson from Hillegeist et al. (2004) with poor AR of 0.12.

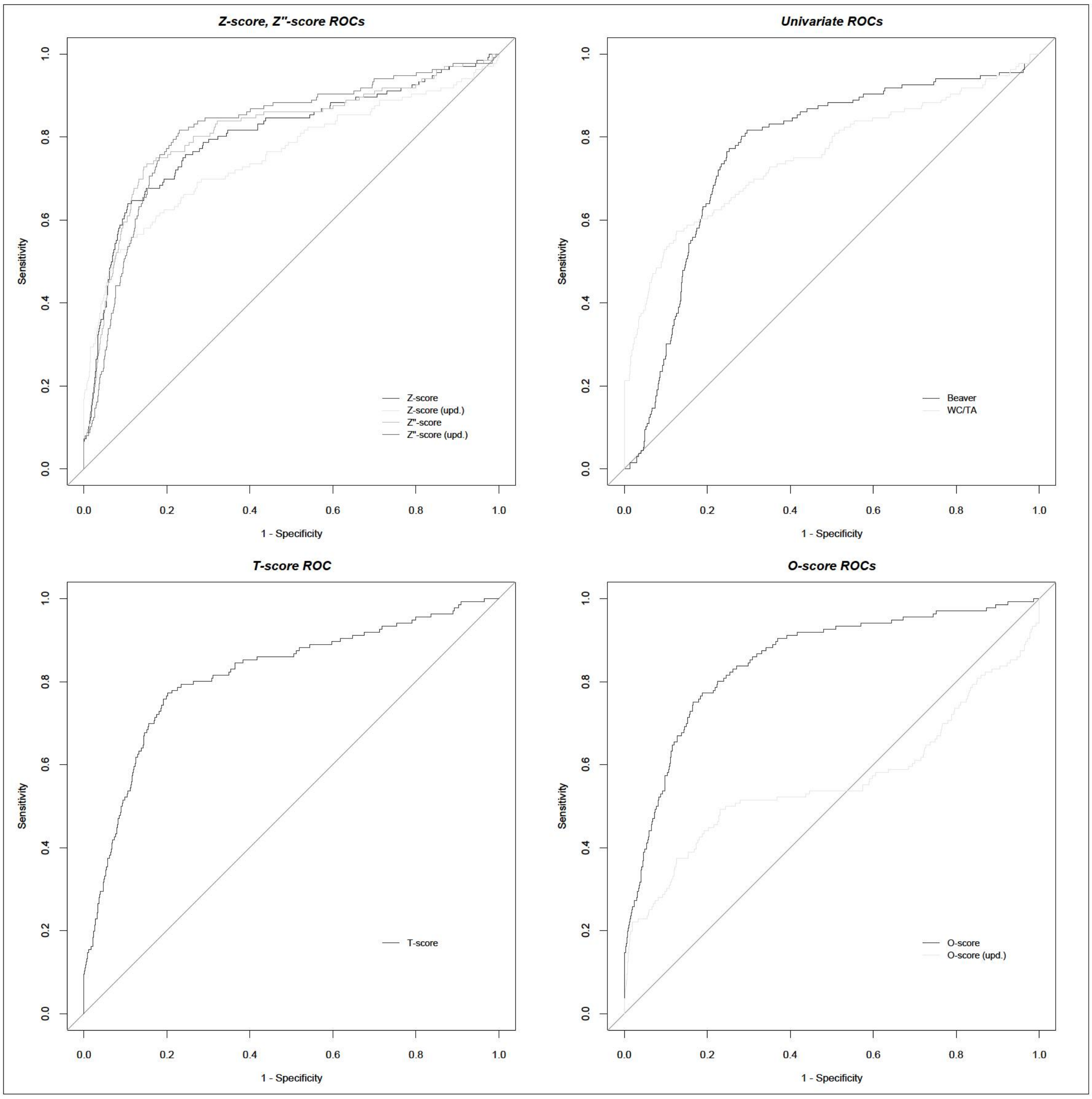
Table 12: Area under the ROC curve and accuracy ratios

Model	*AUC*	*SE(AUC)*	*Z*	*AR*
Z-score	80.2%	2.3%	13.22***	0.60
Z-score (upd.)	75.4%	2.4%	10.48***	0.51
Z"-score	81.4%	2.2%	14.02***	0.63
Z"-score (upd.)	81.7%	2.2%	14.21***	0.63
T-score	81.4%	2.2%	14.05***	0.63
O-score	85.0%	2.1%	16.81***	0.70
O-score (upd.)	56.0%	2.6%	2.35**	0.12
WC/TA	75.4%	2.4%	10.48***	0.51
Beaver	76.7%	2.4%	11.15***	0.53

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively. AUC is estimated using trapezoidal rule, standard errors are calculated following Hanley and McNeil (1982). Z shows the statistic on deviation from a null hypothesis that AUC is 0.5 (random model).

The AUC values achieved by our multivariate models in the setting of Europe between years 2012 - 2017 are relatively similar to values reported in prior studies. Reisz and Perlich (2007) report AUC at 78% for the original Altman model, using US data from 1988 - 2002, which is similar to Wood (2012) who reports AUC at 78.7% for the UK between 2006 - 2009. Moreover, our results support results from Wood that original Z-score performs better than updated formula. However, Altman et al. (2017) report relatively lower AUC for both Z''-score and his updated Z''-score 74.7% and 74.3% respectively for mostly on European private companies from 2007 - 2010.

Figure 2: ROC curves



Notes: 45°line is representing a random model.

The test of significance presented in Table 13 show us that all models are significantly more powerful than an updated O-score model. Moreover, all models except for univariate model from Beaver have significantly better predictive abilities than WC/TA and updated Z-score model. Hence, we can categorize these 3 models as low performing. On the other hand, there are only 3 models significantly better than Beaver's ratio – Ohlson model (significance at 1%), updated Z''-score (at 10%) and T-score (at 10%). Further, from Table 13 can be seen that statistical superiority of O-score has been proven only for the 4 worst performing models.

Table 13: Test of significance between models

Z	*WC/TA*	*Z-score*	*Z-score (upd.)*	*T-score*	*Z''-score*	*Z''-score (upd.)*	*Beaver*	*O-score*
Z-score	-2.21**							
Z-score (upd.)	0.00	2.16**						
T-score	-3.00***	-0.56	-3.02***					
Z''-score	-2.94***	-0.84	-2.81***	0.02				
Z''-score (upd.)	-2.70***	-0.72	-2.79***	-0.12	-0.19			
Beaver	-0.38	1.12	-0.38	1.71*	1.48	1.72*		
O-score	-3.35***	-1.63	-3.37***	-1.39	-1.33	-1.30	-3.38***	
O-score (upd.)	6.11***	8.19***	6.00***	7.92***	8.44***	8.60***	6.43***	9.45***

Notes: ***, **, * significant at the 1%, 5%, 10% level respectively. Test is following Hanley and McNeil (1983) comparing AUC of 2 models, where null hypothesis is that models are being equal.

Looking at results, it can be said that Altman's original formula is relevant in current economic conditions in Europe. The model proved to be significantly better than a random model, implying a sign of predictive power. Z-score achieved relatively high 80.2% AUC, as the area under curve higher than 80% could be considered according to studies which are testing the bankruptcy prediction for short-term horizon as a robust value.[19] Furthermore, Z-score performs statistically better than a univariate model and one of its variable WC/TA. Although the superiority of Altman (1968) Z-score over the Beaver cash-flow ratio was statistically not significant (Z=1.12), the superiority of other accounting-based models has also not been statistically proven. Therefore, it can be stated that by using the Altman Z-score user is getting a robust performance which is at worst comparable with the Beaver ratio, nevertheless, the performance may be comparable with other in analysis superior accounting-based multivariate models.

The original formulas for all O-score, T-score, Z-score, and Z''-score surprisingly performed very well and showed its robustness given changing economic and political conditions over the years. On the other hand, performance of updated models could be considered as disappointing. Only one from three updated models showed performance

[19] The best performing market-based models do not exceed 90% AUC – see Bauer and Agarwal (2014), Agarwal and Taffler (2008). Moreover, Altman et. Al (2017) defines AUC>80% as a good result.

similar to its original version. Country-wise results are inconsistent – the best model is estimated using US data, on the other side, T-score (UK data) and updated Z''-score (predominantly European data) performed very well. The geographical effect will be probably difficult to find, since the data are not comparable - US has different bankruptcy policies, which, however, seems to be more unified than policies in Europe, which could cause better overall sample for a model estimation.

Overall, we have found no evidence that calibrated models show better predictive abilities and therefore we cannot support the hypothesis that calibrated models are performing better. It maybe does not make sense to solely update models without further consideration of adding or changing variables. Agarwal and Taffler (2007) provide an interesting discussion concluding that seeking an update of coefficients for models is generally incorrect. They suggest that if the original model will start to show signs of age, instead of re-estimating, new model, unconstrained by prior model ratios, should be built. Our results are leaning towards Agarwal and Taffler's remarks – looking back to the coefficients of updated models – it seems strange to omit ratios from original formula as each ratio were designed to add a unique predictive power.

Moreover, the cut-off point for Z-score model determined by Altman (1968), which was set to be 2.99 shows specificity of 36.72% and sensitivity of 88.97%, with overall sensitivity and specificity of 125.69%. The best cut-off point from the perspective of a minimal sum of errors (error I + error II), Z-score of 0.22 shows combined specificity and sensitivity of 153.46% (given equal misclassification costs), indicating that Altman's threshold is ineffective.

Therefore, for the practical usability, we have made a statistical table (Table 18 in appendix), where 50 cut-off points are presented, from 100% determination of bankruptcy (treating error II as costless) to the most effective cut-off (for equal misclassification costs), so Z-score model can be more effectively applied by practitioners as they can choose an ideal cut-off point given their preferences about misclassification costs. The statistical table was also made for the best performing model – Ohlson (1980) and can be found in the appendix as well (Table 19).

7.2. Results of economic value test

For different misclassification costs, we proposed a test of economic value, where banks are competing for loans. Results of the test are shown in Table 14 presenting revenue, profitability and other descriptive metrics for selected 9 models under the assumptions of simplified competitive loan environment described in section 6.2.

In the credit market, 29 856 loans were provided, where on average a company was serviced by 1.5 banks. Banks reached the return on assets ranging from 0.63% to 1.21%, with the average credit spread ranging from 0.77% to 1.28% demonstrating high competition between banks. The worst four performing models from ROC analysis produced the biggest losses, because of inconsiderate lending, where the rate of default in the portfolio was ranging from 0.52% to 1.16%. The rate of default was in these four instances significantly higher compared to the 0.17% to 0.31% rate for rest of the market participants.

The biggest revenue is generated by the bank using the updated O-score model, however the model has the biggest loss, which is not surprising given its very poor discriminative performance. On the other hand, bank utilizing Z''- score model has the smallest market share, showing difficulties to compete in the market. The best three performing banks make decisions based on O-score (the best performing model in ROC analysis), updated Z''-score (2nd) and Z-score (5th) respectively.[20] Moreover, the worst 3 models from ROC analysis are, besides Z''-score model, the least profitable relative to its assets.

Table 14: Economic value

	Z-score	*Z-score (upd.)*	*T-score*	*O-score*	*O-score (upd.)*	*Z"-score*	*Z"-score (upd.)*	*Beaver*	*WC/TA*
Credits	2 518	3 464	2 685	3 543	5 437	1 605	3 522	3 992	3 090
Market share	7.3%	11.4%	8.4%	11.3%	24.5%	3.6%	10.9%	13.1%	9.5%
Defaults	6	18	8	6	63	5	8	41	18
Defaults/credits	0.24%	0.52%	0.30%	0.17%	1.16%	0.31%	0.23%	1.03%	0.58%
Avg. credit spread	1.20%	1.21%	1.10%	1.28%	1.23%	0.77%	1.20%	1.64%	1.23%
Revenue (€m)	87.62	137.65	92.00	144.80	301.95	28.03	130.36	214.91	116.42
Loss (€m)	-9.95	-32.09	-7.60	-7.60	-133.96	-5.30	-7.98	-75.34	-32.48
Profit (€m)	77.67	105.55	84.40	137.20	168.00	22.73	122.38	139.57	83.95
ROA	**1.07%**	**0.93%**	**1.01%**	**1.21%**	**0.69%**	**0.63%**	**1.12%**	**1.07%**	**0.88%**

Notes: Revenue = market size * market share * average credit spread; Loss = market size * (136/19596) * share of defaulters * average share on defaulted loans * LGD; Profit = revenue - loss. Return on assets = profit divided/ (market size * market share), where market size is €100 billion, LGD is 45%.

[20] Z-score ROA is 1.0664%, which is slightly higher than Beaver's 1.0656%.

The test showed that the best two performing models from ROC analysis have the best performing capabilities given different misclassification costs in a simplified economy. Besides poorly performing bank using Z''-score model, results seem to be relatively stable, implying that better performing models tend to reach higher profitability relative to the amount of assets lent. Moreover, the results hold after robustness check, where we omit the worst performing model from ROC analysis (updated Ohlson).[21]

Although the test was proceeded using simplified conditions, which were set for tractability reasons, we provided solid evidence to conclude that we are unable to reject the hypothesis that most accurate models will generate the biggest profitability as the original Ohlson and Z''-score models showed the best performance for both ROC and Economic value tests. Furthermore, Altman Z-score showed robust performance with a very low default rate of 0.24% and overall return of assets 1.07%, which is 88.4% performance of bank utilizing O-score, compared to 52% of the worst bank.

[21] O-score and Z''-score were still the best performing and overall ranking was almost the same, where only Z-score and Beaver model switched positions.

8. Conclusion

The thesis has focused on the Altman (1968) Z-score and other credit-scoring models, which utilizing financial statements for predictions. The financial statements as a tool for bankruptcy prediction confirmed its predisposition as from the 25 financial ratios tested 24 of them differed for bankrupted and non-bankrupted companies at least at 5% significance level.

The study provides a comparison of the well-known Altman Z-score model with other famous accounting-based models such as O-score from Ohlson (1980), UK based Z-score model from Taffler (1983) or Altman's extension of Z-score for also private firms called Z''-score. The primary goal was to test whether the original Z-score formula is relevant after 50 years of use. The test was conducted on the European environment in order to provide meaningful insight into the less researched market compared to the USA, where model originated. We have found evidence that Altman Z-score model discriminatory power is still relevant as the model showed solid predictive abilities given its age. The predictive power seems to be at least as good as univariate Beaver (1966) model and significantly better than only its part – in literature popular ratio WC/TA. The relevance of Altman was underlined by the fact that even O-score model, which was the best performing model was not significantly better than Z-score.

Moreover, versions of models with updated coefficients did not perform substantially better than original formulas, supporting the Agarwal and Taffler (2007) remarks that updating the coefficients of original models using recent data unlikely improves the discriminatory power. Hence, we, similarly as Agarwal and Taffler (2007) incline towards building a new model if the original model starts to show signs of age. However, that was not the case of Z-score model, which proved its relevance and revealed to be significantly better than its updated version from Wood (2012). Since the performance of Z-score proved to be satisfactory, we provided a table of new cut-off points for financial practitioners as the original cut-off point appeared to be inefficient.

Finally, thesis tackled the issue of model performance given different misclassification costs, showing that the models which are having overall higher discriminatory abilities tend to be more powerful in an environment where misclassification costs are not equal. Nevertheless, the issue of misclassification cost is a complex topic which deserves its own detailed researched since a test of economic value was based on simplified assumptions.

References

[1] Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, *32*(8), 1541-1551.

[2] Agarwal, V., & Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: Does it really have predictive ability? *Accounting and Business Research*, *37*(4), 285-300.

[3] Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting*, *28*(2), 131-171.

[4] Altman, Edward I (1983). Corporate financial distress: a complete guide to predicting, avoiding, and dealing with bankruptcy. *Wiley, New York*

[5] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23(4), 589-609.

[6] Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of banking & finance*, *1*(1), 29-54.

[7] Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, *40*, 432-442.

[8] Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 4, 71-111.

[9] Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial education*, 1-42.

[10] Bharath, S. T., & Shumway, T. (2004). Forecasting default with the KMV-Merton model. *Working Paper Series*, University of Michigan

[11] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of political economy, 81(3), 637-654.

[12] Bureau of Business Research (1930). A Test Analysis of Unsuccessful Industrial Companies. *Bulletin No. 31. Urbana: University of Illinois Press*.

[13] Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting research*, 167-179.

[14] Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative analysis*, *7*(2), 1477-1493.

[15] Eidleman, G. (1995). Z-scores – a guide to failure prediction. *The CPA Journal*. 12(9), 52-53.

[16] Eisenbeis, R. A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal of Banking & Finance*, *2*(3), 205-219.

[17] Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing rating accuracy. *Risk*, *16*(1), 82-86.

[18] Falkenstein, E., Boral, A., & Carty, L. (2000). RiskCalc for private companies: Moody's default model. *Moody's KMV.*

[19] Frank, R. E., Massy, W. F., & Morrison, D. G. (1965). Bias in multiple discriminant analysis. *Journal of Marketing Research*, 250-258.

[20] Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Human Genetics*, *8*(4), 376-386.

[21] Fitzpatrick, P. J. (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms. The CPA Journal, 12(3), 598-605.

[22] Guerard Jr, J. B., & Schwartz, E. (2007). *Quantitative corporate finance*. Springer Science & Business Media.

[23] Hanley, J. and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143 (1): 29–36.

[24] Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, 148(3), 839-843.

[25] Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of accounting studies*, *9*(1), 5-34.

[26] Jackson, R. H., & Wood, A. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review*, *45*(3), 183-202.

[27] Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. Journal of accounting research, 380-395.

[28] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. The Journal of finance, 29(2), 449-470.

[29] Merton, R. C. (1973). Theory of rational option pricing. The Bell Journal of economics and management science, 141-183.

[30] Merwin, C. L. (1942). Financing small corporations in five manufacturing industries, 1926-36. NBER Books.

[31] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.

[32] Padmalatha, S. (2011). *Management of Banking And Financial Services, 2/E*. Pearson Education India.

[33] Reisz, A. S., & Perlich, C. (2007). A market-based framework for bankruptcy prediction. *Journal of Financial Stability*, *3*(2), 85-131.

[34] Scott, E. (1978). On the Financial Application of Discriminant Analysis: Comment. Journal of Financial and Quantitative Analysis, 13(1), 201-205.

[35] Smith, R. F. (1935). Changes in the Financial Structure of Unsuccessful Industrial Corporations, by Raymond F. Smith... and Arthur H. Winakor... University of Illinois.

[36] Sobehart, J. R., S. C. Keenan, and R. M. Stein. 2000. "Benchmarking quantitative default Risk Models: A validation methodology." New York. Moody's Risk Management Services.

[37] Stein, R. M. (2002). Benchmarking default prediction models: Pitfalls and remedies in model validation. Moody's KMV, New York, 20305.

[38] Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. Journal of Banking & Finance, 29(5), 1213-1236.

[39] Taffler, R. J. (1983). 'The assessment of company solvency and performance using a statistical model'. *Accounting and Business Research*, 15(52): 295–308.

[40] Tudela, M., & Young, G. (2003). Predicting default among UK companies: a Merton approach. *Financial Stability Review*, June 2003.

[41] Vernimmen, P., Quiry, P., Dallocchio, M., Le Fur, Y., & Salvi, A. (2014). *Corporate finance: theory and practice*. John Wiley & Sons.

[42] Walter, J. E. (1957). Determination of technical solvency. *The Journal of Business*, *30*(1), 30-43.

[43] Wood, A. P. (2012). The performance of insolvency prediction and credit risk models in the UK: A comparative study, development and wider application.

[44] Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics*, *6*(1), 34-45.

[45] Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves". *BMC Bioinformatics*, 12, p. 77.

Appendix

Table 15: Sample description – Geographic distribution

Country	*NB*	*B*	*Overall*
Austria	194	-	194
Belgium	465	1	466
Bulgaria	304	-	304
Croatia	238	14	252
Cyprus	-	1	1
Czech Republic	47	-	47
Denmark	547	1	548
Estonia	77	-	77
Finland	489	2	491
France	2370	23	2393
Germany	2304	33	2337
Greece	834	2	836
Hungary	101	1	102
Ireland	274	-	274
Italy	806	4	810
Latvia	83	2	85
Lithuania	49	-	49
Luxembourg	201	1	202
Malta	71	-	71
Netherlands	519	2	521
Poland	2264	19	2283
Portugal	196	-	196
Romania	478	4	482
Slovakia	129	-	129
Slovenia	121	-	121
Spain	507	5	512
Sweden	1450	9	1459
United Kingdom	4342	12	4354

Notes: NB means non-bankrupted, B means bankrupted companies.

Table 16: Sample description – Industry distribution

Industry	*NB*	*B*	*Overall*
Chemicals, rubber, plastics, non-metallic	1848	7	1855
Construction	674	5	679
Education, Health	332	-	332
Food, beverages, tobacco	840	5	845
Gas, Water, Electricity	567	4	571
Hotels & restaurants	386	3	389
Machinery, equipment, furniture, recycling	3321	23	3344
Metals & metal products	925	5	930
Other services	5444	39	5483
Post & telecommunications	543	5	548
Primary sector	774	10	784
Public administration & defense	54	-	54
Publishing, printing	717	2	719
Textiles, wearing apparel, leather	495	5	500
Transport	540	-	540
Wholesale & retail trade	1629	19	1648
Wood, cork, paper	371	4	375

Notes: Industries are divided according to ORIBS database.

Table 17: Estimated correlation of two AUCs

r	*WC/TA*	*Z-score*	*Z-score (upd.)*	*T-score*	*Z''-score*	*Z''-score (upd.)*	*Beaver*	*O-score*
Z-score	0.58	x						
Z-score (upd.)	0.93	0.56	x					
T-score	0.64	0.54	0.64	x				
Z''-score	0.63	0.81	0.59	0.64	x			
Z''-score (upd.)	0.51	0.59	0.54	0.59	0.79	x		
Beaver	0.08	0.10	0.07	0.28	0.06	0.22	x	
O-score	0.21	0.11	0.22	0.31	0.22	0.31	0.41	x
O-score (upd.)	0.19	0.26	0.16	0.11	0.22	0.23	0.16	0.14

Notes: r is calculated according to Hanley and McNeil (1983).

Table 18: Altman Z-score (1968) cut-off points

	Threshold	Specificity	Sensitivity	Spec+Sens		Threshold	Specificity	Sensitivity	Spec+Sens
1	24.20	2.42%	100.00%	102.42%	26	1.60	65.37%	81.62%	146.99%
2	21.80	2.68%	99.26%	101.94%	27	1.59	65.61%	80.88%	146.49%
3	12.05	5.34%	98.53%	103.87%	28	1.50	67.69%	80.15%	147.83%
4	11.60	5.62%	97.79%	103.42%	29	1.38	69.96%	79.41%	149.38%
5	6.56	11.90%	97.06%	108.96%	30	1.32	71.34%	78.68%	150.02%
6	5.86	13.88%	96.32%	110.20%	31	1.31	71.43%	77.94%	149.37%
7	5.50	15.11%	95.59%	110.70%	32	1.29	72.09%	77.21%	149.30%
8	5.26	16.09%	94.85%	110.95%	33	1.20	73.74%	76.47%	150.21%
9	4.89	17.91%	94.12%	112.03%	34	1.12	75.49%	75.74%	151.23%
10	4.63	19.39%	93.38%	112.77%	35	1.09	76.10%	75.00%	151.10%
11	4.38	20.99%	92.65%	113.64%	36	1.07	76.41%	74.26%	150.67%
12	4.03	23.62%	91.91%	115.53%	37	1.06	76.50%	73.53%	150.03%
13	3.70	26.94%	91.18%	118.11%	38	1.01	77.35%	72.79%	150.15%
14	3.47	29.41%	90.44%	119.85%	39	0.98	77.94%	72.06%	150.00%
15	3.12	34.60%	89.71%	124.31%	40	0.97	78.08%	71.32%	149.40%
16	2.98	36.91%	88.97%	125.88%	41	0.96	78.21%	70.59%	148.80%
17	2.76	40.58%	88.24%	128.81%	42	0.85	80.65%	69.85%	150.51%
18	2.75	40.71%	87.50%	128.21%	43	0.82	80.99%	69.12%	150.10%
19	2.63	42.83%	86.76%	129.60%	44	0.79	81.71%	68.38%	150.09%
20	2.53	44.62%	86.03%	130.65%	45	0.60	84.98%	67.65%	152.63%
21	2.49	45.53%	85.29%	130.82%	46	0.58	85.33%	66.91%	152.24%
22	1.99	56.18%	84.56%	140.74%	47	0.57	85.49%	66.18%	151.66%
23	1.98	56.69%	83.82%	140.51%	48	0.56	85.60%	65.44%	151.04%
24	1.91	58.12%	83.09%	141.21%	49	0.33	88.49%	64.71%	153.20%
25	1.91	58.14%	82.35%	140.50%	**50**	**0.22**	**89.49%**	**63.97%**	**153.46%**

Notes: 50 possible cut-off points from 100% prediction accuracy of bankruptcy, where investors do not care specificity, to the point where misclassification costs are being equal (maximization of specificity and sensitivity).

Table 19: Ohlson (1980) cut-off points

	Threshold	Specificity	Sensitivity	Spec+Sens		Threshold	Specificity	Sensitivity	Spec+Sens
1	0.04%	1.27%	100.00%	101.27%	19	38.51%	68.07%	86.76%	154.83%
2	0.93%	7.55%	99.26%	106.82%	20	40.13%	69.04%	86.03%	155.07%
3	1.51%	10.44%	98.53%	108.97%	21	41.18%	69.67%	85.29%	154.97%
4	2.04%	12.75%	97.79%	110.55%	22	42.10%	70.16%	84.56%	154.72%
5	5.61%	24.87%	97.06%	121.93%	23	47.29%	72.94%	83.82%	156.76%
6	5.87%	25.62%	96.32%	121.95%	24	49.07%	73.86%	83.09%	156.95%
7	8.64%	32.75%	95.59%	128.34%	25	50.55%	74.61%	82.35%	156.96%
8	9.89%	35.64%	94.85%	130.49%	26	52.39%	75.43%	81.62%	157.05%
9	13.69%	43.01%	94.12%	137.13%	27	53.67%	75.99%	80.88%	156.87%
10	17.38%	48.95%	93.38%	142.33%	28	57.44%	77.50%	80.15%	157.65%
11	19.58%	51.96%	92.65%	144.61%	29	5.66%	77.63%	79.41%	157.04%
12	25.37%	58.29%	91.91%	150.21%	30	58.86%	78.11%	78.68%	156.79%
13	28.41%	60.88%	91.18%	152.06%	31	59.93%	78.55%	77.94%	156.49%
14	31.06%	63.07%	90.44%	153.51%	32	66.72%	81.18%	77.21%	158.38%
15	31.54%	63.43%	89.71%	153.14%	33	67.87%	81.57%	76.47%	158.04%
16	32.65%	64.28%	88.97%	153.25%	34	68.96%	82.07%	75.74%	157.81%
17	35.11%	65.85%	88.24%	154.08%	**35**	**72.65%**	**83.45%**	**75.00%**	**158.45%**
18	36.59%	66.89%	87.50%	154.39%					

Notes: 35 possible cut-off points from 100% prediction accuracy of bankruptcy, where investors do not care specificity, to the point where misclassification costs are being equal (maximization of specificity and sensitivity).

Table 20: Economic test – loan participation

	Z-score	*Z-score (upd.)*	*Taffler*	*Ohlson*	*Ohlson (upd.)*	*Z''-score*	*Z''-score (upd.)*	*Beaver*	*WC/TA*
Avg. Size of loan	2.89	3.29	3.13	3.20	4.51	2.26	3.09	3.28	3.07
% of "stand-alone loans"	57%	64%	61%	63%	88%	44%	61%	64%	60%
% of syndicated loans	43%	36%	39%	37%	12%	56%	39%	36%	40%

Notes: Stand-alone loan means a loan where only 1 bank participates. Average size of loan is in million €.