The accuracy and information content of two traditional statistical methods in forecasting bankruptcy: evidence from European countries

MSc thesis in the area of Business Administration
Name: Bas Machielsen
Student no.: s1131044
University: University of Twente
Faculty: School of Management and Governance
Supervisors: Prof. dr. R. Kabir
Dr. X. Huang

Contents

1. Introduction 1

2. Literature review 4

  2.1 Statistical bankruptcy prediction models 4

    2.1.1 Discriminant analysis 5

    2.1.2 Nonlinear probability models 9

    2.1.3 Distance-to-default probability model 14

    2.1.4 Hazard models 20

  2.2 Intelligent and miscellaneous bankruptcy prediction models 26

  2.3 Performance evaluation 31

    2.3.1 Classification tables 31

    2.3.2 Receiver operating characteristics 32

    2.3.3 Other goodness-of-fit measures 33

    2.3.4 Information content 34

    2.3.5 Economic value when misclassification costs are different 35

  2.4 Summary and implications for this study 37

3. Hypotheses, Methodology, Variables, Data 41

  3.1 Hypotheses 41

  3.2 Methodology 44

    3.2.1 Hypothesis 1: Econometric method performance 44

    3.2.2 Hypothesis 2: Stationarity in the predictor variables 46

    3.2.3 Hypothesis 3: Performance relative to simple benchmark 47

    3.2.4 Hypothesis 4: Robustness to alternative sets of predictors 48

  3.3 Variables 49

    3.3.1 Dependent variables 49

    3.3.2 Independent and control variables 49

    3.3.3 Alternative bankruptcy prediction models 51

  3.4 Data 52

  3.4 Summary of predictor variables and alternative models 56

4. Results 57

  4.1 Descriptive statistics 57

  4.2 Hypothesis tests 60

    4.2.1 Hypothesis 1: Econometric model performance 60

    4.2.2 Hypothesis 2: Stationarity in the predictor variables 64

4.2.3 Hypothesis 3: Performance relative to simple benchmark 70
4.2.4 Hypothesis 4: Robustness to alternative sets of predictors....................................73
5. Conclusion and discussion..77
5.1 Research question 77
5.2 Summary of results 78
5.3 Limitations and suggestions for further research..80
Appendix..83
Appendix A: Mathematical specification of discriminant analysis....................................83
Appendix B: Mathematical specification of maximum-likelihood estimation....................83
Appendix C: Annual Default Rates 84
Appendix D: Ratio’s used for Factor Analysis..85
Appendix E: Additional descriptive statistics..86
Appendix F: Complete parameter estimates – hypothesis 1 90
Appendix G: Complete parameter estimates – hypothesis 2, 3, 4 92
Appendix H: Goodness of fit measures for MDA and logistic regression 95
References...96

Abstract: This study encompasses a treatment of underlying theory and specifications of various bankruptcy prediction models. It assesses the accuracy and information content of some of these models for publicly listed firms in the European Union. The models used to forecast bankruptcy include Altman (1968)'s Z-score model and the logistic regression model by Ohlson (1980). It is investigated what predictors are best, whether the model performance declines over time and which model does the best job of forecasting bankruptcies. The performance is assessed in an estimation sample (2005-2007) and in a hold-out sample (2011-2013). Model performance is evaluated alongside several criteria. It is found that both models lose a significant amount of predictive accuracy when applied out-of-sample, but they still carry information content. Many tests show that a significant amount of information is left uncaptured by the models. This generalizes through various sets of predictors. The findings are subjected to several benchmarks and robustness tests.

Keywords: *corporate failure, bankruptcy prediction, credit risk, z-score, logistic regression, discriminant analysis*

ACKNOWLEDGEMENTS

This Master Thesis came to be as a final result required for obtaining the degree of Master of Science in Business Administration at the University of Twente, Enschede, The Netherlands.

I would like to express my gratitude towards, and admiration for the scientific guidance of my two supervisors, prof. dr. Rezaul Kabir and dr. Xiaohong Huang, both of whom offered much worthwhile advice and suggestions to substantially improve my research.

Finally, I would like to thank my family and friends for their interest, support and patience.

Of course, all errors and mistakes remain solely the responsibility of the author.

Bas Machielsen,

November 2015

1. Introduction

The topic of bankruptcy prediction is a long standing topic in academia and has been extensively researched since the inception of Altman's Z-score model (see Altman, 1968). Since then, several models have been developed and used in practice to predict bankruptcy. In recent years, especially during the ongoing economic crisis, increased attention has been paid to bankruptcy prediction models. The main problem that will be addressed in this study is to what extent two commonly-used statistical models are able to predict corporate bankruptcy. The central research question is **'**what is the accuracy and information content of two statistical models predicting bankruptcy for a sample of listed non-financial European firms?'

In the scientific literature, bankruptcy prediction models are usually divided into statistical bankruptcy prediction models, and non-statistical bankruptcy prediction models (see Kumar and Ravi, 2007). In this study, the focus lies on statistical models because by far the largest body of bankruptcy forecasting features statistical models (see Bellovary et al., 2007, Westgaard, 2007). However, it will also feature reviews of non-statistical bankruptcy prediction models, such as intelligent models, and some miscellaneous models. The research design is based predominantly on a combination of the methods of Altman (1968) and Ohlson (1980) but will feature some minor elements proposed and developed by, among others, Shumway (2001), Wu, Gaunt and Grey (2010) and Bauer and Agarwal (2014). The performance of these models is evaluated by assessing both in-sample and out-of-sample accuracy and information content. Accuracy serves as a metric for how many firms in the sample a model can predict correctly. Information content measures whether one model score contains more information about bankruptcy than another variable (or set of variables). The latter is assessed by means of the hazard modeling approach of Shumway (2001), employed by e.g. Chava and Jarrow (2004), Campbell et al., (2008) and in a slightly different way by Wu, Gaunt and Gray (2010) and Bauer and Agarwal (2014). Thus, it builds on various research designs commonly used in academia to forecast bankruptcy.

The topic of bankruptcy prediction is important for many reasons. Nowadays, it is possible to obtain a variety of information about the risk status a company is exposed to, from many different sources, such as professional agencies, mass media, analysts, etc. In the process of evaluating a great amount of information, many people usually rely on some analyst's judgment (Tsai, 2009). However, various arbitrary factors can influence the result of the analysis. Bankruptcy prediction tools based on quantitative methods can be used by various actors as an alternative. It may be valuable to partners of a firm under scrutiny to inquire whether corporations can pay their debts. It may serve as an early warning to the corporation itself, so that the firm can enact reform policies when the forecast of financial distress is made. The ability to differentiate between sound corporations and troubled ones will reduce the (expected) cost of corporate failure or corporate financial distress. Financial distress models have been used extensively to scrutinize companies from various industries, ranging from manufacturing firms (e.g. Altman, 1968, Taffler, 1984) and corporations (e.g. Peel, 1986, Trujilo-Ponce et al., 2013) to banks (e.g. Martin, 1977, Lane et al., 1986, Canbas et al., 2005).

Among the interest groups in the prediction of financial distress or bankruptcy are creditors, auditors, stockholders and senior management, and employees of a firm (Kumar and Ravi, 2007, Bauer and Agarwal, 2014). Financial distress or eventually bankruptcy would affect them all alike. Bankruptcy prediction can also be used to assess the probability of bankruptcy that takes into account characteristics of specific sub-populations (e.g. SME's, private companies, listed companies, companies in the manufacturing industry, etc.), tuned to changes in the macro environment. Taking recent developments into account (i.e. the financial crisis of 2008, the debt crisis of 2010, implementation of Basel-III and the recession across Europe), the importance of scrutinizing banks and corporations has arguably become even more interesting and urgent for policy makers as well as for the general public.

Up until now there have been different, contradictory empirical results concerning the best-performing models and their inputs (compare e.g. the results of Hillegeist et al., 2004, Agarwal and Taffler, 2008, Trujillo-Ponce et al., 2013, Bauer and Agarwal, 2014). Research that compared one bankruptcy prediction model against another, often conflated evaluating the performance of particular sets of predictor variables with evaluating the performance of the econometric model (e.g. Grice and Dugan, 2003), violating a ceteris paribus clause. The literature in the past often evaluated the performance on a rather limited basis, i.e. either the accuracy or the information content, but not both. In most cases, there is also a lack of a proper benchmark with which the results are compared to. This research aims to evaluate model performance on the basis of multiple criteria and subjects the findings to several robustness tests. Furthermore, there is not much empirical research about model validity during the recent financial crisis. This research attempts to follow the line of Mensah (1984) and Begley et al. (1996) by addressing the problem of whether the model accuracy and information content change under changing macroeconomic circumstances, as changing circumstances can impact the stationarity of predictor variables as well as firm-invariant default risk. In this study, the events after credit crisis of 2007 and the subsequent European debt crisis of 2010 represent these changed circumstances.

The foremost contribution of this study is that it provides an empirical test in a cross-European context in two periods of different macroeconomic circumstances (i.e. the 2004-2007 and 2010-2013 period). Consequently, it provides for rigorous model robustness tests (cf. Grice and Ingram, 2001, Grice and Dugan, 2003), because if the models capture essential information about bankruptcy, model performance should remain stable over time, and the model should not only be expected to work in a country-specific context, but also in a cross-country context. Furthermore, existing empirical research predominantly focuses on U.S. (e.g. Altman, 1968, Ohlson, 1980, Shumway, 2001, Hillegeist et al., 2004) or UK (Agarwal and Taffler, 2007, 2008, Bauer and Agarwal, 2014) listed firms – except for one study, no research attempted to look at the EU as a whole as of yet. The models evaluated in this study will incorporate firm-level predictors as well as macroeconomic predictors, and inquire whether the inclusion of macroeconomic factors adds to or harms model performance. The results will be subjected to several benchmarks and controls to test the robustness along each evaluated performance characteristic. Third, this study features model evaluation in two dimensions – accuracy and information content - whereas many previous studies solely focused on predictive accuracy (e.g. Altman, 1968, Ohlson, 1980 but also Begley, Ming and Watts, 1996) or information content (e.g. Hillegeist et al, 2004). Finally, this study features many sets of predictor variables. Practitioners who have an interest in good bankruptcy prediction models might find it useful to see which set of variables performs best. From a scientific point of view, this is an investigation of whether model performance is sensitive to different sets of predictor variables.

This thesis is structured as follows: in chapter two, I present the theory underlying the different statistical bankruptcy prediction models. I review the econometric specification and assumptions of these models, and give a comprehensive overview of empirical results. Then, I elaborate on the methods used to evaluate bankruptcy prediction models. Finally, this chapter features a brief overview of the studies that feature major contributions in the bankruptcy prediction literature over time, and the implications for this present study. The subsequent chapter (3) elaborates on the research design. I provide several arguments based on the scientific literature for testing a variety of hypotheses. Afterwards, I describe estimation and evaluation procedures, by which each hypothesis can be tested, and provide arguments for the specific methodological choices I make. I conclude this section by providing details on the data and choice of variables. In chapter 4, I describe and interpret the results of the tests, relating them to the posed hypotheses. This is followed by a final conclusion and discussion in chapter 5. An appendix is added for several mathematical and other details.

2. Literature review

In this chapter, a systematic overview of the development of and improvements on bankruptcy prediction models used in academia will be given. Following Kumar and Ravi (2007), I make a distinction between research utilizing statistical bankruptcy prediction models, and research utilizing so called intelligent bankruptcy prediction models, including, among others, research utilizing neural networks, case-based reasoning and decision trees (Kumar and Ravi, 2007). The main focus will be on the statistical techniques, but the review will feature a comprehensive review of several intelligent techniques. The literature review seeks to explain the various methodologies and models utilized, to accentuate similarities and differences, and offer a systematic history of the introduction and methodological development of various models.

First, the focus will be on research using statistical techniques. Each technique will be treated extensively, followed by some general critique and limitations. Second, a section will be devoted to the different ways of evaluating the accuracy of such a model, and the different ways the models have been evaluated in the past. Third, a short, comprehensive section will treat the evaluation of the information content of a model and what is meant by information content. Fourth, a brief statement about the evaluation of the economic value of a model will be provided. Furthermore, in treating each model separately, this review seeks to give a comprehensive account of the empirical results yielded. The empirics sections will feature a short comparison between the empirical results yielded by statistical models with the empirical results yielded by intelligent techniques and the implications for future research. Finally, a summary table containing the most important contributions and findings, sorted by methodology, will be provided and the implications of recent research will be discussed.

2.1 Statistical bankruptcy prediction models

Similar to Bauer and Agarwal (2014), I propose to conceptualize the research focusing on statistical techniques into four different categories with regards to the kind of methodology (estimation procedure) used: (i) discriminant analysis models (e.g. Altman, 1968, Deakin, 1972, Grice and Ingram, 2001), (ii) logistic regression and similar models (Martin, 1977, Ohlson, 1980, Zmijewski, 1984), (iii) Distance-to-default models (e.g. Vassalou and Xing, 2004, Hillegeist et al., 2004, Shumway & Bharath, 2008), and (iv) hazard models (e.g. Shumway, 2001, Chava and Jarrow, 2004).

In treating the empirics of the models, I will accentuate differences between accounting-based and market-based predictors, because there is disagreement over whether any of these models should incorporate accounting-based information, market-based information, or both. Sloan (1996) argues that accounting data can be used to complement market data, since market prices are not an indication of company accounts in the present, but rather a prospect in the future.

In the first wave of research, researchers used mainly accounting variables in their models to predict bankruptcy. There are however, a few objections to merely taking accounting data into account. Agarwal and Taffler (2008) argue that accounting data are only reported periodically while they are subject to standards and might be influenced by individual opinion and bias.

Also, accounting numbers can be manipulated by the management. Nowadays, common practice is combining accounting and market data, for example in simple discrete time logit models following Shumway (2001). Chava and Jarrow (2004) use a mixture of accounting and market-based ratios consisting of profitability, liquidity as well as market volatility or market price. Campbell et al. (2008) use both accounting and market information by using ratios that contain both accounting and market data into one ratio.

In other words, there is no strict separation of accounting and market data anymore, and most research incorporates the informational benefit of both. Whether it is actually useful to do so, (i.e. whether it improves accuracy or information content) is however an empirical matter and will be discussed.

2.1.1 Discriminant analysis

Since the research of Altman (1968), prediction of financial distress and bankruptcy has incorporated statistical techniques instead of less systematic assessments. In prior research, academics (e.g. Beaver, 1966, 1968) tended towards ratio analysis in assessing and predicting financial distress in particular companies. The research was practically limited to comparing different accounting ratios of different firms in order to assess the degree of financial distress companies were in. Altman (1968) proposed a sound methodology, utilizing multivariate discriminant analysis (MDA) in order to provide a mathematically sound way to forecast financial distress, thereby starting a series of research which provides a rigorous framework of bankruptcy prediction utilizing various statistical techniques.

The purpose of the study of Altman (1968) is to attempt an assessment of the quality of ratio analysis as an analytical technique. The research improves on previous work which "compared a list of ratios individually for failed firms and a matched sample of non-failed firms. Observed evidence for five years prior to failure was cited as conclusive that ratio analysis can be useful in the prediction of failure." The poverty of sole ratio analysis without further data treatment was that "the methodology was essentially univariate in nature and emphasis was placed on individual signals of impending problems."

MDA was used because it classifies "an observation into one of several a priori groupings dependent upon the observation's individual characteristics (Altman, 1968)." This methodology fits the problem because the problem is classifying firms into either the 'bankrupt' class or the 'non-bankrupt' class, making use of some available financing ratios for each individual firm. MDA then derives from the data a linear combination (in this case of financial ratios) which discriminates best between both classes (in this case bankrupt or non-bankrupt).

2.1.1.1 Model description

In general, discriminant analysis combines an *n* number of independent variables into a linear function that discriminates best between two classes, i.e. such that the outcome of this function for the two separate classes is as distinct as possible. The function is of the form (Hair et al., 2006):

$$Z = v_1X_1 + v_2X_2 + \cdots + v_nX_n + C \qquad \text{Eq. (2.1)}$$

where $v_1, v_2, \ldots v_n$ = discriminant coefficients (weights) attached to $x_1, x_2, \ldots x_n$ = independent variables. The discriminant variate Z is the linear combination of independent variables that will discriminate best between the objects in the groups defined a priori. The function thus transforms individual variables to a single discriminant score or Z-value which is then used to classify the object firm. $v_1 \ldots v_n$ are determined such that the separation S between the two distributions (i.e. bankrupt and non-bankrupt) is as large as possible and accordingly that the discriminating ability is maximized.

Discriminant analysis is the appropriate technique for testing the hypothesis that the group means of a set of independent variables for two or more groups are equal. The test for statistical significance of the discriminant function is a generalized measure of the distance between so-called centroids. The centroid is defined as the arithmetic mean discriminant score of a class. Multiple discriminant analysis is unique in one characteristic compared to generic multivariate regression analysis: discriminant analysis is applied when the binary dependent variable is categorical, rather than ratio or interval. A mathematical treatment of the estimation produce can be found in appendix A.

Altman (1968) notes that "the primary advantage of MDA in dealing with classification problems [in comparison with ratio analysis] is the potential of analyzing the entire variable profile of the object simultaneously rather than sequentially examining its individual characteristics." MDA is utilized to establish a function which best discriminates between elements in two mutually exclusive groups: in our case, bankrupt and non-bankrupt firms. Z-scores are not explanatory theories of failure (or success) but pattern recognition devices.

2.1.1.2 Model assumptions

The most prominent assumption underlying discriminant analysis is multivariate normality. Multivariate normality is a generalization of the univariate normal distribution to higher dimensions. This means that the used set of (possibly) correlated real-valued predictors should cluster around a mean value. Multivariate normality is not easily tested for, but standard practice is to test for univariate normality using the Shapiro-Wilk test (see Lo, 1986, Hair et al., 2006). Nevertheless, in practice, violation of these tests is ignored because "a rejection of normality on the basis of such tests does not leave the econometrician with a clear alternate method of analysis (Lo, 1986)."

The second assumption is homogeneity of variances and covariances of the included variables for each class. This means that the variances in and covariances along the different levels of predictors should be roughly the same. This assumption can be tested through Box's M statistic (see Hair et al., 2006). Third, the model assumes no multicollinearity, because predictive power can decrease with an increased correlation between predictor variables. Fourth, MDA assumes independence between entries, i.e. random sampling, in this case that means that the variable entry of one firm is assumed to be independent of scores on that variable for all other firms: they must not be correlated with one another. The size of the smallest group must also be larger than the number of predictor variables. These relatively strict assumptions have made for extensive critique (Ohlson, 1980, Zmijewski, 1984, among others) on the usage and usefulness of discriminant analysis in predicting bankruptcy. In the case of bankruptcy prediction, the most restrictive assumption is probably the assumption regarding independent observations: most research has been conducted within one particular country. In the case of research encompassing multiple countries, nonrandom sampling might lead to biased estimates towards 'median' countries. The discriminating ability of the model appears to be sensitive to outliers. It would also seem doubtful whether financial ratios are in fact normally distributed.

The fact that Altman's bankruptcy prediction model is still used throughout academia leads to the question whether Altman's original model is as useful for predicting bankruptcy in recent periods as it was for the periods in which it was developed and tested by Altman. The assumption in using the model with the same coefficient estimates in a hold-out sample is that the model is robust across (macro) economic conditions that change over time, such as inflation, interest rates, and credit availability (see Nam et al., 2008, Tinoco and Wilson, 2013) and predictor variables must remain stationary. The effect of changing economic factors on the accuracy, magnitude, and significance of model coefficients was evaluated by Mensah (1984), Begley, Ming and Watts (1996) and Grice and Ingram (2001) and all of them reported that the accuracy and structure of the models changed over time.

In practice, researchers report that these assumptions of discriminant analysis are often violated (e.g. Altman, 1968, Altman and McGouch, 1974, Ohlson, 1980, Mensah, 1984). It should be noted that the key is, however, to build an empirically sound model, despite relatively minor theoretical objections. Indeed, much research (e.g. Hillegeist et al, 2004, Wu, Gaunt and Gray, 2010) uses discriminant analysis as a benchmark model, regardless of whether the assumptions are met. Sometimes, these models generally do indeed yield good results (for recent examples, see e.g. Agarwal and Taffler, 2007, 2008) despite the violations of the aforementioned assumptions.

2.1.1.3 Model empirics

In this section I will briefly go through the empirical research that utilizes the MDA-methodology and will look for implications for further research. I will first treat articles that contain *static* tests of the MDA-methodology. This means that the emphasis of this research lies on model performance at one point in time, and not in the model performance over time. Afterwards, I will treat *dynamic* tests, i.e., research explicitly focusing on model performance over time.

Static tests: Discriminant analysis was developed as a multivariate improvement of assessing the quality of ratio analysis in the prediction of bankruptcy as executed in Beaver (1966). Beaver (1966) investigated industrial firms from the Moody's Industrial Manual database. He compared means between bankrupt and non-bankrupt firms, subjected the findings to likelihood-tests and employed a simple classification test based on the likelihood of a firm belonging into either a bankrupt or non-bankrupt group. Beaver (1966) was able to accurately classify 78% of his sample of firms five years before failure based on ratio analysis.

The study of Altman (1968) features the first application of discriminant analysis to bankruptcy prediction. He investigates a U.S. manufacturing sample of 100 bankrupt and 100 non-bankrupt firms and reports that the discriminant analysis model is extremely accurate in classifying 95 per cent of the total sample correctly. In two years before the occurrence of bankruptcy, 72 per cent correct assignment is evidence that bankruptcy can still be predicted two years prior to the event. The errors do increase when the period to bankruptcy increases, so after two years the model cannot significantly distinguish bankrupt from non-bankrupt firms anymore.

Deakin (1972) emulates the research methodology by Altman (1968). He selected 11 failed and 23 non-failed firms at random from the Moody's Industrial manual. He indicated that misclassification errors averaged 3%, 4.5%, and 4.5% for the first, second and third years respectively. He also noted that incorrect classification as failed (type II errors) happens more often than incorrect classification as non-failed (type I errors). The accuracy of the discriminant analysis model declines when applied 4 or 5 years before the occurrence. This is in line with the results of Altman (1968), although Deakin's (1972) results show slightly greater robustness over time.

Begley et al. (1996) applied discriminant analysis and logistic regression (see 2.1.2) models to prediction bankruptcy for 1365 industrial firms, which is a broader application than used by Altman, who conducted his research on a sample of manufacturing firms. They used data from the S&P COMPUSTAT database. The models performed relatively well when they were estimated, but they find that they do not perform as well in more recent periods, even after re-estimation of the coefficients. The original logistic regression model displays the strongest overall performance. These results are important because they suggest other variables might predict bankruptcy better than the ones chosen by Altman (1968).

Pompe and Bilderbeek (2005) applied discriminant analysis and neural networks (see section 2.2) to the problem of bankruptcy prediction. Both methods produced similar results. They use factor analysis to determine the ratios included in the prediction models, and evaluate by means of classification tables (see section 2.3). Their sample included Belgian SME's from 1986 until 1994. Their data comes from the Belgian National Bank. The results showed that almost every ratio they had considered showed predictive power. Thus, an approaching bankruptcy is noticeable in almost every dimension of a firm's financial position. A few ratios, such as cash flow/total debt, achieved results that were close to the results of complete sets of other ratios. They also investigated the stationarity of their selected variables (i.e. constant mean, variance and covariance over time); it was found that the predictive power of the models were not sensitive to nonstationary variables.

Dynamic tests: Mensah (1984) used discriminant analysis for bankruptcy prediction and investigated the effect of change (macro)economic factors on the accuracy of the model coefficients by developing four models using samples from the 1972–1973, 1974–1975, 1976–1977, and 1978–1980 periods, where each period represents a different economic environment. His research features observations from the Wall Street Journal Index from 1972-1980. He found that the accuracy of the models substantially changed and that recalibration of the model coefficients accordingly leads to superior accuracy. This implies that the model should be re-estimated in every period of time and that the model is generally not robust to changes in the macroeconomic environment. Thus, these results contradict Deakin's (1972) and Altman's (1968) earlier results of a significant model robustness and stable coefficients over time. They imply that such research designs introduce a bias.

Mossman et al. (1998) investigate a sample of U.S. non-financial firms. Their accounting-based data ranges from 1980-1991 and comes from Compustat and their market data from CRSP. They test four sets of predictor variables: so-called ratio models, cash flow models, return models and return variation models. These ratios then serves as input in either a logistic regression based on Ohlson (1980) or a MDA model based on Altman (1968) Based on a sample of bankruptcies from 1980 to 1991, results indicate that no existing model of bankruptcy captures the data completely: "the discriminatory ability of the cash flow model remains relatively consistent over the last two to three fiscal years before bankruptcy, while the ratio model offers the best discriminatory ability in the year immediately prior to bankruptcy."

Grice and Ingram (2001) use a U.S. sample from the S&P COMPUSTAT database to find bankrupt firms and select a random sample of non-bankrupt firms in order to examine the generalizability of the Altman (1968) model in a variety of ways: whether the Altman original model is as useful for predicting bankruptcy in recent periods as it was for the periods in which it was developed and tested , whether the model is as useful for predicting bankruptcy of non-manufacturing firms as it is for predicting bankruptcy of manufacturing firms, used in the original Altman sample, and whether the model is as useful for predicting financial stress conditions other than bankruptcy as it is for predicting bankruptcy. The researchers find that the model performs worse in recent periods and that the model performs worse in a non-manufacturing sample, but it performs equally well when predicting distress rather than bankruptcy.

Agarwal and Taffler (2007) investigate bankruptcy for UK manufacturing firms. They conduct a test of the Z-score model based on the methodology of Altman (1968) but use estimation coefficients estimated by Taffler (1983) on a previous sample of UK manufacturing firms. They test the predictive ability of the model via hazard models (see section 2.1.4). The model has clear predictive ability over 25 years of bankruptcy data and dominates more naïve prediction approaches (a profit-before-tax indictor of bankruptcy). They also test the predictive of the models via ROC-characteristics (see section 2.2). They conclude that the z-score approach is not inferior in performance to market- based approaches. It is interesting to note that according to these results, discriminant analysis models can be robust and continue to maintain accuracy over many years. The results show high accuracy and high information content during a period ranging from 1979 to 2003.

2.1.2 Nonlinear probability models

In this section, logistic regression and probit estimation approaches are treated.[1] As a reaction on, inter alia, Altman (1968), researchers noted several econometric issues regarding the sampling approach and the degree to which the assumptions were met in the discriminant analysis. In particular, discriminant analysis provides a linear estimation approach to the problem of bankruptcy, while a linear change in a particular predictor ratio might not lead to a linear change in bankruptcy risk. Martin (1977), Ohlson (1980) and Zmijewski (1984) were the principal architects of a new, less restrictive model used to predict bankruptcy, utilizing maximum-likelihood optimization of a logit function (in the case of Martin, 1977 and Ohlson, 1980) or a probit function (in the case of Zmijewski, 1984) in order to estimate the coefficients of predictor variables, arguing for a nonlinear relationship between several ratio's and the probability of bankruptcy.

2.1.2.1 Logistic regression model description

The logistic regression model involves optimization of the logarithm of the likelihood of any specific outcome and reflects the binary sample space of bankruptcy versus non-bankruptcy. This method is appropriate for estimating a probabilistic outcome in a dichotomous variable (Hair et al., 2006), in this case, whether a firm is bankrupt or not. The independent variables used to estimate this outcome are in the case that we're concerned with, financial ratios. The coefficients are estimated by the method of maximum likelihood-optimization. The mathematical specification of the model can be found in appendix B.

All this optimization needs is a probability distribution function that fits the problem. Logistic regression makes use of the logistic function (Hair et al., 2006):

$$P = \frac{1}{1+e^{-\sum_j \beta_j X_{ij}}} \qquad \text{Eq. (2.2)}$$

This then yields coefficient estimates and a probability of bankruptcy for each firm *i.*

2.1.2.2 Probit estimation model description

Zmijewski (1984) follows a very similar approach to Ohlson (1980), that differs only in that the probability density function to be optimized is not the logistic function, but the probit function (the cumulative standard normal distribution). Eq. (3) is the maximum-likelihood optimizing function, but instead of using the Logit function as Ohlson (1980) does, Zmijewski (1984) employs the very similar cumulative normal distribution function:

$$P = \Phi(\beta_j X_{ij}) \qquad \text{Eq. (2.3)}$$

[1] Apart from the MDA, logit and probit models, several alternative statistical approaches, such as quadratic discriminant analysis (Altman et al., 1977) and mixed logit models (Jones and Hensher, 2004) have been used. These are treated in section 2.2.3. About these different methodological approaches to bankruptcy prediction, Agarwal and Taffler (2007) note that "since the results generally do not differ from the conventional linear discriminant model approach in terms of accuracy, or may even be inferior, and the classical linear discriminant approach is quite robust in practice, (…) methodological considerations are of little importance to users."

Where Φ represents the cumulative normal distribution function, β represents a vector of parameters to be estimated and X_{ij} represents a vector of j independent variables with i observations.

2.1.2.3 Model assumptions

Zmijewksi (1984) examines if there is an estimation bias when financial distress models are used in nonrandom samples, i.e. when a researcher first observes the dependent variable and then draws the sample based on that knowledge. This approach violates random sampling design and causes both parameter and probability estimates to be asymptotically biased. He argues that there are two different forms of bias in this case; “choice-based sample bias” and “sample selection bias”. The former comes from firms which are used (too) much in samples while the latter results from data which are incomplete, i.e. when firms try to hide information about their distress from the public. In other words, the analysis wrongly assumes that the bankruptcy to non-bankruptcy frequency in the sample is equal to that in the population because the assumption of random sampling is violated.

In order to look at this issue in a bit more detail, Zmijewski (1984) provides proof of biased parameters when the likelihood of a firm entering the sample is dependent on variable attributes, and shows that this bias decreases as the likelihood of bankruptcy in the sample approaches that of the population. He argues that the sampling practices utilized by Beaver (1966), Altman (1968), Deakin (1977), among others violate the exogenous random sample assumption of most estimation techniques, and alternative or adjusted techniques are needed to estimate unbiased parameters. The solution is then to weigh the log-likelihood function to be maximized in eq. (3) by the ratio of the population bankruptcy frequency rate to the sample bankruptcy frequency rate of the individual groups. This yields the following likelihood specification (Zmijewski, 1984):

$$L* = \frac{[POP]}{[SAMP]} \sum_{i \in S_1} \log P(X_{ij}, \beta_j) + \frac{[1-POP]}{[1-SAMP]} \log(1 - P(X_{ij}, \beta_j)) \qquad \text{Eq. (2.4)}$$

Where the probability as a function of parameter estimates β_j multiplied by X_{ij} = a vector of predictors, is modeled by the cumulative distribution function. This modification is done because weighting the log-likelihood function adjusts the parameters for the choice-based sample. As the sample selection probability approaches the population probability, the likelihood function approaches the "unweighted" probit likelihood function. In other words, studies using sample selection frequencies close to the population frequency of default, such as Ohlson (1980), and Zmijewski (1984) are not as affected by the aforementioned bias as are studies using 50% sample frequency rates and an unweighted estimation such as by Beaver (1966), Altman (1968) and Deakin (1977), among others.

Furthermore, unlike OLS, logistic regression and probit estimation do not assume linearity in the relationship between the values of the independent variables and values of the dependent, it does not require normally distributed variables, does not assume homoskedasticity, and in general have less rigorous requirements than MDA (see Hair et al, 2006). It does require that observations are independent and it does assume that the independent variables are linearly related to the logit of the dependent variable.

Based on Collins and Green (1982), as well as the methodological explanations by Ohlson (1980) and Hair et al. (2006) the following requirements need to be met:

- The true conditional probabilities are a logistic function (logistic regression) or the cumulative normal distribution (probit regression) of the independent variables.
- No important variables are omitted.
- No extraneous variables are included.
- The independent variables are measured without error.
- The observations are independent.
- The independent variables are not linear combinations of each other (no multicollinearity).

Note that these assumptions for logistic and probit regression are the same, because their estimation procedures are identical, apart from the probabilistic modeling functions.

Mensah (1984) further points out that users of accounting-based models must recognize that such models may require recalibration from time to time to take into account changes in the economic environment to which they are being applied – predictor variables might not exhibit stationarity. This is probably the most restrictive assumption, which is resolved in a model by Shumway (2001), treated in section 2.1.4. Grice and Dugan (2003) specifically test this issue with the models of Ohlson (1980) and Zmijewski (1984) and confirm that performance over time indeed declines and that coefficients better be re-estimated. They indicate that the models' accuracy is significantly lower in recent periods. These results improve when the models are re-estimated, but the magnitude and significance of the re-estimated coefficients differ from those reported in their original application. Hair et al. (2006) note that sample size considerations are primarily focused on the size of each group, which should have an amount of observations at least equal to approx. 10 times the number of estimated model coefficients as a rule of thumb.

2.1.2.4 Model empirics

As in the previous empirics section, I will briefly go through the empirical research that utilizes the logistic regression or probit analysis-methodology and will look for implications for further research. I will first treat articles that contain *static* tests (model performance at one point in time) and then articles focusing on *dynamic* tests (research explicitly focusing on model performance over time).

Static tests: A study by Martin (1977) introduced logistic regression to predict probability of failure of banks based on the data obtained from a U.S. Federal Reserve System database for research on bank surveillance programs. In the relevant sample period (1974-1976), 23 banks were reported as bankrupt and 5575 banks served as the non-bankrupt addition. Using different predictors, he finds several accuracy rates of around 91%. Ohlson (1980) employed the logit model to predict industrial firm failure in the U.S. He attempts to distinguish 105 bankrupt firms from 2058 non-bankrupt firms. The classification accuracy reported by him was 96.12%, 95.55% and 92.84% for prediction within one year, two years and one or two years respectively. Unlike the approach of Altman (1968), Ohlson (1980) follows no explicit matching procedure in sampling. “It is by no means obvious what is really gained or lost by different matching procedures, including no matching at all. At the very least, it would seem to be more fruitful actually to include variables as predictors rather than to use them for matching purposes." Zmijewski (1984) examined two estimation biases for financial distress models on non-random samples. The research is conducted on 40 bankrupt and 800 non-bankrupt industrial firms from the 1972 to 1978 time period. The choice-based sample was examined using unweighted probit and weighted exogenous sample maximum likelihood. In the best case, he finds an accuracy of 97.1%, that is, 97.1% of the bankruptcy statuses was predicted correctly.

Platt and Platt (1991) tested the model used by Ohslon (1980), while employing different, industry-relative ratios as predictors. They used 114 equally matched bankrupt and non-bankrupt companies for the period of 1972 until 1986. They gathered industry and firm ratio's from the Compustat database. They found that using these industry-relative ratios improved the accuracy of the model: their models could predict 80-86% of all firms correctly. Westgaard and van der Wijst (2001) assess default probabilities in a corporate bank portfolio. Their analysis is based on a logistic regression model where financial variables as well as other firm characteristics affect the default probability. They use a logistic regression analysis based on Ohlson (1980) and find that the financial ratios are negative and significantly so at least at a 5% level of significance. The same is true for firm size and firm age. The results indicate a strong relationship between bankruptcy and the variables used in the model and a reasonably good fit of the model to the data.

Lee and Yeh (2004) investigate whether corporate financial distress is related to corporate governance characteristics in a sample of Taiwanese listed firms that encountered financial distress between 1996 and 1999, and they pair-matched these companies with healthy firms. Logistic regression models in approximately the same way as Ohlson (1980) were employed, using several corporate governance measures in addition to some accounting variables to predict financial distress. In all cases, they could classify over 85% of the observations correctly. The evidence suggests that firms with weak corporate governance are vulnerable to negative economic trends, and the probability of falling into financial distress increases.

Astebro and Winter (2012) use a methodology based on Ohlson (1980), although somewhat more sophisticated. They get their data from the Compustat database, covering firms listed on the NYSE, AMEX, and NASDAQ exchanges from 1980-1988. Rather than using a binary logistic regression (i.e. predicting either bankruptcy or non-bankruptcy), they model the outcome as failure, survival as going concern, and acquisition based on accounting ratios. Their results show that accounting variables are significant predictors. The multinomial model performs superior to the binary model, and industry specific dummies add explanatory power to the model.

Dynamic tests: Grice and Dugan (2003) research the accuracy of the Ohlson (1980) and Zmijewski (1984) models under different circumstances than when the models were originally estimated in order to test whether "the models' predictive powers transcend several time periods, industries, and financial conditions outside of those used to originally develop the models." They used data from Compustat's Industrial Annual Research file (CIAR) from

1985 to 1987 in their estimation sample and 1988 to 1991 in their hold-out sample. Their findings indicated that the accuracy of the models increased when the coefficients were re-estimated. The relation between financial ratios and financial distress thus changes over time. The relative importance of the ratios differed over time, i.e. the ratios showed different levels of statistical (in)significance.

Canbas et al. (2005) develop an integrated early warning system for the prediction of bank failure. Principal component analysis was used to explore the basic financial characteristics of the banks, and discriminant, logit and probit models were estimated based on these characteristics to construct a comprehensive methodological framework. Their sample contains 40 privately owned Turkish commercial banks (of which 21 banks failed) during the period 1997–2003) and their financial ratios. The results show that one year prior to bankruptcy, discriminant analysis has a correct classification rate of 90%, logistic regression of 87.5% and the probit model of 87.5%. As in other studies (e.g. Altman 1968, Ohlson 1980, Mensah 1984 and many others) the accuracy declines when ratios are used two or three years prior to bankruptcy.

2.1.3 Distance-to-default probability model

Starting with research by Hillegeist et al. (2004), Vassalou and Xing (2004), Shumway & Bharath (2004), among others, bankruptcy prediction models have been developed that exclusively incorporate data obtained from capital markets. The model that is arguably the most prominent is based on Merton (1974), who developed a distance-to-default forecast model based on capital market data, thus allowing for continuously re-estimated forecasts because of the incorporation of continuously changing capital market data. Relying on the option pricing analysis of Black and Scholes (1973), Merton (1974) proposes considering the value of the equity as a call option on the value of assets of the firm with a strike price equal to the face value of the firm's debt (Trujillo-Ponce et al., 2013).
From this perspective, the company will default if its asset value falls below a certain default boundary related to the company's outstanding debt.

The contingent claims-based model overcomes many of the fundamental shortcomings of accounting-based models. Assuming market efficiency (Fama, 1998), stock prices reflect all historical financial as well as the individual and market-wide perceptions of a firm. Second, market prices are less likely to be influenced by accounting policies. Third, market prices are less likely to contain manipulations and distortions concerning the firm's financial health.

2.1.3.1 Model description

Under the Black and Scholes (1973) and Merton (1974) approaches, debt and equity are considered as a derivative on the assets of a given firm: the value of the firm's equity (V_E) representing the value of the call option on the firm's assets (V_A) with a strike price (D) being the market value of the firm's debt. Mathematically, it yields the following specification (see for instance Hillegeist et al., 2004, Vassalou & Xing, 2004 for a derivation):

$$V_E = V_A N(d_1) - De^{-rT} N(d_2) \qquad \text{Eq. (2.5)}$$

Where $d_1 = \frac{\ln\left(\frac{V_A}{X}\right) + (r + \frac{\sigma_A^2}{2}))\ T}{\sigma_A^2 \sqrt{T}}$ and $d_2 = d_1 - \sigma_A \sqrt{T}$. Eq. (2.6, 2.7)

The risk-free rate is denoted by r, N is the cumulative density function of the standard normal distribution, σ_A is the volatility of the market value of the firm's assets and T denotes the time to maturity.

The market value and the volatility of the firm's assets are not observable. The market value and the volatility of assets that are implied by the value of equity are estimated by simultaneously solving the equation given in eq. 2.5 and the following equation where $\sigma_{E,it}$ is the volatility of the equity:

$$\sigma_{E,it} = \frac{V_{A,it} N(d_1) \sigma_{A,it}}{V_{E,it}} \qquad \text{Eq. (2.8)}$$

So under the model's assumptions, both unknowns can be inferred from the value of equity, the volatility of equity, and several other observable variables by using an iterative procedure to solve a system of nonlinear equations (Shumway and Bharath, 2008) so that the two unknowns can be solved for.

The model can then be specified in a way that the probability of default of a given firm at time *t* is the normal cumulative density function of a z-score depending on the firm's underlying value, the firm's volatility, and the face value of the firm's debt (Shumway and Bharath, 2008):

$$p_t(T) = N\left[-\frac{\ln\frac{V_A}{D} + \left(\mu - \frac{\sigma_A^2}{2}\right)(T)}{\sigma_A \sqrt{T}}\right] \qquad \text{Eq. (2.9)}$$

The approach of Shumway and Bharath (2008) is then to first estimate σ_E from either historical stock returns data (i.e. standard deviation of returns) or from option-implied volatility data. Second, a forecasting horizon and a measure of the face value of the firm's debt are chosen. They note that it is common to use historical returns data to estimate σ_E and to assume a forecasting horizon of 1 year (T=1), and take the book value of the firm's total liabilities to be the face value of the firm's debt. The third step is to collect values of the risk-free rate and the market equity of the firm. After obtaining data on these variables, all variables in equations (2.5) and (2.8) are present, except for V_A and σ_A, the total value of the firm's assets and the volatility of firm value.

Hillegeist et al. (2004) follow an approach where they first simultaneously estimate V_A and σ_A, and afterwards use these values to estimate the expected return on the firm's assets μ while Vassalou & Xing (2004), Shumway and Bharath (2004, 2008), Charitou et al. (2013) and Bauer and Agarwal, (2014) estimate this using the past returns. Hillegeist et al. (2004) note that "while Vassalou and Xing (2004) rely on the DD model as we do, they do not adjust for dividends, and their method for estimating μ will frequently result in negative expected growth rates, a result that is inconsistent with asset pricing theory."

Trujillo-Ponce et al. (2013) follow a somewhat different estimation procedure by assuming that the market value of equity is considered to be the firm's market capitalization during the month of December. The input σ_E (the annualized standard deviation of equity returns) is estimated from the prior year of stock price returns. They consider a 1-year period to maturity of the debt. The assumption is that the amount of debt or the default point is equal to the book value of the current liabilities plus half of the long-term debt. Similar assumptions and definitions are serving as inputs to the models in the studies of Vassalou and Xing (2004) and Du and Suo (2007), Shumway and Bharath (2004, 2008) among others: σ_E is the annualized percent standard deviation of returns and is estimated from the prior year stock return data for each month. For *r*, the risk-free rate, Shumway and Bharath (2008) use the 1-year Treasury Constant Maturity Rate obtained from the Board of Governors of the Federal Reserve System. Other research uses similar appropriate proxies in accordance with their sample. V_E , the market value of each firm's equity is calculated as the product of share price at the end of the month and the number of shares outstanding.

Altogether, this model represents the probability that the value of the company is less than the value of the debt of the company on date *T*. The approaches outlined here are used by Moody's KMV[2], by Shumway and Bharath (2004, 2008), Hillegeist et al. (2004), Vassalou and Xing (2004), Demirovic and Thomas (2007), Du and Suo (2007), Trujillo-Ponce et al. (2013) and is also tested against other models in Duffie, Saita and Wang (2007) and Bauer and Agarwal (2014)

Shumway and Bharath (2004, 2008) as well as Bauer and Agarwal (2014) employ a simpler version of the model previously outlined, to test whether the procedures of estimating the two unknown parameters V_A and σ_A adds value when compared to reduced-form models. The researchers construct a simple alternative probability that does not require solving equations (3) and (4) by implementing an iterative procedure used in for example Hillegeist et al. (2004) and Vassalou and Xing (2004). They construct the 'naïve' model with two objectives. First, to test whether the naïve model captures the same information as the more complex Merton DD predictor does. Second, because it is simple and does not involve elaborate estimation procedures.

In the naïve DD model, the market value of each firm's debt is approximated with a half times the current liabilities plus its long-term debt:

$$naive\ D = \frac{1}{2} * C + L \qquad \text{Eq. (2.10)}$$

The risk of a firm's debt is correlated with their equity risk, so the volatility of a firm's debt is approximated as: $naive\ \sigma_D = 0.05 + 0.025 * \ \sigma_E$ Eq. (2.11)

The five percentage points in this term represent term structure volatility, and the 25% times equity volatility is included to allow for volatility associated with default risk. The approximation of the total volatility of the firm is then:

[2] A world-renowned credit rating agency

$$naive\ \sigma_A = \frac{E}{E+naive\ D}\ \sigma_E + \frac{naive\ D}{E+naive\ D} naive\ \sigma_D \qquad \text{Eq. (2.12)}$$

They set the expected return on the firm's assets is equal to the firm's stock return over the previous year: $naive\ \mu = r_{it-1}$ Eq. (2.13)

This captures some of the same information that is captured by the model utilized by Hillegeist et al. (2004) and Vassalou and Xing (2004). The iterative procedure is able to condition on an entire year of equity return data. Shumway and Bharath (2008) note that "by allowing our naïve estimate of μ to depend on past returns, we incorporate the same information. The naïve alternative model is easy to compute; however, it retains the structure of the Merton DD distance to default and expected default frequency."

It also captures approximately the same quantity of information as the DD probability. By incorporating this model alongside the complex variant of the DD model, the accuracy of this model can be determined. It can be assessed whether the iterative procedure utilized by Shumway and Bharath (2004, 2008), by Hillegeist et al. (2004), Vassalou and Xing (2004) and Charitou et al. (2013) is worth the effort or whether the predictive ability of the model is due to its functional specification. We define the naïve probability estimate as (Shumway and Bharath, 2008)**:**

$$p_{naive} = N\left[-\frac{\ln\frac{E+naive\ D}{naive\ D} + \left(r_{it-1} - \frac{naive\ \sigma_V^2}{2}\right)(T)}{naive\ \sigma_V\sqrt{T}}\right] \qquad \text{Eq. (2.14)}$$

2.1.3.2 Model assumptions

It should be noted that Merton's (1974) model presents certain unrealistic assumptions. It considers the debt-structure of the firm as only one homogenous zero-coupon bond and holds that the bankruptcy cannot be triggered before the debt will mature In addition, this model presumes that the absolute priority rule always holds, i.e., equity holders can only get a payout after debt maturity (Trujilo-Ponce et al., 2013). But arguably the foremost assumption of the DD model in its forms treated here is that it is assumed that the market value of debt can be approximated by the face value of debt when estimating firm asset volatility. In reality however, the use of face value of debt may underestimate true asset volatility and firms that are riskier (or higher leveraged) may be more affected than healthy firms, as volatility underestimation increases with leverage. From this it follows that the model's ability to separate defaulting from non-defaulting firms may be reduced.

In case of the Shumway and Bharath (2008) methodology, the volatility of assets measure assumes that bond volatility is non-zero and increases with equity volatility, so changing debt volatility may be an improvement.

In case of the design based on Hillegeist et al. (2004), Charitou et al. (2013) note that "asset volatility estimation may be affected by capital structure changes as it is based on changes in the total value of equity and debt rather than on returns on equity and debt. Hence, capital raising through equity or debt, stock repurchases or debt repayments may result in misestimating asset volatility.” From this follows the assumption of constant capital structure. The design based on Shumway and Bharath (2008) is based on past stock returns and is therefore not affected by such changes in capital structure (Charitou et al., 2013).

Furthermore, Allen and Saunders (2002) point out that DD models are unable to differentiate between the different durations of debt since they assume a zero-coupon bond for all liabilities. Also, Avramov et al. (2010) argue that distressed firms are prone to suffer from “market microstructure problems such as thin trading or limitations to short-selling which might result in prices deviating from fair values for extended period.” Another key assumption is that some variables required for these models (asset volatility, expected asset returns, and market value of assets) are unobservable and need to be approximated introducing potentially large errors.

The model is also exclusively based on market-data and does not allow for the input of accounting ratios, whereas other models treated in this study do. The arguments for incorporating accounting ratios in contingent claims-based bankruptcy prediction frameworks are a trend in literature are put forward by Sloan (1996), who finds that market prices do not accurately reflect the information from company accounts, so that accounting data can be used to complement market data.

2.1.3.3 Model empirics

In this section, the empirical research examining the performance (both accuracy and information content) of DD models will be covered. In contrast to the two previous empirics sections, the DD model, due to its unique nature, does not feature a distinction between static and dynamic tests: the model does not feature fixed (static) coefficients.

Hillegeist et al. (2004) compare the relative information content (see section 2.2) of the Altman (1968) and Ohlson (1980) models to a market-based measure of the probability of bankruptcy that is based on the Black–Scholes–Merton option-pricing model by means of a hazard model. They use the Moody’s Default Risk Services’ Corporate Default database and the SDC Platinum corporate restructuring database from defaulted firms, and all industrial firms that have the CRSP and Compustat data available to compute the three models. Their sample period ranges from 1980 to 2000. Their sample consists of 14,303 firms. They find that the market-based approach carries superior information to the Altman (1968) and Ohlson (1980) models and is robust to various modifications, including re-estimating the coefficients, making industry adjustments and lagging the respective scores.

Vassalou and Xing (2004) use the DD model to compute a default measure for individual firms and assess the effect of default risk on equity returns. They use the Compustat database and the Federal Reserve Board statistics to obtain all relevant data. Their sample period ranges from 1971 to 1999. They find that default risk is intimately related to the size (SMB) and book-to-market (HML) characteristics of a firm as specified in the Fama-French (1993) three factor model. The results point to the conclusion that both the size (SMB) and growth (HML) effects can be viewed as default effects and that default is a variable worth considering in asset-pricing models, above and beyond size and book-to-market ratio. In addition, they provide new ways to evaluate the models by introducing the area under an ROC-curve (see section 2.3) as proxy for the predictive accuracy of a model.

Demirovic and Thomas (2007) assess whether markets effectively reflect credit risk and test the DD model against information from accounting ratios. The sample comprises all those UK-listed companies that are rated by the three major ratings agencies (Moody, S&P, and Fitch) during the period 1990–2002. They obtain credit ratings via the Bloomberg database, and all other relevant data from the Datastream database. They find that the DD model is the most significant variable in explaining credit risk ratings, but that accounting variables are informative and display statistical significance when added to a model that contains only the distance-to-default measure. This addition however, is contingent on industry and firm size.

Reisz and Perlich (2007) compare their own custom DD model in estimating the probability of default and compare it to a basic version, based on Hillegeist et al. (2004), as well as the Altman (1968) z-score using the ROC curve (see section 2.2). They gather data for 5784 industrial firms in the period 1988–2002 from the Compustat database and show that for a one-year prediction horizon their complex framework outperforms the Merton (1974) framework but, in contrast to Hillegeist et al. (2004), underperforms the z-score model.

Bharath and Shumway (2008) construct a default forecasting model without considering the iterated procedure in computing Merton (1974) DD model used by e.g. Hillegeist et al. (2004). For their non-bankrupt companies, they use data for all non-financial firms that are entered in the Compustat Industrial file and the CRSP daily stock return for NYSE, AMEX, and NASDAQ stocks between 1980 and 2003. They obtain their data for bankrupt companies over the period 1980–2000 from the database of firm default maintained by Edward Altman (the Altman default database). They augment this data with data from 2001 to 2003 from the list of defaults published by Moody's. Altogether, they have a total of 1449 firm defaults covering the period 1980–2003. The naïve probability they use captures both the functional form and the same basic inputs of the DD model, and it performs surprisingly well. The naïve version of the model performs slightly better in hazard models and in out-of-sample forecasts than the full estimation of the model. Several other forecasting variables are also important predictors and show statistical significance when the full estimation of the DD model does not. Although the DD model does not produce a sufficient statistic for the probability of default (as control variables show significance too), its functional form is useful for forecasting defaults. However, hazard models that use the DD model with other covariates have slightly better out-of-sample performance than models that omit the DD metric. They conclude that the DD model is a useful variable for forecasting default, but not a sufficient predictor. Bauer and Agarwal (2014) refer to the good results obtained by Bharath and Shumway (2008) as a justification for incorporating the naïve model into their research.

Agarwal and Taffler (2008) compare market-based and accounting-based prediction models for all non-financial firms in the UK listed on the London Stock Exchange (LSE) at any time during the period 1985–2001. They collect their accounting data from various databases, the exchange of listing and firm stock exchange industrial classifications from the London Business School London Share Price Database (LSPD). The risk-free rates rates, market value of equity and daily stock prices from Datastream, and the list of firm failures from LSPD and minor other sources. They test the performance of the (accounting-based) z-score model against two versions of the (market-based) DD models as used by Shumway and Bharath (2004) and Hillegeist et al. (2004). They assess accuracy via the area under the ROC-Curve, use hazard models to assess information content and assess the economic value of the models. They find that in terms of predictive accuracy, there is little difference between the market-based and accounting-based models. In terms of economic value of each model, the accounting-based approach performs slightly better. Neither the market-based models nor the accounting-ratio-based model is a sufficient explanation for firm failure and both carry unique information about firm failure.

Dionysou et al. (2008) use the "naïve" DD model from Shumway and Bharath (2004, 2008) without solving the required nonlinear equations to assess the ability to forecast default. The data they used comes from 1269 U.S. industrial firms that filed for bankruptcy from 1983 to 2001 and have data available in the Compustat and CRSP databases. They check every bankrupt firm for actual bankruptcy in the Wall Street Journal or in the Internet Bankruptcy Library as having filed for bankruptcy. They augment the bankrupt sample by using 6564 available healthy firms from the same source, resulting to a total sample of 7833 firms. They control for the possibility of going bankruptcy before debt maturity by including a liquidity proxy in their model. The authors find that their model appears to be a valuable tool in forecasting default: it provides sufficient statistic and significant predictive ability. Hereby they corroborate Shumway and Bharath (2008) in that the predictive power of the DD model lies in its functional form.

Charitou et al. (2013) test several versions of the DD model on US firms over the period 1985‑2009. They gather firms in their sample that have financial and market data available in the Compustat and CRSP databases. They show that models using direct market-observable volatility estimate perform better than alternative, more complex models like the iterative-procedure DD model. They apply Cox proportional hazard models (see section 2.1.4) in order to test whether alternative specifications of the option variable estimations used in the literature lead to measures with improved ability to forecast default probabilities. "Our findings suggest the adoption of simpler modeling approaches relying on market data when implementing the DD model. (…) The results verify the main insight noted by Bharath and Shumway (2008) and extended by us in directly estimating volatility from market firm data, confirming that it is not necessary to solve the DD model's simultaneous equations to accurately predict default." The findings further indicate that models based using as many inputs as available generally improve the accuracy and information content.

2.1.4 Hazard models

A contemporaneous trend in the bankruptcy prediction literature is the use of survival analysis (e.g. Shumway, 2001), where, contrary to basic statistical models, the time-to-default of a firm is taken into account and there are more firm-year observations incorporated to explain the bankruptcy. Shumway (2001) denotes his model by the term "hazard model", but the term is used interchangeably with the terms panel logit model, pooled logit model, or Cox regression model with time-varying covariates and sometimes the general survival analysis is used (in fact, hazard models are just 1 particular branch of survival analysis models, see e.g. Hair et al. [2006]). It should be noted that these models have two applications within the bankruptcy prediction literature: the first application is that it serves as a model that directly estimates bankruptcy probabilities, while incorporating the dynamic characteristics of a firm. This is the application that will be focused on here. The second application concerns the evaluation of the information content of a particular bankruptcy prediction model. This is covered in section 2.3.4.

Lane, Looney, Wansley (1986), Shumway (2001), Chava & Jarrow (2004), Beaver et al. (2005), Campbell et al. (2008), as well as Nam et al. (2008) and Wu, Gaunt and Gray (2010) all incorporate hazard models in their research. Shumway (2001) argues that the models of Altman (1968), Ohlson (1980) and Zmijewski (1984) are biased and inconsistent because they only consider data of firms at one point in time and do not account for changes in time.

2.1.4.1 Model description

Arguably, the first paper applying a hazard model to bankruptcy prediction research was the paper of Lane, Looney and Wansley (1986). They use a Cox proportional hazard model, because "the omission of time to failure lessens the usefulness of MDA and its usual alternatives (logit, probit, and regression analyses)" and note that many assumptions of the parametric models are violated. In contrast with other research using hazard models, Lane, Looney and Wansley (1986) only incorporate accounting information as covariates.

The hazard function is specified as: $h(t|z) = e^{\beta * z} * h_0(t)$ Eq. (2.15)

Where β is a column vector of regression coefficients for *z* independent variables, h_0 is the baseline hazard varying with time and h(t|z) the probability that a firm will go bankrupt given the characteristics obtained by z and that the firm is alive at time *t*. Thus, the assumption underlying the Cox proportional hazard model is that the effect of the explanatory variables is to multiply the base hazard, $h_0(t)$, the basic probability of default over time, by some function, Ψ(z), of the individual firm characteristics z. In this case, Ψ(z) = e (β(z)) where *z* represents a vector consisting of predictor variables. Using an exponential function greatly simplifies the estimation of the regression coefficients according to Cox and Oakes (1984) and Hair et al. (2006).

In a basic hazard model, it is assumed that the values of the independent variables (accounting ratios) for a particular firm remain constant over the time interval that the object is in the study. The major contribution of the Cox model is that it incorporates additional information regarding the time to failure provided by the model. This additional information is captured in the estimated survivor function for a given firm. This practice is improved on by Shumway (2001) and several studies that utilize his model, that account for changing covariates over time.

Shumway (2001) argues that because bankruptcy occurs infrequently, forecasters should use samples that utilize data from several years in order to calibrate or estimate their models. Static models can only consider one set of explanatory variables for each firm, even though the characteristics of most firms change from year to year, so they don't take into account the change in any of the covariates. This criticism applies to the methodologies of Altman (1968), Ohlson (1980) and Zmijewski (1984) models, as well as the Cox regression approach of Lane, Looney and Wansley (1986). Most forecasters choose to observe each bankrupt firm's data in the year before bankruptcy. Data on healthy firms that eventually go bankrupt do not serve as input into the model. Because forecasters choose when to observe a firm's characteristics, static models introduce an unnecessary selection bias into their estimates (Shumway, 2001). The point of the Shumway (2001) study is then to construct a model that can include these time-varying predictor variables.

The Shumway (2001) model can be defined most easily as follows:

$$P = \frac{1}{1+e^{-y_{it}}} \qquad \text{Eq. (2.16)}$$

Where

$$y_{it} = \alpha + \beta' X_{i,t-1} = \beta' \begin{bmatrix} X_{1,t-1} & \cdots & X_{1,t-j} \\ \vdots & \ddots & \vdots \\ X_{n,t-1} & \cdots & X_{n,t-j} \end{bmatrix} \qquad \text{Eq. (2.17)}$$

where *P* is the probability of bankruptcy, *X* represents the variables listed and *a* represents the baseline hazard, i.e. the instantaneous risk of bankruptcy at time t given that the firm has survived up until t. This is in essence a pooled logit model, but instead of treating each firm-year as an independent observation, all prior values of the independent variables for a particular firm are included in the information set *n*, which represents the number of independent variables, and *j* represents the number of time periods prior to time *t* for which data are available. The coefficients are being estimated via partial log-likelihood optimization and can be treated as an ordinary log-likelihood optimization to derive valid estimators (Cox and Oakes, 1984, Hair et al., 2006).

Generally, Cox proportional hazard models with time-varying covariates are not presented in this form. However, Shumway (2001) provides an elaborate proof that the maximum-likelihood estimation of a formal Cox proportional hazard model is equal to the maximum-likelihood estimation of the model in equations 2.16 and 2.17.

Beaver et al. (2005) define the same model as:

$$\ln \left[\frac{h_j(t)}{1-h_j(t)}\right] = \alpha(t) + \beta X_j(t) \qquad \text{Eq. (2.18)}$$

where $h_j(t)$ represents the baseline hazard, or instantaneous risk of bankruptcy at time t for company j, given that the firm has survived up until t. $\alpha(t)$ is the baseline hazard and X is a matrix of time-varying observations for firm *j*. β is a vector of to be estimated coefficients. This makes for an easy interpretation: the log of the odds of bankruptcy at a given time *t* equals a linear function of a hazard rate α and a set of predictor variables *X*.

The main issue is to choose the correct explanatory variables, i.e. the ones that have a high explanatory power in forecasting firm bankruptcy. Shumway (2001) argues for combination of accounting ratios and market-driven variables that perform better than merely accounting or merely market variables used in prior studies. He concludes that a simple-hazard model gives indeed a better bankruptcy prediction of firms than a static model does with the same input (explanatory) variables. Campbell et al. (2008) suggest integrating accounting and market information even further by using ratios that contain both accounting and market variables. Hazard models can also incorporate macroeconomic variables that might influence bankruptcy (especially important in periods of crisis). Macro-economic variables are the same for all firms at a given point of time and hence cannot be incorporated into static models. Nam et al. (2008) and Christidis and Gregory (2010) make use of macroeconomic predictors in a hazard model.

2.1.4.2 Model assumptions

The hazard model by Shumway (2001) is identical to a Cox regression with time-varying covariates. This approach and derivative approaches assume that the data serving as inputs as the covariates X_{it} for all firms $i = 1,...,n$ and times $t = 1,...,T$ that the default times for every firm are independent of each other. This conditional independence is a standard assumption in hazard rate modeling.

Another assumption regards the censoring of observations. Censoring means that the data on the firm is observed at time *t* but not at time *t+1*. Time t is usually the last date in the sample period, but it could be otherwise. For example, the firm could experience a merger and vanish from the data set, or it could have no more data available. An assumption of Cox regression is that censoring is non-informative, i.e.in this case, that the causes giving rise to censoring of individual firms are not related to the probability of a firm going bankrupt.

Traditional Cox regression (Cox and Oakes, 1982) requires proportional hazards, i.e. a constant hazard rate dependent on a (set of) covariate(s) *X*. As Shumway (2001) shows, the problem of bankruptcy is not suitable for this assumption. Following Shumway (2001), allowing for time-varying covariates leads to unbiased coefficient estimates, merely to biased standard errors. For specific details, see Shumway (2001). A procedure is outlined in the methodology section to correct for this bias.

2.1.4.3 Model empirics

This section presents an overview of the empirical results obtained by hazard modeling in bankruptcy prediction. It feature single-method studies, but also studies that compared hazard models to other bankruptcy prediction models. As in section 2.1.3, this section features no distinction between dynamic and static models, because the models are now explicitly expected to take into account changes over time due to their econometric specification.

Lane, Looney and Wansley (1986) employ Cox proportional hazard models to forecast bank bankruptcy. They use a sample including all failed banks in the United States from January 1979 through June 1984 for which complete data were available from the Federal Reserve data subscription database. They find that the R-values for the Cox models are 0.449 for the one year prior data and 0.350 for the two years prior data, which indicates a “reasonably good" fit – meaning the ratios employed are useful in distinguishing between bankrupt and non-bankrupt banks. The Cox proportional hazard models are also compared to MDA in their performance. Results from the Cox model compare favorably with those from MDA, especially for the model using two years prior to bankruptcy-data. The superior performance of the Cox models over MDA does not hold under all circumstances, however.

Shumway (2001) introduces and tests a hazard model with time-varying covariates that uses all available information to determine each firm's bankruptcy risk at each point in time. His sample contains 300 bankruptcies in the U.S. between 1962 and 1992 obtained from various databases (Wall Street Journal Index, Compustat, among others). He finds that some of the ratios used by Altman (1968) and Ohlson (1980) are insignificantly related to bankruptcy risk and that several previously neglected market-driven variables are strongly related to bankruptcy probability, among them the firm's market capitalization, its past stock returns, and the idiosyncratic standard deviation of its stock returns. Combining market driven variables and accounting ratios, he constructs a model that is quite accurate in forecasting bankruptcy: it classifies three-quarters of bankrupt firms into the highest bankruptcy probability decile.

Chava and Jarrow (2004) test the hazard model specification of Shumway (2001) against the models of Altman (1968) and Zmijewski (1984). They use a variety of tests, at first only on industrial firms, and second also on financial firms. They use a variety of databases to collect bankruptcy data from companies listed on AMEX, NYSE or the NASDAQ exchanges. Their findings are that the hazard model has superior forecasting performance as opposed to the Altman (1968) and Zmijewski (1984) models. They also stress the importance of including industry effects in hazard rate estimation. They show that bankruptcy prediction is improved using shorter observation intervals, i.e. monthly observations rather than yearly observations. They demonstrate that accounting variables add little predictive power when market variables are already included in the model.

Beaver et al. (2005) employ hazard models based on the methodology by Shumway (2001) to forecast bankruptcy: they include the bankrupt and non-bankrupt firm data for years prior to the final year before bankruptcy occurs. Their sample consists of NYSE and AMEX-listed firms with data available in the Compustat database, from 1962 to 2002. The difference with Shumway (2001) lies in the fact that Shumway identifies and uses market-driven variables in the model that yields the best results, while Beaver et al. (2005) identify that accounting variables do have predictive power. The results show that the models retain strong predictive power over time, while the slight decline in the predictive ability of accounting-based variables is offset by raising predictive ability of market-related variables.

Duffie, Saita and Wang (2007) assess the probabilities of corporate default, incorporating both firm-specific and macroeconomic covariates. Their sample is confined to U.S. firms and observations are from 1980 until 2004 from the Compustat database. Their method allows them to combine survival analysis with time-varying covariates with conventional time-series analysis of covariates, in order to obtain multi-period survival probabilities. They find that the probability of bankruptcy for industrial firms depend significantly on the current state of the economy (i.e. macroeconomic variables have strong predictive ability), and also especially on the current leverage of the firm.

Campbell et al. (2008) use the hazard model methodology as developed by Shumway (2001) and very similar to that of Chava and Jarrow (2004). They use 1600 failures from the U.S. ranging from 1963 until 1998. Their data is obtained from various databases, including Compustat. They test several specifications. The model that obtains the best fit includes both market-based and accounting data. They also test the information content of the DD model. They find that this measure adds relatively little explanatory power.

Partington and Kim (2008) use Cox regression model with time-varying covariates to assess bankruptcy risk for a sample of Australian firms listed on the Australian Securities Exchange from 1989 to 2006.The results show that firms with higher book leverage, less cash flow generating ability and lower market leverage are more likely to fail. They note that their model has predictive power, but there is scope for considerable improvement.

Nam et al. (2008) use a similar methodology on a sample of listed companies on the Korea Stock Exchange. They compare three models: the existing static model based on Ohlson (1980) although with different covariates, a duration model that has time-varying covariate, and a model in which they include covariates that reflect the panel properties of financial statements as well as some macroeconomic factors. The results showed that dynamic models with time-varying covariates are more accurate than static models.

Wu, Gaunt and Gray (2010) compare five key models (Altman, Ohlson, Zmijewski, Shumway, Hillegeist, respectively utilizing MDA, logistic regression, probit estimation, a hazard model and a DD model) and evaluate their performance and build a new model comprising key variables from each of the five models. They obtain their data from New Generation Research, Compustat and CRSP. The sample contains NYSE- and AMEX-listed Compustat firms and covers the period from 1980 to 2006.They also add a new variable that proxies for the degree of diversification within the firm. They compare each of the original econometric specifications based on model-fit criteria such as the Receiver Operating Characteristics area (ROC area) and classification rates and evaluate their information content by means of a hazard model. They find that the MDA model of Altman (1968) performs poorly relative to the models of Ohlson (1980) and Zmijewski (1984). The model of Shumway (2001), which includes market data and firm-characteristics, outperforms models that are based on accounting information. The DD model proposed by Hillegeist et al. (2004) performs adequately, but the Shumway (2001) model performs better. The researchers construct a new hazard model based on the covariates with the highest informative content. That model outperforms all previous models. The researchers note that “this is consistent with the different types of data capturing different aspects of corporate financial distress.”

Christidis and Gregory (2010) employ the hazard model methodology of Shumway (2001). They use bankrupt publicly listed companies firms in the U.K. between 1978 and 2006 and non-bankrupt counterparts from the same source. They show that the incorporation of market variables of the form developed by Chava and Jarrow (2004) and Campbell et al (2008) add considerable predictive power compared with models based merely on accounting data. They also include macroeconomic and industry dummy variables. This adds predictive power, both within- and out-of-sample.

Tinoco and Wilson (2013) develop a custom model, based on a slightly modified logistic regression procedure that predicts financial distress and bankruptcy for listed companies. They use a sample of 23,218 firm-year observations of listed companies during the period 1980–2011 from various sources of data, not confined to any particular country or region. They use a combination of accounting data, market data and incorporate the inflation and interest rate changes as a proxy for macroeconomic developments. "It is tested whether market variables add information that is not contained in financial statements and therefore act as complement in default prediction models." The results show that this is true, i.e. that combining accounting, market and macro-economic data yields the best fit to the data. They note the increase of the area under the ROC curve among several other formal measures (see section 2.2), from 0.88 to 0.92 in a model estimated with data a one-year period from the event. They find statistically significant coefficients on a combination of market- and accounting variables as much as two years prior to the distress occurring and their findings remain robust through a large time period in a very wide sample.

Bauer and Agarwal (2014) conduct a comprehensive test comparing the performance of hazard models based on Shumway (2001) against the traditional accounting-based approach (e.g. Altman, 1968) or the contingent claims approach (e.g. Hillegeist et al., 2004). They analyze a sample of UK firms from 1979 to 2009 and evaluate the models on accuracy, information content and economic merit. The information content tests demonstrate that the hazard models subsume all bankruptcy related information in the accounting-based model as well as the DD model. Their other results show that hazard models based on Shumway (2001) yield the best results in all three evaluation criteria.

2.2 Intelligent and miscellaneous bankruptcy prediction models

Kumar and Ravi (2007) distinguish between several types of intelligent forecasting techniques. The most prominent technique used is neural networks. "In using neural networks, the entire available data set is usually randomly divided into a training (in-sample) set and a test (out-of-sample) set. The training set is used for neural network model building and the test set is used to evaluate the predictive capability of the model (Zhang et al., 1999)." I will provide a specification of the neural networks methodology and review the corresponding empirical evidence, as well as evidence obtained by using other intelligent bankruptcy prediction models.

As in any statistical model, the parameters (weights) of a neural network model need to be estimated before the network can be used for prediction purposes. The process of determining these weights is called training. Zhang et al. (1999) note that "patterns or examples are presented to the input layer of a network. (…) Finally an output value is obtained to match the desired value." The aim of a training set is to minimize the differences between the Neural Networks output values and the known target values for all training patterns.

The neural networks-based methodology found extensive applications in bankruptcy prediction. Kumar and Ravi (2007) note various types of neural networks, in particular the multi-layer perceptron, radial basis function network, probabilistic neural network, cascade correlation neural network, learning vector quantization, self-organizing feature map and others are among some of the popular neural network architectures.

2.2.1 Model description

Zhang et al. (1999) describe the mechanics of neural networks. The available data set is usually randomly divided into a training (in-sample) set and a test (out-of-sample) set. The training set is used for neural network model building and the test set is used to assess the predictive power of the model. They also present a mathematical definition of the neural networks methodology. They present a so-called three-layer perceptron form , containing one layer attaching weights from the input to the hidden layer, and another layer attaching the same weights from hidden to output layer.

$$y = f_2(w_2 f_1(w_1 \vec{x})) \qquad \text{Eq. (2.19)}$$

where $\vec{x}$=(x_1,x_2,…,x_n) is an n-dimensional vector of predictor variables, y is the output from the network, $w_1$and w_2 are the matrices of linking weights from input to hidden layer and from hidden to output layer.

f_l and f_2 are the transfer functions for hidden node and output node, respectively. According to Zhang et al. (1999), the most popular choice for f_l and f_2 is the logistic function:

$$f_1(x) = f_2(x) = (1 + e^{-\vec{x}})^{-1} \qquad \text{Eq. (2.19)}$$

The purpose of neural network training is to estimate the weights in eq. (2.18) such that an overall error measure such as the mean squared errors (MSE) or sum of squared errors (SSE) is minimized. MSE can be defined as

$$MSE = \frac{1}{N} \sum_{j=1}^{N} (a_j - y_j)^2 \qquad \text{Eq. (2.20)}$$

where a_j and y_j represent the target value and network output for the $_j$th training pattern respectively, and N is the number of training patterns. Thus, neural network training is an unconstrained nonlinear minimization problem. The various approaches used in the bankruptcy prediction literature differ mainly in their use of algorithms to solve this problem.

Atiya (2001) suggests a number of reasons why a nonlinear approach such as this, would be superior to a linear approach. For example, there might be saturation effects in the relationships between the financial ratios and the prediction of default. For example, if the earnings/total assets changes say by an amount of 0.2, from 0.3 to 0.1, it would have a far larger effect (on the probability of default) than it would if that ratio changes from, for example, 1.0 to 1.2. Atiya (2001) also argues that there are multiplicative factors as well. For example, the potential for default for a firm with negative cash flow gets more amplified if it has large liabilities, because firms with higher leverage find it harder to borrow money to finance their deficits.

Other approaches besides neural networks include case-based reasoning, which Kumar and Ravi (2007) note "is intuitively similar to the cognitive process humans follow in problem solving. When people confront a new problem, they often depend on past similar experiences and reuse or modify solutions of these experiences to generate a possible answer for the problem at hand." Case-based reasoning works because of its capability to give an explanation for its decision based on previous cases. The nearest neighbor algorithm is often used in combination with case-based reasoning approaches. Rough set theory also has important applications in bankruptcy forecasting. According to Kumar and Ravi (2007) it is based "on the assumption that with any object of the given universe there is some information associated and objects characterized by similar information are indistinguishable or indiscernible." Rough sets are applied for constructing and applying certain decision rules in order to solve classification problems.

Furthermore, support vector machines have also been used to forecast bankruptcy. It encompasses using a linear model to implement nonlinear class boundaries by mapping input vectors nonlinearly into a high-dimensional feature space. In the new space, an optimal separating hyperplane is constructed. The training examples that are closest to the maximum margin hyperplane are called support vectors (Kumar and Ravi, 2007). Decision trees have also been used to predict bankruptcy. Kumar and Ravi (2007) note that a majority of the decision tree algorithms can be used for solving this kind of classification problems. Some algorithms can be used for solving regression problems also. "All these algorithms induce a binary tree on the given training data, which in turn results in a set of 'if–then' rules". These rules are then used to solve a classification problem (Kumar and Ravi (2007)".

2.2.2 Empirical results of research using intelligent techniques

A brief overview is given of the most important results of intelligent techniques methodologies used in bankruptcy prediction – especially the research that features a comparison between intelligent techniques and statistical techniques is reported. For a more extensive review of intelligent techniques, see Kumar and Ravi (2007).

Frydman et al. (1985) used a recursive partitioning algorithm to forecast bankruptcy and compared it with the DA based on Altman (1968) and an economic analysis was conduct on which a judgment was based which model was better. They constructed two variants of discriminant functions. The recursive partitioning algorithm outperformed the discriminant analysis methodology in all respects. Tam and Kiang (1992) compared the performance in predicting bankruptcy of two statistical techniques versus other techniques based on neural networks. Back-propagation neural networks outperformed other techniques for one-year prior data, whereas for two-year prior data, the model of Altman (1968) yielded the best results. In out-of-sample forecasts, back-propagation neural networks yielded the best accuracy. Wilson and Sharda (1994) compared back-propagation neural networks with discriminant analysis, using the exact version of the Altman (1968) model. They concluded that the methodology based on neural networks outperformed other methods both in- and out-of-sample.

Lacher et al. (1995) conducted research on bankruptcy prediction using a cascade-correlation neural networks methodology for classifying financial health of a firm. The variables of Altman (1968) were used. The results of the model were compared with the results of the discriminant analysis methodology used by Altman (1968). They concluded that the Cascor model obtained higher overall classification rates in every scenario. Bell (1997) compared logistic regression based on Ohlson (1980) and back-propagation neural networks in predicting bank failures. The results were that the logit and the back-propagation model were about equally accuracte and neither model dominated the other in terms of predictive ability. Bryant (1997) used case-based reasoning for bankruptcy prediction. He compared it with Ohlson's (1980) logit model. He employed 25 financial variables and estimated the model for one year, two year and three years data. The results were ambiguous about which model was the most accurate, but showed that logit outperformed CBR in terms of less Type-I errors.

Dimitras et al. (1999) used rough set theory for predicting bankruptcy. They used the data collected from large number of Greek firms. The of rough set based approach was compared to the accuracy with the methodology of Altman (1968) and Ohlson (198) and the results showed that the rough set approach outperformed the other two methods. Ahn et al. (2000) constructed hybrid models based rough sets and BPNN for bankruptcy prediction in a sample of Korean firms and compared it with back-propagation neural networks (e.g. Bell, 1997) and discriminant analysis (Altman, 1968). They used classification tables to assess accuracy, and concluded that the hybrid models outperformed both discriminant analysis and back-propagation neural networks.

McKee (2003) compared rough set methodology with actual auditor signaling rates for US companies. The two models he developed (containing different variables) achieved classification accuracy of 61% and 68% on the validation set. Auditors achieved classification accuracy of 66%. Kumar and Ravi (2007) note that this is because "the samples employed here were more realistic than prior studies."

Charitou et al. (2004) developed failure prediction models via logistic regression and neural network (NN) methodology, and also to explore the incremental information content of cash flows in predicting the probability of bankruptcy. They find that Neural Networks achieved the highest overall classification rates for all three years prior to insolvency, with an average classification rate of 78%. The logistic model obtained an average correct classification rate of 76%, although it produced slightly lower type I error rates

Tsai (2009) compares five well-known feature selection methods used in bankruptcy prediction: t-test, correlation matrix, stepwise regression, principle component analysis (PCA) and factor analysis (FA) to select variables that serve as input in a multi-layer perceptron neural networks prediction model. Five related datasets are used to check for robustness. Surprisingly, the t-test feature selection method outperforms the other ones. This is assessed by means of classification tables

Lee and Choi (2013) employ back-propagation neural networks to conduct a multi-industry investigation of the bankruptcy of Korean companies. The industries include construction, retail, and manufacturing. The prediction accuracy of back-propagation neural networks is compared to that of multivariate discriminant analysis based on Altman (1968). Their results show that prediction using industry samples as input outperform the prediction using the entire sample by 6‑12%. The prediction accuracy of bankruptcy using neural networks is greater than that of MDA.

2.2.3 Empirical research using customized models

In order to assess the discriminating and classifying ability of a certain estimation procedure (e.g. MDA, logistic regression, or probit analysis), some benchmark is needed so that one can compare one method of estimation with another. In order to be able to make a comparison between more commonly used and less commonly used methods of estimation, research that uses derivative but very similar techniques to the techniques treated before, such as quadratic discriminant analysis, featured in Altman et al. (1977), will also be treated here. Research utilizing several methodologies with the primary emphasis on a custom model will also be treated in this section.

Altman et al. (1977) test both linear discriminant analysis and quadratic discriminant analysis in an estimation sample and a hold-out sample. The results show that “the quadratic and linear models yield essentially equal overall accuracy results for the original sample classifications, but the holdout sample tests indicate a clear superiority for the linear framework.”

Pindado et al. (2008) propose an approach to estimating the likelihood of financial distress that can be applied to different periods and countries. The model encompasses logit analysis specified for panel data (both fixed and random-effects logit models) in order to eliminate the unobservable heterogeneity. The results show that the model is stable in terms of coefficients and significance. The findings of this study show that, compared even with a re-estimated Z-score model, the new model performs much better. It has stable predictions over time and in different countries, and retains coefficient significance over time and in different countries.

Lyandres and Zhdanov (2013) assess bankruptcy by constructing and empirically testing models that includes a measure of investment opportunities. The results show that investment opportunities are significantly related to the likelihood of bankruptcy. Augmenting existing bankruptcy prediction models by measures of investment opportunities, in particular the market-to-book ratio, the ratio of estimated “true” equity value to its book value, and the ratio of R&D expenses to assets improves the forecasting ability for the models in an out-of-sample test.

Trujilo-Ponce et al. (2013) analyze whether accounting-based or market-based default prediction models better explain corporate credit risk through linear regression analysis. They find that these two approaches do not differ significantly as predictors of credit risk. A comprehensive model that combines accounting- and market-based variables is the best option to explain the credit risk, suggesting that both types of data are complementary. This is in line with Shumway (2001), Wu, Gaunt and Gray (2010) but seems to contradict the findings of Chava and Jarrow (2004). They also find that the explanatory power of credit risk models is particularly strong during periods of high uncertainty, i.e. in the recent financial crisis. Finally, the models show higher explanatory power if the credit rating is used as the proxy for credit.

All in all, the results seem to imply that at most, marginal increases in model accuracy can be realized through utilizing different, less commonly used estimation procedures.

2.3 Performance evaluation

According the framework of Agarwal and Taffler (2008) as well as Bauer and Agarwal (2014), there are several dimensions along which the accuracy of a bankruptcy prediction model can be evaluated. Accuracy is concerned with evaluating a model concerns the model's ability to discriminate between failures and non-failures. I have divided these into three different categories. These are (1) classification tables, (2) Receiver Operating Characteristics, (3) other goodness-of-fit measures. Model accuracy used to be assessed by means of classification tables (e.g. Altman, 1968, Zmijewski, 1984, Hair et al., 2006) but nowadays, the area under the ROC-curve is most often used (e.g. Charitou et al., 2013, Bauer and Agarwal, 2014).

Then, I show how information content is evaluated. Information content is concerned with whether bankruptcy prediction models carry information about actual bankruptcy in a statistically significant way. That is to say, does an incremental increase in a bankruptcy prediction score correlate well with an increase in actual bankruptcy risk, and does it so under all circumstances? The approach here is based on an econometric model by Shumway (2001), and only later used for this purpose by Chava and Jarrow (2004), followed by many others.

Finally, economic value added tests (used in this context first by Agarwal and Taffler, 2008) are more relevant for businesses and credit providers that actually use bankruptcy prediction models. They are used to construct scenarios in which a bank judges to offers loans to creditors with randomized characteristics (some of which will go bankrupt) by using a particular bankruptcy prediction model. Various performance indicators, such as return on assets or return on risk-weighted assets can then be compared to assess a prediction model's economic value.

2.3.1 Classification tables

Beaver (1966) evaluated his model by means of t-tests, to see whether there was a significant difference in financial ratios between a bankrupt and non-bankrupt firm. Due to its univariate nature, this cannot be applied to multivariate techniques like discriminant analysis or logistic regression. Altman (1968) was the first to introduce classification tables in the bankruptcy predicition literature. In general, discriminant analysis is still evaluated in this way (see Hair et al., 2006).

Classification tables show have two rows and two columns, indicating how many of bankrupt firms are classified by the model as bankrupt and as non-bankrupt (type I error), and how many non-bankrupt firms are classified by the model as bankrupt (type II error) and non-bankrupt. The performance is then assessed by choosing the cut-off point such that the amount of misclassifications (the sum of type I and type II errors) is minimized. This assumes implicitly that the cost of misclassifying a bankrupt firm is equal to misclassifying a non-bankrupt firm. A further evaluation of this issue will be featured in section 2.3.5. He ordered his observations estimated by the model from a small z-score to a large z-score and then chooses the optimal cut-off point as the point where the total number of misclassifications was minimized (i.e. assuming that the type I and type II error are symmetric). This approach, or an approach very similar, is used by Ohlson (1980) and Zmijewski (1984). Martin (1977) reports -2 log-likelihood statistics as well as classification tables for several logistic regression models and reports solely classification tables for discriminant analysis models.

Begley, Ming and Watts (1996) estimate an optimal cut-off point in a discrimant analysis model and optimal cut-off probability in their logit models to minimize type I and type II error and report classification tables. They also re-estimate the coefficients of the models previously employed by Altman (1968) and Ohlson (1980) and test for statistically significant differences in coefficients via Wald chi-squared tests and accuracy via classification tables. Grice and Ingram (2001) report classification tables with accuracy rates for both the distressed group and the non-distressed group in several periods and tests for a statistically significant difference in the same manner as Altman (1968). Chava and Jarrow (2004) use a similar kind of classification method by calculating the probabilities forecasted by the model, and then sorting the companies are grouped into deciles based on the default probabilities. It is then assessed what percentage of firms in each decile forecasted by the model actually went bankrupt. Campbell et al. (2008) investigate a sample spanning several years and distribute the total portion of bankruptcies occurring over the years. They plot for each year what proportion of firms has actually gone bankrupt against the proportion of bankruptcies that their model predicts.

The problem associated with classification tables is that it consists of an arbitrary approach (mostly involving conjectures) at determining a cut-off point. In the case of a z-score model, a cut-off point is a score where observations higher than the cut-off point should be classified as non-bankrupt and a score lower than the cut-off point should be classified as bankrupt. In studies like Altman (1968) and others, cut-off points are determined involving conjectures and comparisons of classification tables yielded by those conjectures. The second issue is that the measure reflects a static measure of accuracy, rather than a dynamic one where a flexible cutoff point is taken into account, i.e. what happens to type I and type II errors when the cut-off point is shifted. These issues are both addressed when measuring receiver operating characteristics.

2.3.2 Receiver operating characteristics

As of now, the technique that is dominant among evaluating the goodness-of-fit of a bankruptcy prediction model is Receiver operating characteristics. It is a tool to compare each of the models based on model-fit criteria and thus serves as a substitute for statistics like Pseudo R-squared, Cox and Snell (1968)'s R-squared, and others. The ROC curve is a widely used technique for assessing several models in their predictive accuracy in the bankruptcy prediction literature (used by i.a. Chava and Jarrow, 2004, Vassalou and Xing, 2004, Agarwal and Taffler, 2008, Lyandres and Zhdanov, 2013, Bauer and Agarwal, 2014, Kim and

Partington, 2015). A ROC-curve is constructed that takes into account the dynamics of the cut-off points and thus indirectly the (erroneous) assumption that Type I errors and type II errors are equally costly.

Bauer and Agarwal (2014) explain the principle behind constructing the ROC-curve. In order to construct the ROC curve, each year they sort the sample firms from high to low default probability. "For each integer x of the highest default risk firms (x%), we calculate the percentage of failed firms (number of firms that failed within the next year divided by the total number of failures in the sample)." Then, the numbers are cumulated for the total sample period. The plot of x% of highest default risk firms against its percentage of failed firms results are then displayed in the ROC curve. The quantitative metric is the area under the ROC-curve, called the AUC-statistic (Area Under (ROC-Curve). A similar approach is also followed by Vassalou and Xing (2004) and Christidis and Gregory (2010).

The ROC score is also discussed in Agarwal and Taffler (2008) and Lyandres and Zhdanov (2007). A scaled version of this score, known as the accuracy ratio is used by Moody's and is discussed in Vassalou and Xing (2004). A perfect model has an AUC score of 1, and a model that has no discriminatory power whatsoever has an AUC score of 0.5. Thus, a higher AUC score are examined indicates that the model is better able to discriminate between bankrupt and healthy firms.

Shumway and Bharath (2008) sort their data from highest to lowest probability of going bankrupt and consider per decile what percentage is predicted correctly by the model. They use other approaches more or less identical to ordinary classification tables (paragraph 2.2.1) throughout their study. Wu, Gaunt and Gray (2010) incorporate the original economic specifications of the model and evaluate that against each other. That means that each model (in their case, the models of Altman, Ohlson, Zmijewski, Shumway and Hillegeist et al.) is estimated and then applied to the data. They plot the occurrence of Type I errors (classifying a bankrupt firm as healthy) and Type II errors (classifying a healthy firm as bankrupt) according to model scores. For each model, the cutoff that minimizes the sum of Types I and II errors is then manual selected. These are compared to each other, where the model with the smallest portion of errors is the most accurate. The relation of this method with the ROC-curve approach is that choosing a minimum-error cut-off point is a static measure, while the ROC-curve takes changes of the cut-off point into account and represents the change in accuracy when the cut-off point is changed. Therefore, the ROC-curve is a more meaningful evaluation than just a static metric.

2.3.3 Other goodness-of-fit measures

Hair et al. (2006), Anderson (2007), as well as Tinoco and Wilson (2013) provide a number of alternative tests for the area under the ROC-curve test for evaluating model goodness-of-fit. Already discussed is the AUC-statistic. The Gini rank correlation coefficients and Kolmogorov–Smirnov statistics are widely used analysis tools by scoring analysts to assess the predictive accuracy of in-sample and hold-out tests (Tinoco and Wilson, 2013, Anderson, 2007). The advantage of these tests is that they are easy to interpret and calculate, as both can be derived from the AUC.

The Gini-rank coefficient is a way of measuring model goodness of fit. Tinoco and Wilson (2013) note that "the Gini coefficient is very similar to the AUC; the difference is that the former calculates only the area between the curve and the diagonal of the Lorenz curve, unlike the latter, which calculates the full area below the curve." Anderson (2007) argues that the Gini rank coefficient has been used frequently by credit scoring analysts, who employ it as a measure of "how well a scorecard is able to distinguish between goods and bads" (as quoted in Tinoco and Wilson, 2013) where "the end result is a value representing the area under the curve." As a reference point, in the context of professional credit scoring analysis, a Gini coefficient equal to or above 50% is a very satisfactory level in a retail environment (Tinoco and Wilson, 2013).

Anderson (2007) and Tinoco and Wilson (2013) also mention the Kolmogorov-Smirnov test. The two sample Kolmogorov-Smirnov test is a nonparametric test that compares the cumulative distributions of two data sets (such as the z-scores or default probabilities of both the bankrupt and non-bankrupt group) and measures "the maximum vertical deviation between two empirical cumulative distribution functions (good and bad) in credit score modeling." This test should be interpreted as that a significant result would mean that there is a difference between a bankrupt and a non-bankrupt cumulative distribution function, hence the model has discriminating power. [3]

2.3.4 Information content

The relative information content is a way to assess the explanatory power of a given model. The goal of this kind of evaluation is to assess the incremental information about bankruptcy captured by different models. The procedure of an information content evaluation of a model is as follows. At first, a series of bankruptcy probabilities predicted by one of the models is entered as time-varying covariate in a hazard model. This means that z-scores or logit-scores for consecutive years are included as independent variables in a hazard model, specified in section 2.1.4. It should be noted that the function of this hazard model is not, as in section 2.1.4, to serve as a bankruptcy probability model, but rather as an *evaluation* device for assessing the information content of other bankruptcy forecasting models. For a precise methodological specification, see section 3.2 as well as e.g. Chava and Jarrow, 2004 and Shumway and Bharath (2008).

The procedure, outlined by Shumway and Bharath (2008), is relatively simple. For each time period *t* that a firm is in the sample, a bankruptcy probability for each firm *i* is calculated according to a specific bankruptcy prediction model. These observations are then pooled together and serve as one independent variable in a hazard model. The according dependent variable is always 0 for non-bankrupt firms, but also 0 for bankrupt firms for all *t* at which they were not yet bankrupt, and only takes on the value of 1 at the time the firm actually went

[3] This test should not be confused with the Kolmogorov-Smirnov univariate normality test.

bankrupt. Applying this procedure also implies that we are using "event time" rather than "calendar time", as explained by Kim and Partington (2015). Then, the coefficient belonging to this bankruptcy probability is estimated and the statistical significance of its estimation coefficient is assessed[4], and the significance of the model as a whole. The interpretation of the outcome of this procedure is that if the coefficient shows statistical significance, the bankruptcy probability metric is highly correlated with the actual bankruptcy data. Bauer and Agarwal (2014) use an approach describe here, applied on the hazard-, accounting-, and contingent claims-based models by taking the probability of failure from each model as independent variable in a discrete-time hazard model with time-varying covariates. For each firm-year observation, three separate probability measures are computed using the coefficients estimated earlier, and these then serve as input into a hazard model, to assess statistical significance and relative information content. The question whether one probability metric subsumes the influence of the other metrics, or rather augments them can be answered this way.

Shumway and Bharath (2008) also use a Cox proportional hazard model with two covariates to test the information content of the DD model against the Naïve DD model (eq. 2.9 against eq. 2.14). They report the estimates of several Cox proportional hazard models with time-varying covariates with the DD model and its naïve counterpart, a few control variables, with, among others, the bond spread included as an additional explanatory variable. The addition of control variables allows evaluating model robustness: it can be tested whether a simple metric subsumes or adds relatively much explanatory power next to the bankruptcy probability metric. Hillegeist et al. (2004), as well as Agarwal and Taffler (2008), Wu, Gaunt and Gray (2010), among others, use a discrete-time hazard model to evaluate the explanatory power of each of the variables used in each separate model: they take as covariates in a hazard model the ratios used in each model in order to assess statistical significance, to see whether that particular set of ratios has explanatory power in explaining bankruptcy.

Chava and Jarrow (2004) use hazard models to test the significance of the predictor variables used by Altman (1968) and Zmijewksi (1984), as well as a re-estimation of the predictors used by Shumway (2001). Charitou et al. (2013) also perform Cox proportional hazard models to test the information content of several models based on Shumway and Bharath (2008) in combination with the annual default rate as a control variable. Trujillo-Ponce et al. (2013) use linear regression models of credit default swap spread on the DD model along with several other control variables as a way of assessing corporate credit risk.

2.3.5 Economic value when misclassification costs are different

A final way to evaluate a prediction model's performance is to simulate an economic scenario, where the performance of the models is assessed when costs of misclassifying a failed firm is different to the cost of misclassifying a firm that does not fail are taken into account. This scenario supposes that there is a bank with a certain amount of money available

[4] By using this particular estimation procedure, we introduce a bias in the standard error. For details and a correction for this bias, see Shumway (2001) as well as section 2.1.4 and 3.2 in this study.

to lend, and using each of the models to assess whether to lend to customers or not based on the bankruptcy probability metric. All firms (thus, bankrupt and non-bankrupt) in the sample serve as hypothetical customers and the bank decides via the model which loan demands will be granted by what premium. The banks are then compared in terms of their hypothetical profit they would have obtained. The simulation thus represents a scenario where the costs associated with the two misclassifications (classifying a bankrupt firm as non-bankrupt and vice versa) are different. While refusal to lend to a subsequently non-failed firm simply leads to the loss of extra revenue, lending to a firm that subsequently fails can lead to substantial losses. Agarwal and Taffler (2008) and Bauer and Agarwal (2014) propose a method in which they use a loan pricing model of Stein (2005) and Blöchlinger and Leippold (2006):

$$R = \frac{p(Y=1 \mid S=t)}{p(Y=0 \mid S=t)} LGD + k \qquad \text{Eq. (2.21)}$$

Where R is the credit spread, $p(Y=1|S=t)$ is the probability of failure for a score of t, $p(Y = 0|S = t)$ is the probability of non-failure for a score of t, LGD is the loss in loan value given default, and k is the credit spread for the highest quality loan. They assume a simple loan market worth £100 billion with banks competing for business each using a different bankruptcy prediction model. They assume all loans have the same price and that the banks reject customers that fall in the bottom 5.0% according to their respective models and quote a spread based on Eq. (2.21) for all the other customers. “The customer chooses the bank which quotes the lower spread. If the quoted spreads are equal, the customer randomly chooses one of the banks (or equivalently, the business is split equally between the banks. Each year, we independently sort our sample firms on their probability of failure based on different bankruptcy prediction models and group them into 100 categories for each of the models.” Then, they assess the economic value of using each of the models by evaluating bank profitability, return on assets and return on risk weighted assets.

2.4 Summary and implications for this study

This table will present a short and comprehensive summary of some the most important articles in bankruptcy prediction literature and the results and contribution they brought about. The table is sorted by methodological approach and focuses on novel methodological techniques and sampling practices, not on empirical corroboration.

Table 1: Overview and short summary of key research papers in bankruptcy prediction

Author:	Contributions:	Results:
Panel A: Discriminant analysis and precursors		
Beaver (1966)	Introduced pair-matched sampling.	78% of his sample of firms was classified correct, five years before failure based on ratio analysis.
Altman (1968)	Introduced MDA to forecast bankruptcy, multivariate context.	Extremely accurate, classifying 95 per cent of the total sample correctly.
Deakin (1972)	Tested for the univariate relevance of the ratios used in MDA.	Misclassification errors averaged 3%, 4,5%, and 4,5% for the first, second and third years respectively.
Mensah (1984)	Investigated the effect of change in (macro) economic factors on the accuracy of the model.	He finds that the models are fundamentally unstable and need to be re-estimated in order to retain their predictive capability.
Panel B: Probit estimation and logistic regression		
Martin (1977)	Introduction of new method, highly reliable sample.	Accuracy rates of around 91%.
Ohlson (1980)	Dropped pair-matched sampling, used proportions of population.	96.12%, 95.55% correctly classified within one year and two years, respectively.
Zmijewski (1984)	Showed that logit and probit must be weighed according to population proportion.	In the best case, he finds an accuracy of 97.1%, that is, 97.1% of the bankruptcy statuses was predicted correctly.
Panel C: Distance-to-default model		
Hillegeist et al. (2004)	Introduce DD model, compares with original and updated Altman (1968) and Ohlson (1980) models. Use hazard model to evaluate.	The market-based DD model provides more information about the probability of bankruptcy than the Z- and O-Scores, robust to various modifications.
Vassalou and Xing (2004)	Estimate default risk using DD model, introduce theoretical framework based on the Fama-French (1993) risk factors. Evaluate performance in multiple dimensions.	Size and BM effects are intimately related to default risk. Small firms earn higher returns than big firms, only if they also have high default risk. Value stocks earn higher returns than growth stocks, if their risk of default is high.
Shumway and Bharath (2008)	Introduce a simplified version of DD model and test it against the sophisticated version and control for several other default forecasters.	The naïve version performs better in hazard models and out-of-sample forecasts than the sophisticated model. Several other forecasting variables are also important predictors.
Agarwal and Taffler (2008)	Introduce three different dimensions in which to evaluate the model: accuracy, information content and economic value.	In terms of accuracy and information content, the models perform about equally well. The z-score model leads to greater bank profitability.
Panel D: Hazard models		
Lane, Looney and Wansley (1986)	Introduced survival analysis to forecast bankruptcy.	Results indicate a “reasonably good” fit. Cox model compares favorably to MDA.
Shumway (2001)	Introduces a hazard model, combines both market- and accounting-data.	Some of the ratios used by Altman (1968) and Ohlson (1980) are insignificantly related to bankruptcy risk and several market-driven

		variables are strongly related to bankruptcy probability.
Chava and Jarrow (2004)	Compare hazard models with probit and MDA. Add industry-specific variables, include monthly instead of yearly observations and extend the model to financial firms.	Show that hazard models are superior to other models in accuracy as well as information content. Industry groupings are shown to add predictive power.
Beaver et al. (2005)	Investigate the performance of hazard models and the incremental information content of various market-based and accounting-based variables	The robustness of hazard models is strong over time. The decline in the predictive ability of the accounting data is compensated for by market-based variables. A model which combines the two appears to perform the best.
Campbell et al. (2008)	Identify additional factors that explain distress risk. Test the hazard model in various ways.	They find that the value and size effects do not compensate for the risk of financial distress.

Panel E: Intelligent and miscellaneous models

Atiya (2001)	Introduced influential neural networks design, incorporated market-data based on the Merton (1974) distance to default measure	Out-of-sample prediction accuracy of 85.5% for a three-year-ahead forecast, and 87% for a one-year-ahead model.
Duffie, Saita and Wang (2007)	Construct a model including several macroeconomic proxies alongside DD model.	Corporate default depends significantly on the current state of the economy, especially on leverage of the firm. Distance to default is an important predictor.
Lyandres and Zhdanov (2013)	Provide a theoretical argument for the inclusion of and provide a measure for investment opportunities in their model.	Find that investment opportunities are related to the probability of bankruptcy and the inclusion of that factor improves prediction results.

The preceding literature review shows that contemporary research has provided for new evidence regarding what econometric specification fits the problem best and what predictor variables are most powerful in any of the used prediction models.

With regards to the first issue, we can say that the literature provides no strong consensus. Nam et al. (2008) note that restrictive presuppositions and structural limitations of traditional methods were detrimental to progress in the bankruptcy prediction literature. The traditional methods (MDA, logistic regression, probit analysis) are mainly based on a dichotomous classification of failure versus non-failure. The critical drawback of this type of classifying methods is that a dataset is assumed to be composed of two distinct and separate populations. In this case, classification accuracy is relatively high in the modeling phase because the fate of each firm is already known in the estimation procedure. However, conventionally, the prediction power decreases sharply when applied to a hold-out sample, i.e. for true forecasting purposes. It seems that in practice, bankrupt and non-bankrupt firms come from the same population.

Discriminant analysis, probit analysis and logistic regression all assume independence of entries: there exists no unobserved factor that causes certain observations to be correlated with one another. Formally, this is assumption is not likely to be met when using panel data with observations from different countries, because structural unobserved factors in a given year, given country or given industry are likely to cause these observations to be correlated. Furthermore, as mentioned these specifications are based on the presupposition that bankrupt and non-bankrupt firms come from two distinct populations. By contrast, the hazard model considers the samples to be drawn from an identical population instead of two distinct ones. Since the data of non-failed firms can be regarded as censored data, which is sampled from an identical population, the approach of the hazard model is more appropriate. The hazard model specification also does allow for dependence of variable entries.

Shumway (2001) shows that most of the existing methods that used a single-period classification model, with multi-period data from financial statements, have inconsistent and biased estimated parameters. Using a hazard model, by including all relevant observations over time, then corrects for the bias. While taking into account the inevitable bias in the logit, probit and discriminant analysis models due to the set-up of the models, Zmijewski (1984) has shown that asymptotic bias in these models, i.e. oversampling bankrupt firms can be overcome by adjusting the sample selection accordingly.

Theoretically, the DD model makes use of an excellent framework based on option-pricing (Merton, 1974) and efficient markets theory (Sloan, 1996, Fama, 1998). The advantage of this model is that it is in essence continuous, it can be updated at every moment and produce a probability of bankruptcy based on that information, while the Shumway (2001) approach is a discrete-time hazard model. Additionally, it has solid foundations in conventional economic theory. From a purely econometric point of view then, it should be clear that the model of Shumway (2001) is superior to the MDA, probit and logit approaches, but it is no use comparing this to the DD model, as that makes strictly no use of conventional econometrics. The only way to compare the latter with the Shumway (2001) approach is by empirical performance.

Empirically, however, the matter is not so obvious. Some researchers (Reisz and Perlich, 2007, Agarwal and Taffler, 2007, 2008) argue for the sustained relevance and predictive accuracy of discriminant analysis as econometric method despite the presence of bias, providing evidence that a discriminant analysis-based model performs well over a period of 25 years. Other research (e.g. Hillegeist et al, 2004, Wu, Gaunt and Gray, 2010) provides conflicting evidence, showing that Z-score models systematically underperform DD model and discrete-time hazard models. Yet, almost every study that attempts to evaluate several bankruptcy prediction models includes either the Altman (1968), Ohlson (1980) or Zmijewski (1984) model as a benchmark (see f.i. Hillegeist et al., 2004, Chava and Jarrow, 2004, Shumway and Bharath, 2008, Bauer and Agarwal, 2014).

Several alternative econometric specifications, such as neural networks (e.g. Atiya, 2001) or a multinomial logit model (e.g. Jones and Hensher, 2004) have also yielded good results, whereas others (e.g. quadratic discriminant anlaysis as in Altman et al., 1977) have not shown improvement relative to the three aforementioned standard models.

With regards to the choice of predictors, it is generally accepted that the Altman (1968), Ohlson (1980) and Zmijewski (1984) variables are fairly good predictors. However, research such as Shumway (2001) and Wu, Gaunt and Gray (2010) construct models in which some of these predictors fail to show significant predictive power. Various researchers (e.g. Mensah, 1984, Zavgren, 1985, Bilderbeek and Pompe, 2005) use factor analysis and retain the factors as predictors. This seems to yield comparable or worse model accuracy to the traditional Altman (1968), Ohlson (1980) or Zmijewski (1984) variables.

Looking at large cross-industry samples, there is some evidence (e.g. Chava and Jarrow, 2004) in support of the significant predictive ability of industry effects. With regards to samples over long time periods, Christidis and Gregory (2010) and Tinoco and Wilson (2013) argue and provide evidence for the inclusion of macroeconomic control variables.

As briefly mentioned in the introduction, the main contribution of this study to the literature is that it will feature a cross-EU sample. In table 2 below, I show a brief summary of recent bankruptcy prediction research conducted in Europe. The great majority of European studies have been conducted in the UK. A review by Bellovary et al. (2007) concludes that in countries such as Italy, Greece, Germany, Sweden, Austria, Finland, Spain, the Netherlands, Belgium, the latest research ranges from the late ‘90s to the early ‘00s. Doumpos et al. (2015) conducts a study using a DD-based custom model in a cross-EU sample (although it also includes non-EU member Switzerland). As far as I know, no other study attempted to estimate and test a bankruptcy prediction model in a cross-country sample.

Table 2: Recent bankruptcy prediction studies in Europe

Authors:	**Model(s):**	**Country:**
Bilderbeek and Pompe (2005)	MDA, Neural networks	Belgium
Agarwal and Taffler (2007)	MDA	UK
Agarwal and Taffler (2008)	DD, MDA	UK
Christidis and Gregory (2010)	Hazard model	UK
Tinoco and Wilson (2013)	Hazard model	UK
Bauer and Agarwal (2014)	DD, Hazard, MDA	UK
Doumpos et al. (2015)	DD (custom)	7 EU countries + Switzerland

Finally, this study takes up the issue first noted by Mensah (1984) and elaborated on and evaluated by Platt and Platt (1991), Grice and Ingram (2001) and Grice and Dugan (2003) regarding stationarity of predictor variables and accordingly, the generalizability of bankruptcy prediction models. As explained, discriminant analysis, probit analysis and logistic regression need a fair degree of stationarity in the predictor variables so as to retain accuracy. In the Shumway (2001) hazard model specification, this is less the case, as coefficient estimates should absorb changing firm-invariant bankruptcy risk over time, as well as changing means, variances and covariances between predictors over time, and should remain accurate under changing circumstances. However, researchers such as Christidis and Gregory (2010) and Tinoco and Wilson (2013) have shown that even in hazard models, macroeconomic indicators can improve model performance. This study will be a continuation of aforementioned research by attempting to test several hypotheses that relate to this problem.

3. Hypotheses, Methodology, Variables, Data

In this section, I will comprehensively elaborate on the research questions, and provide several hypotheses. Afterwards, it elaborates on the research design to test these hypotheses. The research design section will cover the general methodology, mentions important methodological issues in detail and provides justification of particular methodological choices. Then, I will present the model variables in detail and substantiate the arguments for the inclusion of several alternative bankruptcy prediction models to test model robustness. Then, I will describe the data and sample source used. This section will also address the criteria for inclusion of a firm in the sample. Finally, I will provide a short summary of the predictor variables used in the models, and several alternative bankruptcy prediction models that serve as benchmark.

Many studies roughly similar to this one (e.g. Altman [1968], Ohlson [1980], Zmijewski [1984], Grice and Ingram [2001], Agarwal and Taffler [2007], Bauer and Agarwal [2010]) have been conducted, either in a framework where only one bankruptcy probability model was used, or in a framework where they tested more than one model. The implicit purpose of conducting research similar to the last option is of assessing the fruitfulness of a particular econometric method rather than assessing the influence of some or more particular variables that might be correlated highly with future bankruptcy. To address the first issue is the first and foremost purpose of this study, finding a particular set of variables that correlates highly with future bankruptcy is a secondary purpose. The main research question that I am attempting to answer is **'**what is the accuracy and information content of various statistical techniques predicting bankruptcy for a sample of listed non-financial firms in the European Union?'

Altogether, four separate questions are addressed: (i) which bankruptcy prediction model has the highest accuracy in predicting bankruptcy? (ii) Which model carries the highest information content regarding bankruptcy? (iii) Does model performance improve, stay constant or decline under a changing macroeconomic environment in Europe? (iv) Does the particular set of ratios included in the model have an impact on either accuracy, information content, or both?

3.1 Hypotheses

This study features bankruptcy prediction models estimated by logistic regression and discriminant analysis. These two particular methods are chosen, despite the fact that they are biased econometrically, because of two reasons: they continue to perform reasonably well empirically (Reisz and Perlich, 2007, Agarwal and Taffler, 2007, 2008, Wu, Gaunt and Gray, 2010, Lee and Choi, 2013) and because of data availability: it is difficult to find time-series data on, for instance, past market capitalization for currently bankrupt firms, so the DD model would be impossible to use, whereas the Shumway (2001) hazard model and variables would be problematic, as empirical findings by Shumway (2001), Christidis and Gregory (2010), Wu, Gaunt and Gray (2010) and others indicate that market variables are essential to the performance of the model.

Discriminant analysis and logistic regression have both been an ad-hoc methodology suited to the problem, but have yielded excellent results many times. The research of Agarwal and Taffler (2007) and Reisz and Perlich (2007) shows that the Z-score model cannot be written off as a significant and reliable model in bankruptcy prediction, and a similar argument applies to logistic regression. Following Ohlson (1980) and Zmijewski (1984) among others, I argue that, ceteris paribus, the logistic regression model will bring about better (i.e. more) accurate results than a discriminant analysis model because of the less strict assumptions and a better econometric fit.

Specifically, discriminant analysis assumes a linear relationship between a change in one of the predictors and a change in the bankruptcy risk metric, while logistic regression and probit analysis allow for nonlinearity. A constant increase or decrease in a financial ratio might affect the bankruptcy probability in a nonlinear manner, which is consistent with logistic regression but not with MDA. Furthermore, the variance-covariance matrices of the predictors should be the same for both groups when using discriminant analysis, which is an assumption that almost never holds empirically. The output of the MDA model is a score which has a limited intuitive interpretation, since it is an ordinal ranking metric (see Ohlson [1980], Collins and Green [1982], Hair et al. [2006] among others). In contrast, in logistic regression no assumptions of this kind have to be made regarding prior probabilities of bankruptcy and/or the distribution of predictors (Ohlson, 1980, Hair et al., 2006). Hence, logistic regression should yield parameter estimates closer to the true population parameter with smaller estimated standard errors and should therefore perform superior to MDA.

Hypothesis 1a: *Everything else the same, a logistic regression model will show a greater accuracy than a discriminant analysis model in predicting bankruptcy.*

Hypothesis 1b: *Everything else the same, a logistic regression model will show a more significant information content than a discriminant analysis model in predicting bankruptcy.*

A good bankruptcy prediction model will not lose its accuracy and information content over time. However, as indicated in research by Altman (1968), Mensah (1984), Grice and Ingram (2001), Grice and Dugan (2003), coefficients of static models do need to be re-estimated in order to retain their accuracy over time. It seems likely that the models should lose accuracy when predicting out-of-sample, as the out-of-sample data represents a wholly different economic environment, given past empirical research. Mensah (1984) identifies three reasons why the financial characteristics of firms might change over time: (i) changes in the rate of inflation can impact firms either by increasing the costs of production and marketing which cannot be passed on with higher prices or, if passed on in higher selling prices, result in a drop in demand. (ii) periods of high interest rates or worsening credit availability may induce failure by raising borrowing costs in excess of profit margins and (iii) period of initial descent into a recession as well as when the recession ends and recovery begins, firms that may be forced into bankruptcy are those which cannot survive a sustained drop in sales.

Platt and Platt (1991) argue that model robustness is likely to exhibit economic differences from the periods in which the models were originally developed if factors such as relationships between the dependent and independent variables change, the average range of the independent variables changes over time and relationships among the independent variables change (e.g. a general increase in corporate leverage over time). Platt and Platt (1991) suggest that these changes are attributable to "shifts in the business cycle, corporate strategy, the competitive nature of the market, and technology." As we will see, the sample used in this study will provide an excellent opportunity to test this kind of hypothesis, because of the 2007 financial crisis and subsequent 2010 European debt crisis. However, Agarwal and Taffler (2007, 2008) note that their z-score model retains accuracy over a large period of time. Taking in consideration that as the larger part of empirical studies indicates that model accuracy will decline over time, this will be the default hypothesis.

Hypothesis 2: *The bankruptcy prediction models will not retain their accuracy and information content over time.*

Furthermore, the models take a significant amount of time and information to compute (Bauer and Agarwal, 2014). In general, information and the effort to obtain this are costly, and therefore it is important to know whether the estimated models outperform very simple competing indicators: in other words, whether it's worth the effort gathering the information needed for a particular model and consequently, its robustness to several easy-to-compute alternative prediction models. One can imagine that a bankruptcy prediction model such as Altman's (1968) should outperform a model where bankruptcy for a given set of firms is predicted as random (Tinoco and Wilson, 2013). Sun (2007) juxtaposes statistical bankruptcy prediction models against auditors' forecasts. Shumway (2001) argues that because the market equity of firms that are close to bankruptcy is typically discounted by traders, firm size is a bankruptcy predicting variable by itself. Similarly, Agarwal and Taffler (2007) argue that a simple profit-before-tax indictor can be a bankruptcy predictor. Bankruptcy prediction models that cannot beat these or other simple benchmarks in terms of predictive accuracy or information content have no or little predictive power. Several studies lack this benchmark or robustness test (e.g. Altman, 1968, Zmijewski, 1984, Grice and Ingram, 2001, Grice and Dugan, 2003). Other empirical research shows however, that good models, ranging from discriminant analysis (Agarwal and Taffler, 2007, 2008) to probit estimation (e.g. Sun, 2007), to logistic regression (Wu, Gaunt and Gray, 2010) to hazard models (Kim and Partington, 2015) significantly outperform simpler benchmark bankruptcy prediction models in both accuracy and information content.

Hypothesis 3a: *A bankruptcy prediction model will outperform a simple competing bankruptcy prediction model in terms of accuracy.*

Hypothesis 3b: *A bankruptcy prediction model will subsume information content of a simple competing bankruptcy prediction model.*

The last question that will be addressed concerns the particular sets of ratios used to compose a bankruptcy prediction model. Various researchers have used the Altman (1968), the Ohlson (1980) or the Zmijewski (1984) variables, or combinations of them: among others Mensah (1984), Grice and Ingram (2001), Shumway (2001), Hillegeist et al. (2004), Sun (2007), Wu, Gaunt and Gray (2010). Other research (e.g. Vassalou and Xing, 2004, Shumway and Bharath, 2008, Charitou et al., 2013) incorporated mainly market variables in their models.

Pompe and Bilderbeek (2005) categorize ratios in four separate categories: profitability ratios, activity ratios, liquidity ratios and solvency ratios. In their study, they show that model accuracy is not very sensitive to the particular set of ratios included in the model, as long as out of each category, one predictor is present. Shumway (2001) found that ratios originally incorporated by Altman (1968) were statistically insignificant in a multivariate test. The empirical research is ambiguous with respect to this issue, so the null hypothesis will be the default option.

Hypothesis 4: *model accuracy and information content will not be sensitive to a change of ratios used in the model estimation procedure, as long as one liquidity, solvability, profitability and activity predictor are included.*

3.2 Methodology

3.2.1 Hypothesis 1: Econometric method performance

Hypothesis one holds that ceteris paribus, logistic regression performs better in terms of accuracy (h1a) and information content (h1b) than discriminant analysis. There is an important methodological issue when testing this hypothesis. Note first that hypothesis 1 is explicitly about prediction. Therefore, it does not suffice to compute the AUC-statistics in the estimation sample, because that measures classification accuracy, not predictive accuracy. Also, in order to measure predictive accuracy or information content of a particular econometric model, it is required to separate the noise coming from possible nonstationarity of the predictor variables from the noise coming from using a particular econometric method. As far as hypothesis 1 is concerned, I am interested in the latter. That means a hold-out sample must contain observations exclusively from the same period, so as to rule out possible noise due to nonstationarity in the predictor variables. Therefore, the model coefficients will be estimated with observations from 2005 and tested on a hold-out sample on observations from 2006-2007. A robustness check is conducted. It indicates whether the results are influenced by the (arbitrary) choice of an estimation sample consisting of observations 2005 and a hold-out sample consisting of 2006-2007 observations. In the robustness check, I will re-estimate the parameters on an estimation sample of 2005-2006, and evaluate only in 2007.

Hypothesis 1a: The first class of bankruptcy prediction models, based on Altman (1968) are multiple discriminant analysis (MDA) models, as explained in section 2.1.1 and specifically, equation 2.1, which represents a function that will estimate weights $v_1 \dots v_n$ according to the procedure formalized in appendix A:

$$Z = v_1X_1 + v_2X_2 + \dots + v_nX_n + C \qquad \text{Eq. (2.1)}$$

The second class of bankruptcy prediction models are logistic regression models, based on Ohlson (1980) specified in equation 2.2. This represents a function that will estimate coefficients according to the procedure outlined in appendix B:

$$P = \frac{1}{1+e^{-\sum_j \beta_j X_{ij}}} \qquad \text{Eq. (2.2)}$$

Now, hypothesis 1a will be tested by first computing all bankruptcy probabilities according to each model, and then comparing the AUC-statistic (see section 2.3.2) of an MDA model with a logistic regression model, containing the same variables and the same estimation sample. A small difficulty arises: in order to compare the accuracy of the models, logistic regression will yield a probabilistic output, but the discriminant analysis will not. Therefore, following Bauer and Agarwal (2014), the score resulting from the discriminant analysis model will be transformed to a probability using

$$p = \frac{e^{z-score}}{1+e^{z-score}} \qquad \text{Eq. (3.1)}$$

so that the final input will be between 0 and 1.

Agarwal and Taffler (2007) provide a statistic for assessing whether two AUC statistics are significantly different from one another:

$$z = \frac{B-A}{\sqrt{se(A)^2+se(B)^2}} \qquad \text{Eq. (3.2)}$$

Standard errors of the AUC estimation are computed using the Hanley and McNeil (1982) property that the AUC-statistic represents the probability that a randomly chosen bankrupt subject is (correctly) rated or ranked with greater suspicion than a randomly chosen non-bankrupt subject. Under the assumption of probabilities being sufficiently precise to avoid 'ties', with discrete data points, the AUC-computation features approximation procedures, and hence has a standard error. The precise specification of the standard error can be found in Hanley and McNeil (1982) or Agarwal and Taffler (2007).

Hypothesis 1b: I test the relative information content of each model by a hazard model (also known as multi-period logit) approach. This procedure involves estimating the following model (eq. 2.16 and 2.17) with as input bankruptcy probabilities generated by a particular bankruptcy prediction model *X* for a number of consecutive firm-year observations (Shumway, 2001):

$$P = \frac{1}{1+e^{-y_{i,t}}} \qquad \text{(Eq. 3.3)}$$

where

$$y_{i,t} = \alpha + \beta_1 X_{i,t-1} + \sum_{j=2}^{M} \beta_j Z_{j,i,t-1} + \varepsilon_{i,t} \qquad \text{(Eq. 3.4)}$$

Where y_{it} represents the bankruptcy status (either non-bankrupt, 0 or bankrupt 1) of firm *i* at time *t… t-j*. Then, *X* represents a probabilistic input generated by a bankruptcy prediction model for each firm *i* at time *t-1*. Consequently, this specification pools all probabilities of bankruptcy for firm *i* according to bankruptcy prediction model *X* and includes all observations so far as they are available in the dataset (from time *t-1…t-j*). This means data from bankrupt firms in the sample are also included, with the dependent variable being 0 as long as they were not yet bankrupt, and 1 at the time they were (see e.g. Agarwal and Taffler, 2007, Partington and Kim, 2015 for a similar procedure).[5]

Concretely, since this particular hypothesis is only tested on a hold-out sample in 2006 and 2007, this means that data from firms that went bankrupt in 2007 are also present in the hold-out sample as non-bankrupt firms in 2006, and non-bankrupt firms in 2007 are also included as non-bankrupt firms from 2006.

Furthermore, Z represents the inclusion in the information content test of M-1 alternative predictors (see section 3.4.3). Shumway (2001) notes that for bankruptcy prediction purposes, a hazard model estimated as a pooled-logit model needs a baseline hazard rate as predictor. A short description and justification for my particular choice for a baseline hazard rate can be found in Appendix C. This variable will serve as the only alternative predictor when testing hypothesis 1.

A small difficulty arises: probabilistic inputs are not in line with the econometric assumptions of the hazard model. I follow Hillegeist et al. (2004) and Bauer and Agarwal (2014) in solving this problem by transforming the estimated probabilities from the logistic regression model into general scores:

$$score = \ln\left(\frac{p}{1-p}\right) \qquad \text{Eq. (3.5)}$$

The procedure requires that for each separate bankruptcy prediction model[6], the statistical significance of the particular coefficient is assessed, and if it shows significance, then the model carries meaningful information about bankruptcy. A procedure like this allows for a statistical judgment of the information content of the bankruptcy prediction models. By executing this procedure, separately for each class of bankruptcy prediction models, hypothesis 1b can be tested. Hypothesis 1b is rejected if the level of statistical significance for models estimated using discriminant analysis is systematically lower than the level of statistical significance of those models estimated using logistic regression.

3.2.2 Hypothesis 2: Stationarity in the predictor variables

Hypothesis 2 holds that the models will not retain their accuracy and information content over time. Hold-out sample tests represent rigorous tests of the models' accuracy and information content. A critique raised by Grice and Ingram (2001) mentions that in many previous studies, the hold-out sample accuracy rates reported in prior studies are potentially upwardly biased (meaning the hold-out sample accuracy rates are higher than the rates users should expect when they apply the models) because time periods for the estimation and hold-out sample are

---

[5] Shumway (2001) notes that "the test statistics produced by a logit program are incorrect for the hazard model because they assume that the number of independent observations used to estimate the model is the number of firm years in the data. (…) Dividing these test statistics by the average number of firm years per firm makes the logit program's statistics correct for the hazard model."

[6] So, continuing using notation from eq. 3.2, the set of bankruptcy prediction models $M_1 \ldots M_n$ consists only of one model M when testing this hypothesis.

not substantially different and because the hold-out sample consisted of firms from the same restricted set of industries as those in the estimation sample. Now, this specific test addresses both biases by using a hold-out sample from a different period and using a wide variety of industries in both estimation and hold-out samples.

First, unlike in hypothesis 1, discriminant analysis models and logistic regression models will be estimated using the full estimation sample from 2005-2007 with various sets of predictors (see section 3.4.3). Then, for all observations in the estimation sample, from 2005-2007, bankruptcy probabilities will be calculated according to each model. The accuracy of each model is assessed again through the AUC-statistic (see section 2.3.2). Then, using the same bankruptcy prediction model, the bankruptcy probabilities are also computed in the hold-out sample, from 2011-2013. The AUC-score of each model in the hold-out sample is then compared to the AUC-score of its counterpart in the estimation sample.

Similarly, for the information content test, first, all bankruptcy probabilities according to one particular model in the 2011-2013 holdout-sample will be computed. Then, eq. 3.3 is estimated with a bankruptcy probability scores for each firm-year observation. The evaluation now consists of assessing the level of statistical significance of a particular model using the (out-of-sample) 2011-2013 data. The information content test will include a particular bankruptcy predictor model X, and the annual default rate Z_1, but no other alternative prediction model.

Using the terminology from section 2.1.1.4 and 2.1.2.4, applying the accuracy and information content tests on the 2011-2013 observations without re-estimating the parameters, are 'dynamic tests'. If the alpha-level of statistical significance >10%, then these models have lost information content, if it is <10%, then these models have retained their information content over time. This procedure will be conducted for each bankruptcy prediction model separately.

Hypothesis 2 is rejected if the AUC-score in the 2011-2013 sample is lower than the AUC-score in the 2005-2007 sample, and if the coefficient of the bankruptcy prediction model in the 2011-2013 sample in the information content test fails to show statistical significance.

3.2.3 Hypothesis 3: Performance relative to simple benchmark

Hypothesis 3 holds that a bankruptcy prediction model containing various predictors and estimated with sophisticated econometric techniques should outperform any simple alternative in terms of both accuracy and information content. As the literature review shows, only a handful of empirical research employs benchmarking, i.e. juxtaposing the results of a bankruptcy prediction model against another bankruptcy prediction model (as in e.g. Wu, Gaunt and Gray, 2010) or against a naïve alternative (as in e.g. Agarwal and Taffler, 2007). In most studies (e.g. Altman, 1968, Grice and Dugan, 2003), models are being tested without asking the question of whether another sophisticated or simple model might do better, though in recent times we have seen much more comparisons (Wu, Gaunt and Gray, 2010, Agarwal and Taffler, 2007, Bauer and Agarwal, 2014). This research explicitly addresses that question by attempting to robustness check the results obtained from testing hypothesis 2.

Hypothesis 3a: In the accuracy test, we will compare the AUC-score obtained by each bankruptcy prediction model[7] in the 2011-2013 hold-out sample with the AUC-score of several "naïve" alternative models. These are mentioned in section 3.3.3. If the AUC-score of alternative models is significantly higher than the AUC-score of the bankruptcy prediction model, hypothesis 3a is rejected.

Hypothesis 3b: In the information content test, a generalized version of the information content test of hypothesis 2 is conducted, again in the 2011-2013 hold-out sample. Again, the annual default rate is added as a baseline hazard approximator, but instead of using only one particular bankruptcy prediction model as a predictor (X), the naïve alternative models (Z_j) are added to the same model. The alternative bankruptcy prediction models are presented in section 3.3.3.

Using the notation from eq. 3.4 again, the bankruptcy prediction models included in the test now encompass the model in question X, being the particular bankruptcy prediction model that is tested, and all alternative prediction models, Zj being alternative models. As such, it is a more generalized version of hypothesis 2. Hypothesis 3b is rejected if the coefficient of the bankruptcy prediction model fails to show statistical significance or if the coefficient of an alternative naïve model shows a greater level of statistical significance than the coefficient of the bankruptcy prediction model.

3.2.4 Hypothesis 4: Robustness to alternative sets of predictors

Hypothesis 4 holds that a model's accuracy and information content is not sensitive to the particular sets of ratios, as long as certain dimensions are present in a bankruptcy prediction model. Pompe and Bilderbeek (2005) propose to identify four dimensions as liquidity ratios, activity ratios, profitability ratios and solvability ratios. I test this hypothesis by estimating several models (in the 2005-2007 period) with combinations of variables that include at least one activity variable, one profitability variable, one solvability variable and one liquidity variable. Furthermore, I will estimate models excluding and including both miscellaneous and macroeconomic variables. Finally, I will use factor analysis to select ratios. This will be elaborated on in section 3.4. These models are then subjected to the same accuracy and information content tests described under hypothesis 2, and their scores are compared with the standard specifications mentioned before. If the AUC-scores of these models are significantly lower than the standard alternatives, or if in the information content tests, these models show a lower level of statistical significance, then hypothesis 4 is rejected.

[7] Estimated as explained in section 3.2.2, i.e. using the 2005-2007 data.

3.3 Variables

3.3.1 Dependent variables

The dependent variable is in all models 1 if a firm went bankrupt in the year of the observation (as defined earlier) and 0 if a firm does not go bankrupt. In the estimation procedure of the bankruptcy prediction models, each firm is only observed once. In the information content evaluation procedure, a firm, either bankrupt or non-bankrupt, contributes multiple observations, as explained in section 2.1.4 and section 3.2.

3.3.2 Independent and control variables

Independent variables: When evaluating the models on their accuracy and information content, the independent variables are always bankruptcy probability scores. However, in this section I am concerned with what predictors are actually included in the bankruptcy prediction model so as to produce a bankruptcy prediction model, and consequently, a score.

For the sake of a general and robust analysis, I use several sets of ratios as possible predictors and I will test the hypotheses separately for all sets. This way, I can rule out that the obtained results are due to bad predictor choice. First, I will use a set of ratios not selected on a theoretical or strictly empirical basis, but rather on the basis of their performance in prior studies. This practice is also followed by Zmijewski (1984), Grice and Ingram (2001), Shumway (2001), Wu, Gaunt and Gray (2010), Agarwal and Taffler (2007, 2008), among others. The variables originated in the study of Altman (1968) and are 1. Working capital/total assets, 2. Retained earnings/total assets, 3. EBIT/total assets, 4. Market equity/total liabilities, 5. Sales/total assets. These variables either exactly match the variables used by Altman (1968), Deakin (1972), Zavgren (1985), Grice and Ingram (2001), Grice and Dugan (2003) and Wu, Gaunt and Gray (2010). These studies reported relatively high accuracy rates for the models using small samples and short windows of time.

I also use the Zmijewski (1984)'s variables. They include 1. the ratio of net income to total assets (NI/TA), 2. the ratio of total liabilities to total assets (TL/TA), and 3. the ratio of current assets to current liabilities (CA/CL). These variables are also used frequently by researchers to evaluate any bankruptcy prediction model (Shumway, 2001, Grice and Dugan, 2003, Wu, Gaunt and Gray, 2010).

Additionally, I use the set of Ohlson (1980) variables, which are also used by Begley, Ming and Watts (1996), Grice and Dugan (2003), Canbas et al. (2005), Wu, Gaunt and Gray (2010). They are 1. LTA = Log(total assets/GNP price-level index), 2. TLTA = Total liabilities divided by total assets, 3. WCTA = Working capital divided by total assets, 4. CLCA = Current liabilities divided by current assets, 5. OTLETA = 1 If total liabilities exceed total assets, 0 otherwise, 6. NITA= Net income divided by total assets, 7. IOTL = Funds provided by operations (income from operation after depreciation) divided by total liabilities, 8. NEN2

= 1 If net income was negative for the last 2 years, 0 otherwise and 9. CHIN= $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where NI_t is net income for the most recent period.

Furthermore, I will use ratios on an empirical basis, entirely based on Pompe and Bilderbeek (2005). In some studies on bankruptcy prediction, factor analysis has been used to select combinations of ratios for use in the models (see e.g. Mensah, 1984, Zavgren, 1985). With factor analysis, the object is to describe the covariance relationships among many variables in terms of a few underlying factors. In this study, I will also use factor analysis, identifying factors from all the ratios listed in appendix D. The criterion chosen for deciding how many factors to retain was that the factors should account for at least 70% of the total variance. Then, for each factor retained, the most important contributor to the factor will be included as variable in the model. These variables are: 1.Net income/total assets, 2. Working capital/operating revenue, 3. Current assets/total assets 4. Current assets/current liabilities, 5. Turnover/current assets, 6. Natural log of total assets, 7. GDP growth (%), 8. Market cap/total assets.

Finally, I use two sets of alternative predictor variables, each predictor representing activity, liquidity, profitability and solvability, so as to test model sensitivity to particular ratio choice. First, I will use 1. Working capital/operating revenue (liquidity), 2. EBIT/market cap (profitability), 3. Current liabilities/total assets (solvability), 4. Market cap/turnover (activity). Second, I will use 1. Working capital/sales (liquidity), 2. Net income/market cap (profitability), 3. Long-term debt/total assets (solvability) and 4. Working capital/number of employees (activity).

Control variables: Nam et al. (2008), Christidis and Gregory (2010) and Tinoco and Wilson (2013) argue that there exists a firm-invariant bankruptcy risk that relates to the macroeconomic environment. Christidis and Gregory (2010) mention that it is possible that macroeconomic indicators can add to information in prices insofar as market prices reflect expected cash flow s and discount rates, whilst economic variables may tell us something about downside risk, which is of key importance in bankruptcy prediction models.

When these variables are included as controls, they may account for the difference in firm-invariant default risk, i.e. differences in macroeconomic environment in the 2005-2007 period and the 2010-2013 period.

Similarly, Tinoco and Wilson (2013) argue that the incorporation of time variant data into credit risk models that captures changes in the macro-economic environment is important in two main respects. First it adds a dynamic element to the models that acts to adjust risk scores (likelihood of insolvency) in relation to changing macro-economic conditions. Second, they recapitulate that such models would have a built-in facility to stress test probability of default estimates across the portfolio. Therefore, I include two macroeconomic variables as controls: 1. the inflation rate, as measured by the change in consumer price index (%) and 2. the real GDP growth rate (%).

As a general rule, for each firm-level variable used in this study, the 1st and 99th percentile will be winsorized and set to that value, in order to prevent excessive influence of outliers. The macro-economic and dummy variables will be left unwinsorized.

3.3.3 Alternative bankruptcy prediction models

I introduce various competing bankruptcy prediction models in the information content tests, in particular to test hypothesis 3. These variables are thus at a different level than the variables presented in the previous subsection. In the previous subsections, I presented sets of variables to be included in a particular bankruptcy prediction model. In this subsection, I present alternative bankruptcy prediction models, that will serve as a benchmark. Hence, these variables are not included in a bankruptcy prediction model, but serve as a competitor to them. The rationale behind including these alternative bankruptcy prediction models in the accuracy and information content evaluation is that a good model should outcompete any simple, random or naïve bankruptcy prediction indicator such as firm size as measured by its market capitalization (Shumway, 2001), a random model (Agarwal and Taffler, 2007) or market beta's (Bauer and Agarwal, 2014) in both information content and predictive accuracy: if a simple indicator is correlated more highly with actual bankruptcy than an entire bankruptcy prediction model, one would be better off using that.

Additionally, bankruptcy prediction models require a significant amount of (costly) information and effort to estimate, so one might want to know the marginal gains in accuracy and information content by employing a bankruptcy prediction model rather than a single, naïve indicator. The control variables employed in this study are based on Shumway (2001), Agarwal and Taffler (2007) and Bauer and Agarwal (2014).

Following Agarwal and Taffler (2007, 2008), I introduce two naïve bankruptcy indicators: one that classifies all firms as non-failures, and a simple accounting-based model that classifies firms with negative profit before tax in the last available year (PBT<0) as potential failures and those with PBT>0 as non-failures and compute the AUC-score for these two indicators and then compare it with the AUC-score of the bankruptcy prediction models. Following Shumway (2001) I also add a size proxy to incorporate market perceptions of a particular firm, because the market equity of firms that are close to bankruptcy is typically discounted by traders, so size becomes a bankruptcy prediction variable by itself. Shumway (2001) uses market capitalization, but due to the difficulty of obtaining time-series data on bankrupt firms' stocks, I compute various proxy market capitalization variables as (Francis and Schipper, 1999):

$$Proxy\ Market\ Cap_t = 0.25 * Total\ assets_t + 6.70 * Earnings_t \qquad \text{Eq. (3.6)}$$

$$Proxy\ Market\ Cap_t = 0.78 * Total\ assets_t + 4.28 * Earnings_t \qquad \text{Eq. (3.7)}$$

$$Proxy\ Market\ Cap_t = 0.73 * Total\ assets_t + 3.77 * Earnings_t \qquad \text{Eq. (3.8)}$$

following empirical estimates over a 42-year period of Francis and Schipper (1999). The three combinations of coefficients represent respectively the average, the latest year and the year with the highest coefficient of determination. Every one of these proxies ought to explain around 70% of the variation in market capitalization. They will all separately be included in the evaluation procedure as competing bankruptcy prediction models. These estimates were the last available and only well-performing estimates that were available, to my knowledge.

Lastly, following Bauer and Agarwal (2014), I add a growth proxy that should serve as a naïve bankruptcy predictor for small growth opportunities and bad market perceptions. Bauer and Agarwal (2014) use a firm's market-to-book ratio, but due to aforementioned data constraints, I use a sales growth indicator [($sales_t$-$sales_{t-1}$)/$sales_{t-1}$] as a substitute.

3.4 Data

As a data source, I use ORBIS, an international database by Bureau van Dijk for financial data of many listed and non-listed firms throughout the world. Additional information is gathered through annual reports of the companies present in the sample, and data for macroeconomic indicators comes from the World Bank Database. Companies that went bankrupt from 2005 to 2007 are included in the estimation sample, with any of those years as a target year of bankruptcies, so that the latest available data comes from 2004-2006. This is then augmented by parallel observations of non-bankrupt companies in the same period.

The hold-out sample includes companies that went bankrupt in 2011 until 2013, so that the latest available data comes from 2010 until 2012. This is also augmented with parallel observations of non-bankrupt companies in the same period. Omitting bankruptcies from the years 2008-2010 from my sample allows me to clearly separate pre-crisis from post-crisis observations.

The convenience of using the ORBIS database is that it is largely standardized, i.e. there is no differentiation of the exact legal status of a bankrupt firm in a particular country – therefore, it allows the researcher to sort firms on a bankrupt status without explicitly having to take into account the legal position of a firm. From this follows that a firm in my sample is registered as bankrupt if and only if its status entry into ORBIS is "bankruptcy" or "dissolved (bankruptcy)" within the sample period.

The bankrupt firms in the sample must meet the following criteria:

1. The firm status must be listed under 'bankruptcy' or 'dissolved (bankruptcy)' in the database and must have gone bankrupt between 2005 and 2007 or 2011 until 2013.
2. The firm must be an industrial company.
3. The firm must have been publicly listed during the years it was present in the sample.
4. All required data must be (directly or indirectly) available.
5. The firm must be in the European Union as of 2015.

This yields a sample of 1234 bankrupt firms that experienced their bankruptcy between 2005 and 2007, and 1136 between 2011 and 2013. Financial firms are excluded by default, following common practice (e.g. Altman, 1968, Shumway, 2001, Shumway and Bharath, 2008) and because financial characteristics for financial firms are in generally different from those of industrial firms.

These observations will be augmented with 7299 randomly selected non-bankrupt firms in the estimation sample and 6993 firms in the hold-out sample by augmenting the bankrupt sample by approx. a factor 7 because "it is by no means obvious what is gained by pair-matched sampling" (Ohlson, 1980) and because of the econometric issues regarding the estimation of the log-likelihood function, elaborated on in section 2.1.2 and in Zmijewski (1984). Ratios of bankrupt to non-bankrupt firms in other studies range from 1:1 (e.g. Altman, 1968), to approx 1:5 (e.g. Sun, 2007) to approx. 1:100 (Begley, Ming and Watts, 1996, Bauer and Agarwal, 2014).

The non-bankrupt firms must meet the following criteria:

1. The firm status must be listed under 'active' in the database.
2. The firm must be an industrial company.
3. The firm must be publicly listed during the years it was present in the sample.
4. All required data must be (directly or indirectly) available.
5. The firm's asset size must be the average asset size of the bankrupt firms ±3SD (following Altman [1968]).
6. The firm must be in the European Union as of 2015.

It is clear that inter-country and inter-firm comparability may be a problem because of different fiscal years, different accounting standards, etc. One of the great advantages of the ORBIS database is that this problem is bypassed because of the common international format of balance sheets. Furthermore, ORBIS clearly differentiates firms that have been dissolved via a merger and via a bankruptcy. This differentiation protects me from including data of e.g. merged firms in a set of bankrupt firms.

Similar to Mensah (1984) and Grice and Dugan (2003), in the testing of hypotheses 2, 3 and 4, two subsets of samples are used in analyses for this study. The first one provides an estimation sample, and the second one is a hold-out sample. The hold-out sample and the estimation sample represent different economic climates, as the estimation sample includes observations in the period of 2005-2007 and the hold-out sample includes observations in the period of 2011-2013. This cut-off point in the sample gives me the possibility to evaluate model performance and robustness of performance under different macro-economic circumstances, as the financial crisis hit the real economy in Europe in 2008. By excluding the years 2008, 2009 and 2010 from my analysis, I have cut the estimation and hold-out samples into clear pre-crisis and post-crisis samples, so there is no noise present as to whether a firm went bankrupt because of the macroeconomic effects or because of idiosyncratic circumstances. Below is table 3, summarizing average GDP-growth and inflation in the EU (28) from 2005-2007 and from 2011-2013. P-values indicate the results of a two-sided t-test. Following the 3 criteria set up by Mensah (1984), (i) the average rate of inflation has certainly changed, but (ii) the average interest rates have hardly changed (although the variance has greatly increased) and (iii) the economic boom has turned into a bust, with almost all European economies not yet at GDP levels before the financial crisis. Therefore, there is sufficient reason to suppose macroeconomic conditions have changed, in such a way that it could introduce possible noise into the estimation procedure if observations from 2008 onwards were included.

Table 3: Avg. inflation rates, interest rates and GDP growth rates in the EU

Indicator:	Mean 2005-2007	Mean 2010-2013	p-value
Inflation rate	3.15	2.70	<0.10
Interest rate	4.34	4.70	>0.10
GDP growth rate	4.69	0.57	<0.01

Using a hold-out sample like this provides a much more rigorous test of bankruptcy prediction models than is generally the case (e.g. Grice and Ingram, 2011, Grice and Dugan, 2003). Since the primary application of z-score models is to forecast future events, the most valid test of their performance is to measure their true predictive performance. A tumultuous macroeconomic environment provides a tougher challenge to a model than a relatively stable economic environment.

Table 4 below shows the sample composition. Panel A features the 2005-2007 sample characteristics and panel B features the 2011-2013 sample characteristics. Both subsamples show similar characteristics. Sectors are present in roughly the same proportions, as are countries. The number of bankrupt firms is slightly skewed towards 2007 in the 2005-2007 subsample, and slightly skewed towards 2011 in the 2011-2013 subsample. It is clear that in both subsamples, Belgian, Spanish and French firms show the highest presence.

Given point 5 in the sample selection criteria of non-bankrupt firms, it is obvious that the sample should include in ‘median’ European countries at the highest frequency. If observations from firms located in a particular country are correlated with one another, my sample might lead to biased parameter estimates towards ‘median’ EU countries, as these are most present in my sample, because it violates the independent observations assumption mentioned in section 2.1.1.3 and 2.1.2.3. Following Tinoco and Wilson (2013), country-specific macroeconomic variables are added as a way to correct for this bias towards median countries in het sample, so as to serve as a variable that captures possible changing macroeconomic circumstances that affect country-level bankruptcy risk. A test whether the results are robust to a sample composition including countries with different characteristics (i.e. poorer or richer EU-countries) at a higher frequency is included in section 4.

Table 4: Sample composition

Panel A: 2005-2007 Sample

Country:	**No. Obs.**	**Sector:**	**No. Obs.**	**Year of bankruptcy***	**No. Obs.**
Austria	7	Chemicals	316	2005	375
Belgium	1425	Construction	944	2006	340
Bulgaria	53	Education, Health	157	2007	519
Czech Republic	155	Food, Beverages, Tobacco	269		
Germany	95	Gas, Water, Electricity	56		
Denmark	2	Hotels, Restaurants	369		
Estonia	53	Machinery, equipment, furniture, recycling	628		
Spain	1156	Metals	414		
Finland	3	Other services	1943		
France	3685	Post, telecom	32		
Great Britain	37	Primary sector	188		
Greece	471	Public administration	1		
Croatia	15	Publishing, writing	195		
Ireland	2	Textiles	216		
Italy	748	Transport	367		
Lithuania	4	Wholesale & retail trade	2251		
Luxemburg	32	Wood, cork, paper	125		
Latvia	1				
Netherlands	13				
Poland	54				
Portugal	292				
Romania	94				
Sweden	68				
Slovenia	13				
Slovakia	50				
Total:	8528	**Total:**	8528	**Total:**	1234

Panel B: 2011-2013 sample

Austria	2	Chemicals	336	2011	525
Belgium	1482	Construction	933	2012	494
Bulgaria	49	Education, Health	156	2013	117
Cyprus	1	Food, Beverages, Tobacco	273		
Czech Republic	125	Gas, Water, Electricity	45		
Germany	112	Hotels, Restaurants	373		
Denmark	3	Machinery, equipment, furniture, recycling	575		
Estonia	60	Metals	327		
Spain	1110	Other services	1772		
Finland	1	Post, telecom	21		
France	3018	Primary sector	190		
Great Britain	36	Public administration	4		
Greece	451	Publishing, writing	166		
Croatia	25	Textiles	162		
Ireland	3	Transport	370		
Italy	953	Wholesale & retail trade	2202		
Lithuania	4	Wood, cork, paper	101		
Luxemburg	25	Misc.	123		
Latvia	6				
Netherlands	12				
Poland	62				
Portugal	281				
Romania	87				
Sweden	140				
Slovenia	10				
Slovakia	63				
Total:	8129	**Total:**	8129	**Total:**	1136

Note: * = bankrupt firms only

3.4 Summary of predictor variables and alternative models

Altogether, the following table presents the variables used in this study. Panel A,B,C,D, E and F give the different sets of predictor variables used and that serve as input in computing a bankruptcy probability (in the case of logistic regression) or bankruptcy score (in the case of discriminant analysis). Each of those sets of predictors, except those in panel D, are also augmented by two macroeconomic predictors (in panel G) and serve as a different set of predictors. Panel H gives variables at a different level, i.e. alternative bankruptcy prediction models used in the accuracy tests (i.e. the AUC-analyses) and in the relative information content tests.

Table 3: Independent variables, alternative bankruptcy prediction models incorporated in this study

Panel A: Altman (1968) independent variables	**Panel B: Ohlson (1980) independent variables**
A1. Working capital/total assets	O1. Natural log of total assets
A2. Retained earnings/total assets	O2. Total liabilities/total assets
A3. EBIT/total assets	O3. Working capital/total assets
A4. Market cap/total liabilities	O4. Current liabilities/current assets
A5. Sales/total assets	O5. 1 If total liabilities exceed total assets, 0 otherwise
	O6. Net income/total assets
	O7. Income from operations after depreciation/total liabilities
	O8. 1 If net income was negative for the last 2 years, 0 otherwise
	O9. Change in net income for the most recent period
Panel C: Zmijewski (1984) independent variables	**Panel D: Factor analysis independent variables**
Z1.Net income/total assets	F1.Net income/total assets
Z2.Total liabilities/total assets	F2. Working capital/operating revenue
Z3. Current assets/current liabilities	F3. Current assets/total assets
	F4. Current assets/current liabilities
	F5. Turnover/current assets
	F6. Natural log of total assets
	F7. GDP growth
	F8. Market cap/total assets
Panel E: Alternative set of variables (1)	**Panel F: Alternative set of variables (2)**
B1.Working capital/operating revenue	C1.Working capital/sales
B2.EBIT/market cap	C2.Net income/market cap
B3. Current liabilities/total assets	C3.Long term debt/total assets
B4.Market cap/turnover	C4.Working capital/number of employees
Panel G: Macroeconomic variables	**Panel H: Alternative bankruptcy prediction models**
M1. GDP growth (%)	1. PBT<0 indicator (Agarwal and Taffler, 2007)
M2. Inflation rate (%)	2. All firms non-failure (Agarwal and Taffler, 2007)
	3. Proxy market cap (Shumway, 2001)
	4. Sales growth (Bauer and Agarwal, 2014)

4. Results

In this section, I provide the results of my tests. First, in table 4, I present summary statistics: mean, median, SD, minimum, maximum for every variable used in the 2005-2007 estimation sample and the 2011-2013 hold-out sample, which is used for testing hypothesis 2, 3 and 4. Descriptive statistics for the data used to test hypothesis 1 can be found in appendix E, as can the raw descriptive statistics for the 2005-2007 and 2011-2013 observations prior to listwise exclusion. Then, I show the results of the tests, each hypothesis separately. Only the final bankruptcy prediction model performance is reported. The estimation procedure and specific parameter estimates are left out. A complete summary of parameter estimates for all models can be found in appendix F (for hypothesis 1) and appendix G (for hypotheses 2, 3 and 4). Then, I show the results of each test and will interpret this in light of the hypotheses.

4.1 Descriptive statistics

Table 4 presents descriptive statistics of the data that are used to estimate the models that are tested throughout hypotheses 2, 3 and 4. Hypothesis 1, for reasons mentioned in section 3.2.1, used only data from the 2005-2007 period. The descriptive statistics that are used testing this particular hypothesis can be found in appendix E (table E1). The descriptive statistics reported on the following page stem from the fact that I use unbalanced panel data. Hence the amount of entries *N* differs per variable. Since the data consists mainly of financial ratios, it is also possible that inverse ratios, such as CACL and CLCA have a different *N*, because anything divided by zero is censored. However, one can observe that in such cases, the variable median value is not generally influenced by this. More elaborate statistics, such as the (list-wise) descriptive statistics of the data that are used estimating the models, can also be found in appendix E (table E2). There, the predictors are also sorted by specific set of variables.

The data looks comparable to the data used by e.g. Grice and Dugan (2003) and Wu, Gaunt and Gray (2010), although in some of the sets of alternative predictors, some predictors appear to have a rather large variance. For instance, retained earnings to total assets appears to have a rather negative minimum value, but this is consistent with e.g. Wu, Gaunt and Gray (2010), who report even a mean negative retained earnings to total assets ratio for both bankrupt and non-bankrupt firms. Average firm size appears to be somewhat larger than in other studies. Other ratios used by Altman (1968), Ohlson (1980) and Zmijewski (1984) are grossly similar in magnitude to Shumway (2001), Grice and Dugan (2003), and Wu, Gaunt and Gray (2010).

Some of the alternative predictors, taken from Pompe and Bilderbeek (2005) appear to have a rather large standard deviation. All sets of predictors have, when balancing the data, approximately 7000 observations, with the exception of set F, due to limited availability of the no. of employees for each firm.

Table 4: Descriptive statistics (Unbalanced)

Panel A: Estimation sample (2005-2007)

	WCTA	RETA	EBITTA	MCTL	SATA	TA*	TLTA	CLCA	OTLETA	NITA	IOTL	NEN2	CACL
Mean	0.229	0.010	0.037	1.065	1.633	5.751	0.646	1.198	0.080	0.014	0.121	0.141	2.456
Median	0.214	0.021	0.041	0.168	1.390	1.946	0.633	0.775	0.000	0.025	0.065	0.000	1.284
SD	0.273	0.182	0.175	4.171	1.332	10.052	0.373	2.172	0.273	0.177	0.475	0.348	5.064
Min	-0.546	-0.950	-0.778	0.004	0.000	0.032	0.010	0.005	0.000	-0.950	-1.892	0.000	0.054
Max	0.904	0.447	0.553	34.712	7.583	58.535	2.537	18.066	1.000	0.447	2.505	1.000	41.048
N	8310	7477	8411	7743	7457	8501	7765	8496	7765	8414	7672	8528	8441

	WCOR	CATA	TUCA	MCTA	EBITMC	CLTA	MCTU	WCSA	NIMC	LTDTA	WCNE	CHNI
Mean	0.259	0.680	0.004	0.199	1.573	0.540	2751.146	0.276	0.932	0.095	88.897	0.011
Median	0.153	0.759	0.001	0.102	0.354	0.513	127.825	0.156	0.204	0.013	6.636	0.038
SD	0.581	0.276	0.012	0.276	5.569	0.352	11802.648	0.613	4.485	0.162	880.464	0.600
Min	-0.971	0.019	0.000	0.003	-13.104	-0.210	1.692	-0.777	-16.129	0.000	-2196.489	-1.000
Max	4.341	1.000	0.085	1.696	36.220	2.179	97895.781	4.599	26.756	0.807	43230.320	1.000
N	7094	8500	7454	7766	8398	8499	7276	7193	8401	7766	6226	8329

	GDPR	INFL	BANK
Mean	2.733	2.220	0.144
Median	2.375	1.821	0.000
SD	1.403	0.967	0.352
Min	0.000	0.000	0.000
Max	11.087	10.107	1.000
N	8528	8528	8528

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents unbalanced data. Variable definitions are as follows: WCTA=Working capital/total assets, RETA=Retained earnings/total assets, EBITTA=EBIT/total assets, MCTL=Market capitalization/total liabilities, SATA=Sales/Total assets. TA=Total assets (the natural logarithm is included as predictor), TLTA=Total liabilities/total assets, CLCA=Current liabilities/current assets, OTLETA=1 if total liabilities exceed total assets, 0 otherwise, NITA=Net income/total assets, IOTL=Income from operations after depreciation/total liabilities, NEN2=1 If net income was negative for the last 2 years, 0 otherwise, CACL=Current assets/current liabilities, WCOR=Working capital/operating revenue, CATA=Current assets/total assets, TUCA=Turnover/current assets, MCTA=Market capitalization/total assets, EBITMC=EBIT/Market capitalization, CLTA=Current liabilities/total assets, MCTU=Market capitalization/turnover, WCSA=Working capital/sales, NIMC=Net income/market capitalization, LTDTA=Long term debt/total assets, WCNE=Working capital/no. of employees, CHNI=change in net income for the last to previous year, GDPR=GDP growth rate (%), INFL=Inflation rate (%), BANK=1 if bankrupt, 0 otherwise. *=in mil euro's.

Panel B: Hold-out sample (2011-2013)

	WCTA	RETA	EBITTA	MCTL	SATA	TA*	TLTA	CLCA	OTLETA	NITA	IOTL	NEN2	CACL
Mean	0.210	-0.019	0.004	1.476	1.448	7.340	0.623	1.315	0.091	-0.017	0.065	0.195	3.318
Median	0.188	0.011	0.025	0.179	1.189	2.219	0.585	0.713	0.000	0.013	0.000	0.000	1.394
SD	0.264	0.220	0.199	6.398	1.241	13.203	0.432	2.916	0.288	0.219	0.860	0.396	8.763
Min	-0.604	-1.310	-1.086	0.004	0.000	0.036	0.005	0.003	0.000	-1.310	-1.390	0.000	0.039
Max	0.860	0.417	0.457	53.217	6.850	79.856	2.988	24.131	1.000	0.417	50.110	1.000	73.868
N	7417	7467	7574	7217	6621	7661	7239	7649	7239	7577	7184	8129	7605

	WCOR	CATA	TUCA	MCTA	EBITMC	CLTA	MCTU	WCSA	NIMC	LTDTA	WCNE	CHNI
Mean	0.310	0.660	0.003	0.214	0.957	0.506	2832.139	0.332	0.455	0.104	91.575	-0.065
Median	0.156	0.734	0.001	0.101	0.198	0.455	186.119	0.161	0.111	0.026	29.252	-0.036
SD	0.879	0.284	0.009	0.323	5.242	0.382	7805.439	0.927	4.610	0.165	546.219	0.613
Min	-1.623	0.014	0.000	0.003	-19.943	0.001	2.126	-1.323	-22.529	0.000	-5695.883	-1.000
Max	6.876	1.000	0.065	2.102	30.801	2.305	39734.737	7.487	23.213	0.827	21236.341	1.000
N	6330	7657	6615	7654	7566	7654	6458	6345	7569	7242	5053	7410

	GDPR	INFL	BANK
Mean	0.677	2.314	0.140
Median	1.710	2.117	0.000
SD	2.450	0.865	0.347
Min	-8.864	-1.094	0.000
Max	8.264	6.094	1.000
N	8129	8129	8129

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents unbalanced data. Variable definitions are as follows: WCTA=Working capital/total assets, RETA=Retained earnings/total assets, EBITTA=EBIT/total assets, MCTL=Market capitalization/total liabilities, SATA=Sales/Total assets. TA=Total assets (the natural logarithm is included as predictor), TLTA=Total liabilities/total assets, CLCA=Current liabilities/current assets, OTLETA=1 if total liabilities exceed total assets, 0 otherwise, NITA=Net income/total assets, IOTL=Income from operations after depreciation/total liabilities, NEN2=1 If net income was negative for the last 2 years, 0 otherwise, CACL=Current assets/current liabilities, WCOR=Working capital/operating revenue, CATA=Current assets/total assets, TUCA=Turnover/current assets, MCTA=Market capitalization/total assets, EBITMC=EBIT/Market capitalization, CLTA=Current liabilities/total assets, MCTU=Market capitalization/turnover, WCSA=Working capital/sales, NIMC=Net income/market capitalization, LTDTA=Long term debt/total assets, WCNE=Working capital/no. of employees, CHNI=change in net income for the last to previous year, GDPR=GDP growth rate (%), INFL=Inflation rate (%), BANK=1 if bankrupt, 0 otherwise. *=in mil euro's.

4.2 Hypothesis tests

4.2.1 Hypothesis 1: Econometric model performance

Hypothesis 1a: Hypothesis 1a holds that, ceteris paribus, logistic regression-estimated models should show higher accuracy than MDA-estimated models. Below is table 5, showing AUC-statistics for each particular bankruptcy prediction model. Column A shows that the particular model has been estimated by discriminant analysis, while column B shows the particular model has been estimated by logistic regression. Each row represents a particular set of variables, i.e. the Altman (1968), Ohlson (1980), Zmijewski (1984), and each of those sets augmented by macroeconomic variables. The factor analysis does not include macroeconomic variables, as they were included in the factor analysis selected procedure, and as a result, only one of them was selected as a predictor.

Table 5: AUC scores of various bankruptcy prediction models. Out of sample tests.

Set of variables:	**A. Discriminant analysis:**	**B. Logistic regression:**	**C. Agarwal and Taffler (2007) test (B-A):**	**N**
Altman (1968)	0.761***	0.844***	6.799***	4615
	(0.010)	(0.007)		
" + macro	0.851***	0.847***	-0.500	4615
	(0.000)	(0.008)		
Ohlson (1980)	0.570***	0.869***	19.102***	5129
	(0.014)	(0.007)		
" + macro	0.575***	0.882***	19.614***	5129
	(0.014)	(0.007)		
Zmijewski (1984)	0.873***	0.866***	-0.707	5262
	(0.007)	(0,007)		
" + macro	0.879***	0.854***	-2.352***	5262
	(0.007)	(0,008)		
Factor analysis	0.868***	0.786***	-0.965	4904
	(0.007)	(0,009)		

Note: AUC statistics of various bankruptcy prediction models in a hold-out sample with observations from 2006-2007. The AUC statistic measures predictive accuracy and rangers from 0.5 (a random model) to 1 (a perfect model). Standard errors in parentheses. Univariate and bivariate statistical significance indicated as follows: ***<1%, **<5%, *<10%. Univariate significance is tested against the null hypothesis AUC=0.5.

The Agarwal and Taffler (2007) test statistic is used to compare two AUC-scores with one another (see section 3.2). Standard error estimates are computed using the Hanley and McNeil (1982) equivalence between the AUC-score and the probability that a randomly chosen diseased subject is (correctly) rated or ranked with greater suspicion than a randomly chosen non-diseased subject. The results of these tests are included in column C. A positive test statistic indicates an outperformance of the discriminant analysis model by its logistic regression counterpart.

It is clear that the logistic regression model does not outperform discriminant analysis systematically. Indeed, particularly when smaller sets of variables are used, discriminant analysis seems to outperform logistic regression (although not significantly) and the performance seems to be very dependent on the set of predictors. Therefore, we can regard hypothesis 1a as rejected. Estimating with logistic regression instead of MDA improves the AUC-score with the Altman (1968) set of variables from 0.761 to 0.844, but estimating the

Zmijewski (1984) and macroeconomic variables with logistic regression makes the AUC-score decline from 0.879 to 0.854. Furthermore, as indicated by econometric theory, discriminant analysis model accuracy is in practice highly sensitive to violation of its assumptions. The Ohlson (1980) variables in particular do not suit the MDA assumptions very well, and model accuracy is only slightly above that of a random predictor. Unreported robustness checks show that AUC scores change only very marginally when the models are estimated in 2005-2006 and evaluated in a 2007 hold-out sample, rather than, as is reported here, estimated in 2005 and evaluated in a 2006-2007 hold-out sample.

Hypothesis 1b: Hypothesis 1b holds that ceteris paribus, MDA-estimated models should carry less information content than logistic regression-estimated models. Below is table 6. Each panel represents two information content tests of a particular set of predictors, i.e. the Altman (1968), Ohlson (1980), Zmijewski (1984) and the factor analysis predictors, each particular set augmented in a second test by macroeconomic predictors, except the factor analysis set. Every column in each panel shows the coefficient estimates and relevant statistics of a separate information content test. In column three and column five are tests of bankruptcy prediction models that include the macroeconomic variables along with the set of predictors mentioned in the panel.

Table 6: Information content test for the Altman (1968), Ohlson (1980), Zmijewski (1984) and Factor analysis-selected variables.

Estimation method:	**Discriminant analysis**		**Logistic regression**	
	Panel A: Altman (1968) predictors			
Constant	0.487***	0.177	-3.480***	-3.444***
	(8.761)	(1.739)	(1199.679)	(1182.128)
Annual Default	-0,248	-0.080	-0.260	-0.014
Rate	(2.500)	(0.252)	(2.593)	(0.007)
Model	-4.543***		4.372***	
	(367.082)		(449.770)	
Model incl. macro		-5.888***		4.226***
		(446.614)		(408.486)
Hosmer-Lemeshow	10.428	38.749	143.585	112.677
(1982) test	(0.236)	(0.000)	(0.000)	(0.000)
-2LL	5196.166	4895.483	5020.871	5113.383
Cox and Snell R^2	0.061	0.085	0.075	0.068
Avg. firm-year obs.	2.13	2.13	2.13	2.13
	Panel B: Ohlson (1980) predictors			
Constant	-2.385***	-2.422**	-3.889***	-4.008***
	(227.415)	(236.848)	(1073.745)	(1056.816)
Annual Default	-0.264*	-0.258	-0.214	-0.130
Rate	(2.811)	(2.648)	(1.553)	(0.566)
Model	-0.531*		4.413***	
	(3.424)		(484.830)	
Model incl. macro		-0.458		4.472***
		(2.569)		(525.361)
Hosmer-Lemeshow	75.610	35.396	63.147	76.095
(1982) test	(0.000)	(0.000)	(0.000)	(0.000)
-2LL	5843.136	5845.161	4710.875	4576.139
Cox and Snell R^2	0.001	0.001	0.085	0.094
Avg. firm-year obs.	2.33	2.33	2.33	2.33

Panel C: Zmijewski (1984) predictors

Constant	0.308**	0.358**	-3.596***	-3.674***
	(4.491)	(5.911)	(1169.961)	(1149.987)
Annual Default	-0.282*	-0.171	-0.261	-0.083
Rate	(2.740)	(1.018)	(2.387)	(0.240)
Model	-6.585***		4.981***	
	(452.558)		(357.295)	
Model incl. macro		-6.935***		5.049***
		(447.044)		(367.641)
Hosmer-Lemeshow	85.121	96.546	220.440	133.452
(1982) test	(0.000)	(0.000)	0.000)	(0.00)
-2LL	5015.303	4973.136	5353.549	5275.558
Cox and Snell R^2	0.084	0.087	0.061	0.067
Avg. firm-year obs.	2.43	2.43	2.43	2.43

Panel D: Factor analysis predictors

Constant	0.305**	-2.623***
	(4.760)	(1104.565)
Annual Default	-0.042	-0.215
Rate	(0.065)	(2.020)
Model	-6.910***	19.491***
	(428.462)	(23.426)
Hosmer-Lemeshow	18.373	1190.969
(1982) test	(0.019)	(0.000)
-2LL	4802.391	6004.344
Cox and Snell R^2	0.095	0.006
Avg. firm-year obs.	2.29	2.29

Note: In model information content tests, I estimate a pooled logit regression with a constant term, the annual default rate and the score from one of the bankruptcy prediction models. Discriminant analysis models should have a negative coefficient, whereas logistic regression models should have a positive coefficient. A good model should itself show statistical significance, but it should also subsume the constant and the annual default rate, indicating that these carry no information beyond the bankruptcy prediction model. This table shows bankruptcy prediction models that are estimated using observations from 2005. The tests are conducted using observations from a 2006-2007 hold-out sample. Out-of-sample results are reported. Chi-squared statistics are reported in parentheses and adjusted according to Shumway (2001), and statistical significance is indicated with stars. With 1 degree of freedom, critical Wald values at 1%, 5% and 10% level are 6.63, 3.84 and 2.71 respectively.

In all cases, except for the discriminant analysis models (due to its specification, elaborated on in appendix F) a positive coefficient indicates that the relevant variable is associated with an increase in the likelihood of bankruptcy. In the discriminant analysis cases, a negative coefficient indicates the aforementioned. For all models, the coefficient on the model score has the expected sign and, bar some bad performing models, is always significant at the 1% level. This indicates that the model has explanatory power beyond the baseline annual default rate. The variables used in the model and the econometric method for combining them into a single bankruptcy score are significantly associated with actual bankruptcies and thus useful in predicting bankruptcies. This is the case in by far most models, and is in accordance with the findings of Hillegeist et al. (2004), Agarwal and Taffler (2007) and Wu, Gaunt and Gray (2010).

Hosmer and Lemeshow (1982) tests assess model goodness-of-fit, which helps decide whether a model is correctly specified. The test sorts all probabilities from low to high in 10 separate groups, where the expected number of events is just the sum of the predicted probabilities over the objects in the group. And the expected number of non-events is the group size minus the expected number of events. A chi-squared test is then conducted to assess the difference between expected and observed counts. A higher value means a greater difference between the model predictions and the observed data, and thus a bad goodness of fit. For example, the discriminant analysis model with the Altman (1968) variables shows a chi-squared value of 10.428, which is not significantly different from zero, indicating there is no difference between observed and expected events and thus indicates a well-calibrated model. A formal definition of the Hosmer-Lemeshow (1982) test is given in appendix H. In most cases, this statistic is fairly high, which means a bad goodness-of-fit, but these results are also present in e.g. Agarwal and Taffler (2007).

An indication of the likelihood of the data occurring, given the parameter estimates, is given by the -2 log likelihood (-2LL) statistic. The -2LL statistic show then the deviance that is still present in the optimal solution. A larger -2LL statistic indicates greater deviance between the data and the model. In all models, -2LL is between 4000 and 6000, which is consistent with Wu, Gautn and Gray (2010) but systematically larger than Agarwal and Taffler (2007). The models that have a large -2LL statistic (e.g. the MDA-estimated Ohlson (1980) and macroeconomic predictors) also perform badly when considering other statistics.

Cox and Snell's R-squared measures a model's predictive power. A high Cox and Snell R-squared implies a huge difference between the likelihood of the data given a model with no predictors and the likelihood of the data given the model in question. Cox and Snell R-squared are unique in the sense that they correct for sample size. It is somewhat stricter than the also often-used Nagelkerke R-squared, and is therefore a more conservative metric of predictive power. The R-squared statistics are generally around or below 0.100 but really bad models stand out, as for example the R-squareds of the MDA-estimated models with the Ohlson (1980) and macroeconomic variables, with an R-squared <1, indicating no difference between a model without variables and the model with this particular bankruptcy prediction model. This is consistent with the low value of the Wald-statistics (3.424 and 2.569 respectively), indicating large standard errors.

The results indicate that the discriminant analysis estimation does not obviously carry less information about bankruptcy then does the logistic regression estimation. This seems true for all sets of variables, except for the Ohlson (1980) variables (where MDA is clearly outperformed), even the sets that performed worse on model accuracy. For example, in the case of the Zmijewski (1984) variables, the R-squared statistics (8.4% vs. 6.1%), and model Wald-statistics (452 vs. 357) are a fair amount larger when estimated using discriminant analysis than when using logistic regression. This implies a smaller standard error in the MDA-estimated model, and slightly greater predictive power.

Model coefficient magnitude, R-squared statistics and Wald-statistics levels are generally comparable to Agarwal and Taffler (2007) and Wu, Gaunt and Gray (2010). Furthermore, the log-likelihood statistics seem to be slightly smaller when predictor coefficients are estimated by logistic regression, but this is not always so (e.g. the Altman (1968) and macroeconomic variables estimated with discriminant analysis do better than its logistic regression-estimated counterpart). Unreported robustness checks show that a re-estimation of the models in 2005 and 2006, and a hold-out sample in 2007 has only a marginal effect on results of the information content tests. Therefore, in the absence of a clear superiority of the logistic regression-estimated models, we can regard hypothesis 1b as rejected.

4.2.2 Hypothesis 2: Stationarity in the predictor variables

Hypothesis 2a: Below is table 7, which shows AUC statistics to test hypothesis 2a. Hypothesis 2a holds that bankruptcy prediction models do not retain their accuracy over time. First, the columns under A show each particular set of models that is estimated using discriminant analysis, while columns under B show models estimated using logistic regression. Every result under i) represents an in-sample AUC-score, whereas everything under ii) represents an out-of-sample AUC score. The same Agarwal and Taffler (2007) test is again used to indicate whether a difference is statistically significant. Each row represents a different set of variables contained in the bankruptcy prediction model, augmented with the macroeconomic variables in the Altman (1968), Ohlson (1980) and Zmijewski (1984) cases.

Table 7: AUC scores of various bankruptcy prediction models

	A. Discriminant analysis:				**B. Logistic regression:**			
Set of variables:	**i. 2005-2007 (in sample)**	**ii. 2011-2013 (out-of-sample)**	**Agarwal and Taffler (2007) test (ii-i):**	**N**	**i. 2005-2007 (in sample)**	**ii. 2011-2013 (out-of-sample)**	**Agarwal and Taffler (2007) test (ii-i):**	**N**
Altman (1968)	0.842*** (0.006)	0.818*** (0.008)	-2.400***	6132	0.853*** (0.006)	0.832*** (0.008)	-2.100**	6132
“ + macro	0.856*** (0.006)	0.770*** (0.009)	-13.775***	6132	0.862*** (0.006)	0.756*** (0.009)	-9.799***	6132
Ohlson (1980)	0.735*** (0.009)	0.739*** (0.010)	0.297	6818	0.883*** (0.005)	0.883*** (0.006)	0.000	6818
“ + macro	0.879*** (0.006)	0.855*** (0.007)	-2.603***	6818	0.896*** (0.005)	0.857*** (0.000)	-7.800***	6818
Zmijewski (1984)	0.857*** (0.006)	0.862*** (0.007)	0.542	7141	0.853*** (0.006)	0.858*** (0.007)	0.542	7141
“ + macro	0.870*** (0.006)	0.827*** (0.007)	-4.664***	7411	0.868*** (0.006)	0.786*** (0.008)	-8.200***	7141
Factor analysis	0.872*** (0.006)	0.796*** (0.009)	-7.026***	6246	0.817*** (0.007)	0.731*** (0.009)	-7.542***	6246

Note: AUC statistics of various bankruptcy prediction models in (i) the sample in which the model has been estimated, and (ii) in a hold-out sample with observations from 2011-2013. Results are sorted by the method which is used to estimate the particular bankruptcy prediction model. The AUC statistic measures predictive accuracy and rangers from 0.5 (a random model) to 1 (a perfect model). Standard errors in parentheses. Univariate and bivariate statistical significance indicated as follows: ***<1%, **<5%, *<10%. Univariate significance is tested against the null hypothesis AUC=0.5.

In section 3.3 it is mentioned that my particular sample features some countries (i.e. Spain, France and Belgium) proportionally more than others. This might lead to a bias towards the characteristics of firms in these countries. Therefore, a robustness check is conducted to evaluate the impact of this bias. The results are shown in table 8. For convenience, only the result of the Ohlson (1980) and macroeconomic predictors is reported. In particular, for each model, the in-sample and out-of-sample AUC scores are computed for each model when applied only on (i) Belgium, France and Spain, and (ii) when applied only on all the other countries.

Table 8: Robustness test for the Ohlson (1980) + macroeconomic variables								
	A. Discriminant Analysis				**B. Logistic regression**			
Countries:	**i. 2005-2007 (in sample)**	**ii. 2011-2013 (out-of-sample)**	**Agarwal and Taffler (2007) test (ii-i):**	**N**	**i. 2005-2007 (in sample)**	**ii. 2011-2013 (out-of-sample)**	**Agarwal and Taffler (2007) test (ii-i):**	**N**
Belgium, France, Spain only	0.881*** (0.006)	0.846*** (0.012)	-2.608***	5600	0.903*** (0.005)	0.835*** (0.012)	-5.230***	5600
All others	0.878*** (0.011)	0.811** (0.011)	-4.603***	2592	0.885*** (0.011)	0.830*** (0.010)	-3.699***	2592

Note: AUC statistics of various bankruptcy prediction models in (i) the sample in which the model has been estimated, and (ii) in a hold-out sample with observations from 2011-2013. The first row shows and compares the AUC scores only on observations from Belgium, France and Spain. The second row shows and compares the AUC score for all other countries. Results are sorted by the method which is used to estimate the particular bankruptcy prediction model. The AUC statistic measures predictive accuracy and rangers from 0.5 (a random model) to 1 (a perfect model). Standard errors in parentheses. Univariate and bivariate statistical significance indicated as follows: ***<1%, **<5%, *<10%. Univariate significance is tested against the null hypothesis AUC=05.

It seems that models are generally unable to hold robustness over time, often experiencing a significant decrease in the AUC-statistic. We must therefore accept hypothesis 2a in that models lose a significant amount of predictive accuracy when applied in a different environment. The inclusion of macroeconomic variables as control could not prevent or offset this loss. That implies that changes in firms' financial characteristics are not correlated with macroeconomic events, but rather follow their own idiosyncratic pattern, or, alternatively, that the effects of the macroeconomy are already incorporate into firms' characteristics. For example, the MDA-estimated model with the Altman (1968) and macroeconomic variables AUC-score declines from 0.856 to 0.770, whereas the MDA-estimated model with just the Altman (1968) variables shows a milder decline (from 0.842 to 0.818). These results are consistent with results by Mensah (1984), Zavgren (1985), Grice and Ingram (2001) and Grice and Dugan (2003), although contradictory findings have been reported by e.g. Reisz and Perlich (2007) and Agarwal and Taffler (2007). However, some particular models do appear to retain a reasonably good predictive accuracy.

It is especially noteworthy that the discriminant analysis-estimated model with the Ohlson (1980) and macroeconomic variables performs well (in-sample AUC of 0.879, out-of-sample AUC of 0,855), given the econometric issues involved. The Ohlson (1980) variables included two dummy variables and one dummy-like variable, which is inconsistent with the univariate normality requirement of discriminant analysis. It even outperforms the Altman (1968) variables, which have been selected to be consistent with discriminant analysis assumptions.

Concerning the bias towards median countries, mentioned in section 3.3, it seems that discriminant analysis is somewhat sensitive to the overpresence of Belgium, Spain and France in the sample, albeit only slightly. The model's in-sample accuracy significantly improves when it is applied on exclusively Spain, Belgium and France, compared to the accuracy for the entire sample as a whole.

For out-of-sample accuracy, this does not at all seem to be the case. In case of the logistic regression-estimated Ohlson (1980) and macroeconomic predictors, the out-of-sample AUC score is 0.857 when applied on the entire sample, versus 0.835 when applied exclusively on France, Spain and Belgium. For example, the MDA-estimated out-of-sample AUC-score of the Ohlson (1980) and macroeconomic predictors is 0.855 when applied on all countries, vs. 0.846 when applied on France, Spain and Belgium exclusively. Logistic regression seems to be somewhat less sensitive to this bias, as predictive accuracy declines less strongly if applied on a sample of all countries other than France, Belgium and Spain. From other, unreported results, it seems that this generalizes. It seems then, that the models cannot retain their accuracy over time, but do not lose predictive accuracy when applied to other countries other than the ones overrepresented in the sample.

Hypothesis 2b: Below is table 9, which shows information content tests of various models to test hypothesis 2b. Hypothesis 2b holds that models will not retain their information content over time. Except for column one, each column represents a different information content test. Column two and three show tests of models that have been estimated with MDA, while columns four and five have been estimated with logistic regression.

The coefficient belonging to 'model' is the one that belongs to the bankruptcy prediction model. An information content test is thus conducted on a bankruptcy prediction model containing particular sets of predictors, which in turn have been estimated with either MDA or logistic regression. Furthermore, column three and column five show tests of models that include the two macroeconomic predictors (GDP growth and inflation rate), in addition to the standard set of predictors mentioned in each panel. Panel D does not include macroeconomic predictors because the set of predictors was determined by factor analysis.

Table 9: Information content tests. 2011-2013 out-of-sample results are reported.

Estimation method:	**Discriminant analysis**		**Logistic regression**	
	Panel A: Altman (1968) predictors			
Constant	0.765***	0.396*	-3.179***	-3.143***
	(11.111)	(3.128)	(195.560)	(194.553)
Annual Default	-0.289	-0.301*	-0.299*	-0.359**
Rate	(2.616)	(2.919)	(2.728)	(4.103)
Model	-6.564***		4.555***	
	(861.932)		(885.820)	
Model incl. macro		-6.307***		3.929***
		(685.541)		(740.070)
Hosmer-Lemeshow	34.672	49.759	96.368	36.065
(1982) test	(0.000)	(0.000)	(0.000)	(0.000)
-2LL	5186.743	5460.546	5228.895	5521.609
Cox and Snell R^2	0.113	0.091	0.110	0.086
Avg. firm-year obs.	1.36	1.36	1.36	1.36

Panel B: Ohlson (1980) predictors

Constant	2.238***	0.448*	-3.967***	-2.867***
	(106.947)	(3.481)	(229.629)	(139.735)
Annual Default	0.400**	-0.337*	-0.061	-0.880***
Rate	(5.086)	(3.080)	(0.090)	(20.004)
Model	-5.150***		5.240***	
	(236.045)		(1032.597)	
Model incl. macro		-7.683***		4.686***
		(726.765)		(943.368)
Hosmer-Lemeshow	222.196	47.446	58.830	46.579
(1982) test	(0.000)	(0.000)	(0.000)	(0.000)
-2LL	6138.396	4944.265	4696.979	4970.685
Cox and Snell R^2	0.034	0.123	0.141	0.121
Avg. firm-year obs.	1.51	1.51	1.51	1.51

Panel C: Zmijewski (1984) predictors

Constant	0.696***	0.724***	-3.265***	-2.963***
	(8.115)	(8.860)	(183.712)	(157.701)
Annual Default	-0.258	-0.477**	-0.252	-0.495***
Rate	(1.800)	(6.264)	(1.730)	(6.949)
Model	-6.871***		4.656***	
	(863.306)		(781.012)	
Model incl. macro		-6.788***		3.942***
		(756.112)		(689.327)
Hosmer-Lemeshow	81.436	70.191	221.121	98.385
(1982) test	(0.000)	(0.000)	(0.000)	(0.000)
-2LL	5286.087	5493.738	5583.347	5837.969
Cox and Snell R^2	0.117	0.103	0.097	0.079
Avg. firm-year obs.	1.58	1.58	1.58	1.58

Panel D: Factor analysis predictors

Constant	-0.159	-1.990***
	(0.523)	(96.396)
Annual Default	0.003	-0.402**
Rate	(0.000)	(5.857)
Model	-6.739***	5.042***
	(624.611)	(154.738)
Hosmer-Lemeshow	116.303	287.117
(1982) test	(0.000)	(0.000)
-2LL	5426.656	6241.196
Cox and Snell R^2	0.096	0.028
Avg. firm-year obs.	1.38	1.38

Note: In model information content tests, I estimate a pooled logit regression with a constant term, the annual default rate and the score from one of the bankruptcy prediction models. Discriminant analysis models should have a negative coefficient, whereas logistic regression models should have a positive coefficient. A good model should itself show statistical significance, but it should also subsume the constant and the annual default rate, indicating that these carry no information beyond the bankruptcy prediction model. This table shows bankruptcy prediction models that are estimated using observations from 2005-2007. The tests are conducted using observations from a 2011-2013 hold-out sample. Out-of-sample results are reported. Chi-squared statistics are reported in parentheses and adjusted according to Shumway (2001), and statistical significance is indicated with stars. With 1 degree of freedom, critical Wald values at 1%, 5% and 10% level are 6.63, 3.84 and 2.71 respectively.

Hypothesis 2b held that models do not retain their information content when applied in a hold-out sample. For all models, the coefficient on the model score has the expected sign (i.e. negative for MDA-estimated models and positive for logistic regression-estimated models) and always shows significance at the 1% level. This indicates that the model has explanatory power beyond the baseline annual default rate. Therefore, hypothesis 2b is rejected. The variables used in the model and the econometric method for combining them into a single bankruptcy score are useful in predicting bankruptcies. The annual default rate however, does often assume significance, as does the constant term. This means that there is information about bankruptcy not captured by the particular bankruptcy prediction model.

Unfortunately, the Hosmer and Lemeshow (1982) statistic is significant in all cases, so no model estimated in hypothesis 2 shows sufficient similarity between the expectation given by the model, and the actual data for the null-hypothesis to be accepted. -2 log likelihood (-2LL) statistics gives an indication of the likelihood of the data occurring, given the parameter estimates. It is remarkable that in the case of the Ohlson (1980) variables, the -2LL statistic is much smaller (indicating more similarity between the model and the data) when using logistic regression (4697) than when using discriminant analysis (6138), but this difference disappears when the macroeconomic predictors are added. In this case, the addition of macroeconomic variables does seem to add relevant information, though the results of the other models indicate that this does not generalize across all sets of predictors. A possible explanation is that the Altman (1968), Zmijewski (1984) and factor analysis predictors already subsumed the effect of the macroeconomy into their firm characteristics, but this is not so for the Ohlson (1980) indicators.

In general, the Cox and Snell's R-squared seems to decline when macroeconomic predictors are included in the models, except for the Ohlson (1980) predictors: when estimated by discriminant analysis, the R-squared increases from 0.034 to 0.123 when macroeconomic predictors are included. However, it declines when estimated by logistic regression from 0.141 to 0.121.

Furthermore, for the reasons cited in hypothesis 2a, I conduct a short robustness test. This is featured in table 10. It shows results similar to the robustness test for hypothesis 2a. While the -2LL ratio of the models worsen (e.g. from 1865 to 2911 in the MDA-estimated Ohlson (1980) + macroeconomic variables case, indicating a worsening likelihood of the actual data given the model), the Wald-statistic actually increases when applied to all other countries, so the effects are ambiguous. Anyhow, both in (i) and in (ii), model coefficients assume a 1% level of significance, so the model carries useful information regarding the prediction of bankruptcy in both sample partitions. Unreported results confirm that this generalizes to all sets of predictor variables present in table 9.

Table 10: Information content robustness test

	Discriminant analysis		Logistic regression	
Constant	0.470	0.192	-2.833***	-2.527***
	(1.282)	(0.416)	(52.666)	(64.773)
Annual Default Rate	-0.882***	0.014	-1.329***	-0.481**
	(7.291)	(0.003)	(16.275)	(3.781)
i. Belgium, France and Spain	-7.416***		4.564***	
	(262.989)		(281.770)	
ii. All other countries		-6.563***		3.892***
		(287.937)		(389.411)
Hosmer-Lemeshow (1982) test	6.266	192.793	35.876	29.574
	(0.617)	(0.000)	(0.000)	(0.000)
-2LL	1865.087	2911.427	1987.765	2807.358
Cox and Snell R^2	0.057	0.146	0.043	0.169
Avg. firm-year obs.	1.50	1.52	1.50	1.52

Note: In model information content tests, I estimate a pooled logit regression with a constant term, the annual default rate and the score from one of the bankruptcy prediction models. Discriminant analysis models should have a negative coefficient, whereas logistic regression models should have a positive coefficient. A good model should itself show statistical significance, but it should also subsume the constant and the annual default rate, indicating that these carry no information beyond the bankruptcy prediction model. For illustrative purposes, this table shows only the result of the model with the Ohlson (1980) + macroeconomic variables. The models are estimated using observations from 2005-2007. The tests are conducted using observations from a 2011-2013 hold-out sample. The table shows results of the test conducted in obersvations from (i) Belgium, France and Spain and (ii) all other countries. Out-of-sample results are reported. Chi-squared statistics are reported in parentheses and adjusted according to Shumway (2001), and statistical significance is indicated with stars. With 1 degree of freedom, critical Wald values at 1%, 5% and 10% level are 6.63, 3.84 and 2.71 respectively.

4.2.3 Hypothesis 3: Performance relative to simple benchmark

Hypothesis 3a: Hypothesis 3 holds that an elaborate bankruptcy prediction model such as the ones in this study should outperform other, simple alternative models in predictive accuracy. Table 11 shows AUC-scores of some of the best performing models in out-of-sample forecasting, and juxtaposes them against the alternative models. Only out-of-sample forecasts are reported, and only several models are reported, some of the best performing models in the previous tests.

Table 11: AUC scores of bankruptcy prediction models and competing alternatives

Panel A: Main bankruptcy prediction models:	**AUC:**	**N**
MDA, Ohlson (1980) and macroecnomic variables	0.855*** (0.007)	6818
MDA, Zmijewski (1984) variables:	0.862*** (0.007)	7141
MDA, Zmijewski (1984) and macroeconomic variables	0.827*** (0.007)	7141
Logistic regression, Ohlson (1980) and macroecnomic variables	0.857*** (0.000)	6818
Logistic regression, Zmijewski (1984) variables:	0.858*** (0.007)	7141
Logistic regression, Zmijewski (1984) and macroeconomic variables	0.786*** (0.008)	7141
Panel B: Alternative prediction models:		
PBT<1 indicator	0.669*** (0.009)	8129
All firms non-failure	0.500 (0.009)	8129
Proxy market cap[a]	0.505 (0,010)	8129
Sales growth indicator	0.505 (0.011)	5705

Note: AUC statistics of various bankruptcy prediction models and some alternative benchmarks in a hold-out sample with observations from 2011-2013. The AUC statistic measures predictive accuracy and rangers from 0.5 (a random model) to 1 (a perfect model). Standard errors in parentheses. Univariate and bivariate statistical significance indicated as follows: ***<1%, **<5%, *<10%. Univariate significance is tested against the null hypothesis AUC=0.5. [a]=Only the best performing proxy market cap is reported, since they performed similarly.

As is clear from table 11, even the worst bankruptcy prediction model outperforms the best simple predictor by a fairly large margin. The logistic regression-estimated model with the Zmijewski and macroeconomic variables performs the worst. Its AUC score is 0.786, which is still a substantial amount above the PBT<1 indicator, which has an AUC score of 0.669. The alternative indicators clearly do not have the capacity to forecast bankruptcy on their own. It seems that in order to construct a good bankruptcy prediction model, a simple metric will not do. A good and elaborate functional specification seems to be a requirement for any bankruptcy prediction model aiming to be accurate. Therefore, we can regard hypothesis 3a as verified.

Hypothesis 3b: Hypothesis 3b holds that models will retain their information content when alternative, simpler bankruptcy prediction models are also included in the information content test. In table 12, I present results from several particular bankruptcy prediction models. Panel A shows three models with particular sets of predictors, estimated by MDA, and Panel B shows three models with particular sets of predictors, estimated by logistic regression. Each column represents a different bankruptcy prediction model. The rows show the particular coefficient belonging to each variate. Other model results are similar, but left out.

One can see that in general, the addition of macroeconomic variables again surprisingly decreases the explanatory power of the model. The discriminant analysis-estimated model with the Zmijewski (1984) variables sees its R-squared decline from 0.148 to 0.127. A similar trend is observed across all models, with the exception of the Altman (1968) variables when estimated by either discriminant analysis or logistic regression. Again, this is consistent with the results obtained in hypothesis 1 and 2, indicating that either the impact of macroeconomic conditions is already incorporated into firm characteristics, or that there does not exist country-specific bankruptcy risk.

Most Hosmer and Lemeshow (1982) test results show that the model in question does not show a great fit – with the exception of the logistic regression-estimated Ohlson (1980) variables, both with and without macroeconomic predictors. We can conclude that there is a significant difference between observed and expected fitted risk profiles for most models, though this is contingent on the particular estimation method and selection of variables.

-2 log likelihood (-2LL) statistics gives an indication of the likelihood of the data occurring, given the parameter estimates. The smaller the statistic, the greater is the likelihood of the model. We can observe that the models that have the highest Cox and Snell's R-squared also have the lowest -2LL statistic. The statistics are in the same order of magnitude as those in the tests of Hillegeist et al. (2004) and Agarwal and Taffler (2007).

Furthermore, for all models, the coefficient on the model score has the expected sign (i.e. positive for MDA-estimated models and negative for logistic regression-estimated models) and always shows significance at the 1% level. All models retain their statistical significance when the alternative prediction models are included. In several unreported estimates I replace the proxy market cap by its logarithm-transformed form, but results are very similar. Therefore, we can regard hypothesis 3b as verified.

In many tests however, some alternative predictors also assume statistical significance, which means that they carry information about bankruptcy uncaptured by the bankruptcy prediction model that is tested. Simple predictors such as market capitalization and a profit before tax smaller than 0 carry information about bankruptcy, in addition to the information carried in the model. Many times, the coefficient on the model is much larger than that on one of the alternative predictors, even though they are both measured on the same scale, indicating the model carries relatively more bankruptcy-related information.

Table 12: Information content tests with alternative prediction models

Panel A: Selection of MDA-estimated models:

Constant	-0.395	0.318	-0.192
	(1.956)	(1.190)	(0.448)
Annual Default Rate	-0.336*	-0.254	-0.459**
	(2.94)	(1.641)	(5.637)
MDA, Ohlson (1980) and macroecnomic variables	-7.067***		
	(395.111)		
MDA, Zmijewski (1984) variables		-7.213***	
		(513.659)	
MDA, Zmijewski (1984) and macroeconomic variables			-6.236***
			(378.598)
PBT<1	0.311***	0.145	0.507***
	(6.662)	(1.329)	(18.033)
Sales growth	0.000	0.000	0.000
	(0.024)	(0.393)	(0.316)
Proxy market cap	0.000***	0.000***	0.000***
	(203.856)	(186.478)	(173.928)
Hosmer-Lemeshow (1982) test	46.411	26.566	18.166
	(0.000)	(0.001)	(0.020)
-2LL	3766.866	3739.487	3967.375
Cox and Snell R^2	0.137	0.148	0.127
Avg. firm-year obs.	1.16	1.18	1.18

Panel B: Selection of logistic regression estimated models

Constant	-3.731***	-4.011***	-3.684***
	(207.261)	(245.586)	(224.386)
Annual Default Rate	-0.704***	-0.231	-0.462**
	(12.329)	(1.381)	(5.953)
Logistic regression, Ohlson (1980) and macroecnomic variables	3.899***		
	(476.694)		
Logistic regression, Zmijewski (1984) variables:		4.188***	
		(417.582)	
Logistic regression, Zmijewski (1984) and macroeconomic variables			2.881***
			(264.182)
PBT<0	0.835***	0.866***	1.163***
	(55.677)	(59.745)	(117.154)
Sales growth	0.000	0.000	0.000
	(0.019)	(0.518)	(0.322)
Proxy market cap	0.000***	0.000***	0.000***
	(199.152)	(174.938)	(164.764)
Hosmer-Lemeshow (1982) test	12.328	61.783	54.610
	(0.137)	(0.000)	(0.000)
-2LL	3747.757	3915.990	4163.061
Cox and Snell R^2	0.140	0.132	0.109
Avg. firm-year obs.	1.17	1.19	1.19

Note: In model information content tests, I estimate a pooled logit regression with a constant term, the annual default rate, the score from one of the bankruptcy prediction models, and a selection of alternative bankruptcy prediction models. Discriminant analysis models should have a negative coefficient, whereas logistic regression models, and all alternative models, should have a positive coefficient. A good model should itself show statistical significance, but it should also subsume the constant and the annual default rate, indicating that these carry no information beyond the bankruptcy prediction model. This table shows bankruptcy prediction models that are estimated using observations from 2005-2007. The tests are conducted using observations from a 2011-2013 hold-out sample. Out-of-sample results are reported. Chi-squared statistics are reported in parentheses and adjusted according to Shumway (2001), and statistical significance is indicated with stars. With 1 degree of freedom, critical Wald values at 1%, 5% and 10% level are 6.63, 3.84 and 2.71 respectively. Only the particular estimation with the best performing proxy market capitalization is reported. The alternative predictor 'all firms non failure' is omitted, because it is a constant.

4.2.4 Hypothesis 4: Robustness to alternative sets of predictors

Hypothesis 4a: Hypothesis 4a holds that model accuracy is robust to a change in predictors, as long as certain dimensions are reflected. The Zmijewski (1984) predictors reflect a profitability, solvability and liquidity dimension, while the Altman (1968) predictors reflect profitability, solvability, liquidity and activity dimensions. Two alternative models are composed of variables that also reflect these dimensions, and these are then compared to the previously reported performance of the Altman (1968) and Zmijewski (1984) predictors. Table 13 includes AUC statistics for Altman (1968), Zmijewski (1984) for reference's sake, and presents two alternative sets of variables, and each of those set of predictors augmented with the macroeconomic variables. In column A, the bankruptcy prediction model is estimated using discriminant analysis, in column B using logistic regression. Out-of-sample forecasts are reported.

Table 13: AUC scores of bankruptcy models with nonstandard predictors

Set of variables:	**A. Discriminant analysis:**		**B. Logistic regression:**	
	AUC:	**N:**	**AUC:**	**N:**
Altman (1968)	0.818*** (0.008)	6132	0.832*** (0.008)	6132
" + macro	0.770*** (0.009)	6132	0.756*** (0.009)	6132
Zmijewski (1984)	0.862*** (0.007)	7141	0.858*** (0.007)	7141
" + macro	0.827*** (0.007)	7141	0.786*** (0.008)	7141
B variables	0.828*** (0.008)	6199	0.831*** (0.008)	6199
" + macro	0.804*** (0.008)	6199	0.775*** (0.009)	6199
C variables	0.761*** (0.011)	4203	0.775*** (0.010)	4203
" + macro	0.674*** (0.011)	4203	0.680*** (0.011)	4203

Note: AUC statistics of various bankruptcy prediction models and some alternative benchmarks in a hold-out sample with observations from 2011-2013. The AUC statistic measures predictive accuracy and rangers from 0.5 (a random model) to 1 (a perfect model). Standard errors in parentheses. Univariate and bivariate statistical significance indicated as follows: ***<1%, **<5%, *<10%. Univariate significance is tested against the null hypothesis AUC=0.5. B variables are Working capital/operating revenue, EBIT/total equity, Current liabilities/total assets, Total equity/turnover. C variables are Working capital/sales, Net income/equity, Long term debt/total assets, Working capital/No. of employees.

From Table 13 we can deduce that model accuracy is not generally robust to changes in the set of predictors, although the results are ambiguous. For instance, the B variables score a higher AUC than the Altman (1968) variables when estimated by discriminant analysis (0.828 versus 0.818 respectively), but not when estimated by logistic regression (0.831 versus 0.832 respectively). Similarly, the C variables are outperformed by the Altman (1968) plus macroeconomic variables when estimated by discriminant analysis, but outperform them when estimated by logistic regression.

We cannot conclude that a change to alternative predictors, when it should in theory reflect the same dimension, has a significant negative effect on model accuracy. It seems that the variables used by Altman (1968) and Zmijewski (1984) have no special significance, and slight alterations or alternative ratios that proxy for liquidity, profitability, activity and solvability can be suitable replacements for these predictors, although the results do not show any consistency. Hypothesis 4a is therefore tentatively accepted. A second remarkable result is that again, macroeconomic predictors worsen rather than improve model accuracy. This means that there is no firm-invariant bankruptcy risk in a particular country, i.e., country-specific bankruptcy risk might not exist or the impact of the macroeconomy might already be subsumed into firm-level characteristics, and confirms my earlier findings in section 4.2.2 and 4.2.3.

Hypothesis 4b: Hypothesis 4b holds that model information content is robust to a change in predictors, as long as a liquidity, activity, profitability and solvability dimension are present. Below is table 14, which shows tests of bankruptcy prediction models estimated with 2 particular sets of variables that reflect these dimensions. Panel A shows these models estimated by MDA, panel B by logistic regression. Each column represents a different information content test, where column three and five are tests on models that contain the original sets of variables augmented with two macroeconomic variables (GDP growth rate and inflation rate).

From the results it appears that all models assume statistical significance at the 1% level and show the expected sign. Therefore, we can regard hypothesis 4b as verified.

Most models, including the models that performed well on the AUC-test, show statistical significance on the Hosmer and Lemeshow (1982) tests, indicating a significant deviance between observed and expected values risk deciles. This is not surprising, since in section 4.2.2 and 4.3.3, many models also failed the Hosmer-Lemeshow (1982) test. The alternative sets of predictors do no better, no matter whether the model is estimated using MDA or logistic regression.

The models do not appear to have larger -2LL statistics than do the models with the Altman (1968), Ohlson (1980), Zmijewski (1984) or FA variables. The C variables, that clearly showed lower AUC-statistics compared to the B variables, do in general have lower -2LL statistics. For instance, the B variables set of predictors has an AUC of 0.828 when estimated by MDA, and 0.831 when estimated by logistic regression, versus a score of 0.761 0.775 for the C variables when estimated by MDA and logistic regression respectively. However, the -2LL statistics of the B variables are 5314 and 5416 when estimated by MDA and logistic regression respectively, versus 4572 and 4445 for the C variables, indicating the data is more likely given the model containing the C variables than containing the B variables.

Table 14: Information content tests on models with nonstandard predictors.

Panel A: Bankruptcy prediction models estimated using discriminant analysis

Constant	5.075***	4.987***	1.275	2.133***
	(391.418)	(372.384)	(21.847)	(44.256)
Annual Default	0.402**	0.615***	-0.451**	-1.589***
Rate	(5.181)	(12.256)	(5.206)	(60.697)
B variables	-5.738***			
	(772.005)			
B variables +		-5.737***		
macro		(643.512)		
C variables			-6.362***	
			(396.207)	
C variables +				-5.165***
macro				(250.492)
Hosmer-Lemeshow	14.722	27.615	256.576	143.343
(1982) test	(0.065)	(0.001)	(0.000)	(0.000)
-2LL	5314.970	5467.903	4572.700	4843.079
Cox and Snell R2	0.105	0.093	0.091	0.058
Avg. firm-year obs.	1.62	1.62	1.54	1.54

Panel B: Bankruptcy prediction models estimated using logistic regression

Constant	-2.958***	-2.574***	-2.877***	-1.292***
	(148.756)	(119.132)	(126.418)	(29.676)
Annual Default	-0.341*	-0.687***	-0.466**	-1.568***
Rate	(3.089)	(12.741)	(5.224)	(57.522)
B variables	4.373***			
	(677.352)			
B variables +		3.945***		
macro		(596.861)		
C variables			4.950***	
			(459.262)	
C variables +				4.050***
macro				(362.018)
Hosmer-Lemeshow	136.129	59.668	77.491	108.264
(1982) test	(0.000)	(0.000)	(0.000)	(0.000)
-2LL	5416.449	5611.931	4445.651	4696.008
Cox and Snell R2	0.097	0.081	0.106	0.076
Avg. firm-year obs.	1.62	1.62	1.54	1.54

Note: In model information content tests, I estimate a pooled logit regression with a constant term, the annual default rate and the score from one of the bankruptcy prediction models. Discriminant analysis models should have a negative coefficient, whereas logistic regression models should have a positive coefficient. A good model should itself show statistical significance, but it should also subsume the constant and the annual default rate, indicating that these carry no information beyond the bankruptcy prediction model. This table shows bankruptcy prediction models that are estimated using observations from 2005-2007. The tests are conducted using observations from a 2011-2013 hold-out sample. Out-of-sample results are reported. Chi-squared statistics are reported in parentheses and adjusted according to Shumway (2001), and statistical significance is indicated with stars. With 1 degree of freedom, critical Wald values at 1%, 5% and 10% level are 6.63, 3.84 and 2.71 respectively. B variables are Working capital/operating revenue, EBIT/total equity, Current liabilities/total assets, Total equity/turnover. C variables are Working capital/sales, Net income/equity, Long term debt/total assets, Working capital/No. of employees.

In general, the Cox and Snell R-squared statistics are lower than those reported in sections 4.2.1, 4.2.2 and 4.3.3, indicating worse predictive power. The results are consistenct with the previously reported results, as there is no significant change in magnitude between the R-squared statistics of these models and the models containing the Altman (1968), Ohlson (1980), Zmijewski (1984) or FA variables, as witnessed in sections 4.2.1, 4.2.2 and 4.3.3.

Finally, it is striking that in nearly all estimations both the constant and the annual default rate also show statistical significance. This means that there is a significant portion of information content regarding bankruptcy that is not carried with the bankruptcy prediction model. In the information content tests of hypothesis 2, this was less so. Furthermore, in accordance with the results from hypothesis 2a, 3a and 4a, which showed that macroeconomic variables decrease rather than increase model accuracy, the Wald-statistic on models with macroeconomic predictors included is generally smaller than the Wald-statistic on models without the macroeconomic predictors.

5. Conclusion and discussion

5.1 Research question

The research question that was addressed in this study is 'what is the accuracy and information content of various statistical techniques predicting bankruptcy for a sample of listed non-financial firms in the European Union?' I attempted to answer this question through posing four subquestions, which then lead to four hypotheses. Now, I attempt to give a brief answer to each of the four subquestions, based on the results obtained in section 4.

(i) Which bankruptcy prediction model has the highest accuracy in predicting bankruptcy?

The model that has the highest accuracy is the discriminant analysis-estimated model with the Ohlson (1980) variables. It attains a AUC-statistic of 0.883, which outperforms several other models in this study, and also compares favorable to models from other studies.

(ii) Which model carries the most information content regarding bankruptcy?

The results with regards to this question are more ambiguous. Arguably, the best model has a high Cox and Snell R-squared, a -2LL ratio as low as possible, and only the model should attain statistical significance, but not the constant term or the annual default rate, as the model should subsume the information that is present in those factors. The model that comes closest to these criteria is the model that includes the Ohlson (1980) variables, estimated with logistic regression (see section 4.2.2).

(iii) Does model performance improve, stay constant or decline under a changing macroeconomic environment in Europe?

Model performance unambiguously declines under a changing macroeconomic environment in Europe. Out-of-sample performance in a different environment (see section 4.2.2) showed that model accuracy as measured by the AUC-stat decreases significantly in almost every model. Model information content did not show such a clear decline, although in many cases, other predictors included show statistical significance, indicating that they carry information about bankruptcy that is not reflected by the bankruptcy prediction model.

(iv) Does the particular set of ratios included in the model have an impact on either accuracy, information content, or both?

The results do not show clearly whether both model accuracy and information content are greatly sensitive to a changing set of ratios, even though theoretically, they should reflect the same dimension. It seems that the Altman, Ohlson and Zmijewski variables do not generally have special significance, and can be replaced by a theoretical substitute set, although this seems to be more so for the Altman (1968) and Zmijewski (1984) predictors, than for the Ohlson (1980) predictors.

From these results, it seems that the best set of variables is the Ohlson (1980) set without the macroeconomic variables, estimated using logistic regression.

5.2 Summary of results

This study examined the accuracy and information content of bankruptcy prediction models using logistic regression and multiple discriminant analysis. Its first and foremost purpose was to decide on whether discriminant analysis or logistic regression was the superior forecasting method. Based on econometric theory, both models appear to be biased and inconsistent. In practice, however, both models often perform reasonably well, recently verified by e.g. Agarwal and Taffler (2007) and Reisz and Perlich (2007) and sometimes outperforming models generally thought to be more accurate (Agarwal and Taffler, 2008). In newer research, it is generally outperformed by other models, such as the Shumway (2001) or the Hillegeist et al. (2004) specifications, but because these models require less data (at one point in time), they are relatively easy to compute and still used as a kind of benchmark. The recent results in Agarwal and Taffler (2008), Wu, Gaunt and Gray (2010) and Bauer and Agarwal (2014) are sufficient reason to carefully examine more broadly model performance over time.

First, this study tries to examine whether logistic regression, ceteris paribus, suits bankruptcy prediction better than multiple discriminant analysis. Although econometric theory would imply that it does (due to the strict assumptions on MDA) in practice this is not necessarily so. Many studies, e.g. Mensah (1984), Platt and Platt (1991), Grice and Ingram (2001), Grice and Dugan (2003) and Wu, Gaunt and Gray (2010) conducted similar exercises. With the exception of Wu, Gaunt and Gray (2010), these studies conflated evaluating the performance of a particular set of ratios with evaluating a particular econometric method. They would for instance evaluate discriminant analysis-estimated model with the Altman (1968) variables, and the logistic regression-estimated model with the Ohlson (1980) variables. If one of those performs superior to the other, one cannot identify the cause: a better econometric method or a better set of predictor variables.

This study aimed to separate these issues and tackle them both. First, when testing hypothesis 1, 4 sets of variables were used to estimate bankruptcy prediction models using *both* logistic regression and discriminant analysis. This way, one can rule out that a particular set of predictors is the cause of a model performing better than another. Furthermore, hypothesis 1 was evaluated in a particularly narrow time period so as to rule out noise coming from nonstationarity in the predictor variables. Any difference between model accuracy and information content, then, can only be caused by a difference in econometric method. The results showed that hypothesis 1 is rejected. Logistic regression does not clearly perform better in either model accuracy or model information content. In contrast with the econometrics, empirically, there is no reason for bankruptcy prediction researchers to a priori discard discriminant analysis or logistic regression as a bad bankruptcy prediction model, or as a benchmark. Consider for instance the results of Bauer and Agarwal (2014). Out-of-sample accuracy of the Bharath and Shumway (2008) DD-model is reported as an AUC-statistic of 0.867. In the present study, the discriminant analysis-estimated model with Zmijewski (1984) variables showed an AUC-statistic of 0.873.

Of course, the aforementioned example omits the fact that the results of Bauer and Agarwal (2014) encompass a long out-of-sample period (1979-2002), while hypothesis 1 uses models estimated with data from 2005 and evaluated in 2006-2007. This issue is evaluated in hypothesis 2: can bankruptcy prediction model retain their accuracy and information content over time?

In the tests of hypothesis 2, models are being estimated in 2005-2007, and evaluated in 2011-2013. This represents a rigorous test of model robustness, because the financial crisis hit the real economy in Europe in 2010, and changed the macroeconomic environment from being characterized by relatively high growth, low unemployment (ante 2007) to one characterized by high uncertainty, many bankruptcies and stagnation (post 2010). The results show that nearly all bankruptcy prediction models show a sharp decline in accuracy when applied to this different period. The models do appear to retain their information content, although the baseline hazard proxy (i.e. the annual default rate) and the constant shows statistical significance in most of the tests too, indicating there is additional information not captured by the bankruptcy prediction models.

The discriminant analysis-estimated model with the Zmijewski (1984) predictors and the logistic regression-estimated model with the Ohlson (1980) predictors performed best in terms of accuracy and information content, as they scored the highest AUC-scores of all models, and subsumed statistical significance of the annual default rate in the information content tests. The logistic regression model scored an AUC-score of 0.882, which outperforms the Agarwal and Taffler (2007) discriminant analysis model in Agarwal and Taffler (2007), the Bharath and Shumway (2008) and Altman (1968) models in Bauer and Agarwal (2014), although in the same study it is outperformed by the Shumway (2001) and Campbell et al. (2008) hazard model specifications.

Hypotheses 3 and 4 then provide tests as to whether these results are robust to the inclusion of simple alternative prediction models (hypothesis 3) and whether the previously obtained results are robust to the inclusion of slightly different predictor variables in the bankruptcy prediction model (hypothesis 4). Hypothesis 3 shows that when accuracy of a bankruptcy prediction model is compared to the accuracy of an alternative, simpler predictor, a bankruptcy model clearly outperforms this predictor by a large degree. In the information content test, all bankruptcy prediction models again retain their significance. However, in nearly all tests, several other predictors also show statistical significance, indicating that these alternative predictors capture information about bankruptcy that the model is unable to grasp.

Hypothesis 4 shows that in some cases, model accuracy sharply deteriorates when using different variables than the traditional sets such as the Altman (1968), Ohlson (1980) and Zmijewski (1984) variables, but in other cases, there is no decline, or even an improvement relative to the Altman (1968) and Zmijewski (1984) sets of predictors. The models do however retain their information content, but again, the baseline hazard rate proxy and constant show statistical significance in many tests, indicating that they carry information about bankruptcy that is not grasped by the bankruptcy prediction models.

Finally, throughout this study, bankruptcy prediction models have been estimated with and without the inclusion of macroeconomic variables. These variables were included as control variables, intended to absorb the changes in macroeconomic environment to ensure that the model remains accurate and informative under changing macroeconomic circumstances. From this perspective, it follows the studies of Nam et al. (2008) and Tinoco and Wilson (2013). The results however, are ambiguous. In some cases, such as the MDA-estimated model with the Ohlson (1980) variables, it is clear that the addition macroeconomic variables to the other predictors clearly improves accuracy and information content, but in other cases, such as the MDA-estimated model with the Zmijewski (1984) variables, inclusion of macroeconomic variables deteriorates both model accuracy and information content.

The practical implications of the results include the following. First, if practitioners want to use the bankruptcy prediction model with the highest predictive accuracy, based on these results, one should use the Ohlson (1980) predictors estimated with logistic regression. Despite econometric flaws, this model attains an AUC-score that outperforms several other specifications reported in other studies (cf. Bauer and Agarwal, 2014). Furthermore, almost all bankruptcy prediction models contain useful information about bankruptcy beyond knowing the default rate of the previous year. Practitioners should keep in mind however, that the use of multiple indicators can add value to their predictions. In many information content tests, the bankruptcy prediction model fails to subsume statistical significance of other alternative predictors, such as a firm's market capitalization, or the aforementioned annual default rate. That means these indicators reflect information about bankruptcy beyond the bankruptcy prediction model. Practitioners could – and probably should – consider alternative measures as a complement to statistical bankruptcy prediction models. Finally, this study provides parameter estimates for in total 36 bankruptcy prediction models. These parameters might serve as a better standard for future research or prediction than the original Altman (1968), Ohlson (1980) or Zmijewski (1984) coefficients, because these are specifically calibrated for a cross-European context on fairly recent data. Given the strong sensitivity of model performance to nonstationarity in the predictor variables, this could be especially important, since anyone using the Altman (1968) coefficients, or any other estimates that are 40+ years old, is bound to end up with a bad forecasting model.

5.3 Limitations and suggestions for further research

This study attempted to evaluate the accuracy and information content of several bankruptcy prediction models that were estimated using logistic regression and discriminant analysis. In many current research, logistic regression and discriminant analysis is considered obsolete or only as a benchmark prediction model. Newer models, in particular, the Shumway (2001) specification, should be better specified according to econometric theory and hence should be used instead of models that give biased parameter estimates. This study could not use the Shumway (2001) model because of lack of the necessary data to estimate a pooled logit model. For the same reason, the DD-model could not be used. Hence, this study features "second choice" models that are easier to compute, since they require only observations at one point in time in a particular sample.

Furthermore, I choose to incorporate macroeconomic variables in this study in order to absorb the change in bankruptcy risk across countries and over time. Several studies before, in particular Nam et al. (2008) and Tinoco and Wilson (2013) had done this before, but the results in this study showed that the inclusion of macroeconomic effects as predictor variables sometimes had a positive effect on model accuracy and information content, but oftentimes also a negative effect. Thus, the evidence on the existence of country-specific bankruptcy risk is conflicting. This study fails to investigate a related alternative, namely, the importance of industry effects. As well as being theoretically plausible, several other studies, Chava and Jarrow (2004) and Lee and Choi (2013) reported that industry effects improved model forecasting.

Then, in the information content tests, this study uses a far smaller number of firm-year observations than is conventionally the case. That is a corollary of using a hold-out sample for only three years. Other studies conducted their analyses sometimes in samples ranging 10-20 years (e.g. Agarwal and Taffler, 2007 and Bauer and Agarwal, 2014, among many others). Thus, due to the small sample period and the censoring of unavailable observations, the firm-year number was rather small and allows for limited inference. Furthermore, survival models assume censoring is random, while it may be the case that data is more often lacking on firms that eventually went bankrupt, introducing a bias towards healthy firms in the parameter estimates.

An additional, and arguably the most important limitation is the lack of theoretical contribution of this paper. Like many other bankruptcy prediction research (e.g. research from Altman, 1968 to Shumway, 2001 or Wu, Gaunt and Gray, 2010), this research is only interested in bankruptcy prediction as such, and lacks connection to established economic theory. The only model reviewed in the literature that has a clear, nontrivial connection to the theoretical literature is the Merton (1974) inspired BSM-Prob model. The merit of this model is that it is derived out of an established and extensive framework in financial economics, whereas all its competitors lack any such basis, beyond rather trivial justification for including a particular predictor in a model. It seems the only justification, in this research as well as many others, is that it works.

A purely pragmatic approach such as the one taken in this study thus leads to the question of the purpose of this kind of research. Based on the arguments above, prediction can be very important for practitioners, who are, of course, looking for a device that discriminates optimally between bankrupt and non-bankrupt corporations. From a scientific point of view, the point is not necessarily to make as many accurate predictions as possible, but to attempt to verify a body of hypotheses, and to reject or amend bodies of hypotheses in accordance with empirical results. Studies such as the present one or most research in this direction in the literature, contribute very little to this aim.

Taking these limitations into account, further empirical research should probably focus on using hazard models as the best class of bankruptcy prediction models, while theoretical interest leads us to consider the BSM-Prob model as the most viable alternative. This study evaluates models using a wide range of predictors, but no model was able to outperform hazard models in accuracy and information content as reported in other studies (Shumway, 2001, Campbell et al., 2008, Bauer and Agarwal, 2014 and others). Furthermore, in some cases, macroeconomic variables appear to add predictive power to the model, but this is highly sensitive to the particular set of variables used in combination with these macroeconomic variables. From the robustness checks conducted in section 4.2.2 we can conclude that the effect of bias towards overrepresented countries in the sample was only marginal. That provides evidence in favor of more research to be conducted in multi-country samples. Perhaps combining various predictors in a multi-level specification that includes industry-effects, country or macroeconomic-effects and firm-level covariates, provides for a model that can outperform conventional sets of variables in a multi-country sample, although research that progresses is prone to the aforementioned critique of lacking theoretical content.

Appendix

Appendix A: Mathematical specification of discriminant analysis

A linear discriminant function with a dichotomous outcome is defined mathematically (Collins and Green, 1982, based on Fisher, 1936) as follows:

$$\max_{\vec{v}} S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(\vec{v}\cdot\vec{\mu}_{y=1} - \vec{v}\cdot\vec{\mu}_{y=0})^2}{\vec{v}^T\cdot\Sigma y{=}1\cdot\vec{v} + \vec{v}^T\cdot\Sigma y{=}0\cdot\vec{v}} \qquad \text{Eq. (A.1)}$$

where $\vec{v} = v_1 \ldots v_n$ represent the discriminant weights for n independent variables. The independent variables of the bankrupt class (denoted by y=1) are distributed with mean $\vec{\mu}_{y=1}$ and variance-covariance matrix $\Sigma_{y=1}$ and the independent variables of the non-bankrupt class (denoted by y=0) are distributed with mean $\vec{\mu}_{y=0}$ and variance-covariance matrix $\Sigma_{y=0}$. Accordingly, $\vec{v}^T$ represents the transpose of the weights used to compute the weighted variance-covariance matrix of each class. Computationally, the linear discriminant function is then obtained as the solution of an eigenvalue problem. When the individual weights are established, discriminating ability of the function can be assessed. The final variable profile is determined by the relative contribution of each variable to the total discriminating power of the function, and the interaction between them. This yields a model which is then applied to the sample itself, the ratios 1 year before the target year serving as input, so that the prediction of the model (classification into either bankrupt or non-bankrupt category) can be compared to the actual outcome.

Hair et al. (2006) also give a statistic for assessing the level of significance for the classification accuracy: $t = \frac{p-.5}{\sqrt{\frac{.5(1.0-0.5)}{N}}}$ where p means proportion of the group correctly classified, and N is the sample size. The statistic is t-distributed. In the case of bankruptcy prediction, the interpretation of a discriminant Z score is that the greater a firm's bankruptcy potential, the lower its discriminant score and vice versa.

Appendix B: Mathematical specification of maximum-likelihood estimation

A log-likelihood function is described mathematically as follows (Ohlson, 1980, Hair et al, 2006):

$$l(\beta) = \sum_{i\in S_1} \log P(X_{ij}, \beta_j) + \sum_{i\in S_2} \log(1 - P(X_{ij}, \beta_j)) \qquad \text{Eq. (B.1)}$$

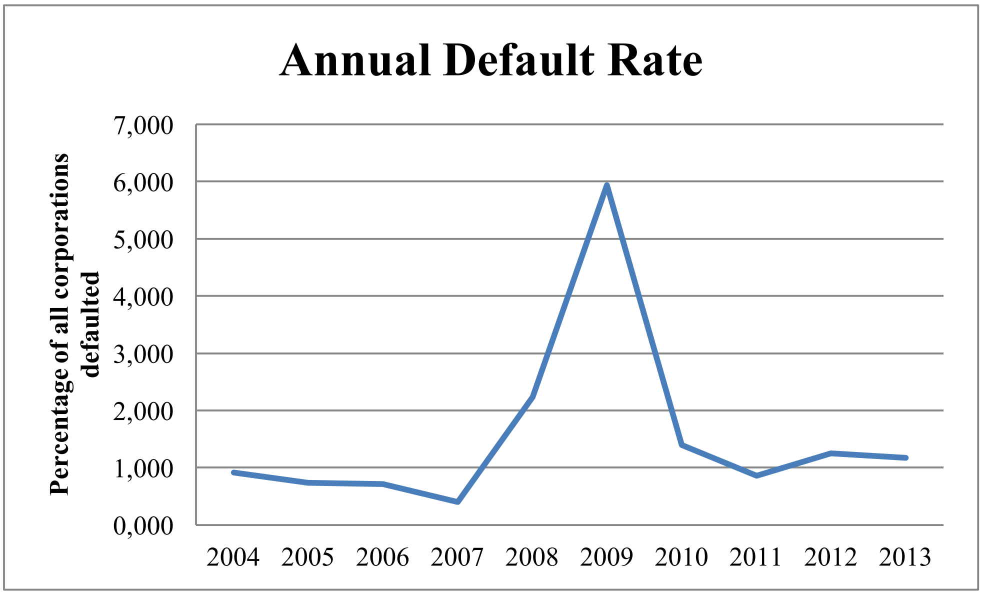
This is the log-likelihood function, which in turn should be maximized so that the parameter estimations best approximate the data. X_{ij} stands for the a j number of independent variables with i observations, β stands for a vector of unknown parameters (to be estimated) and P(Xi, β) stands for the probability of bankruptcy for any given Xi and β. P is a probability density function ($0 \leq P \leq 1$). S_1 stands for the all observation set of bankrupt firms and S_2 stands for the observation set of non-bankrupt firms.

Appendix C: Annual Default Rates

Shumway (2001) and Kim and Partington (2014) mention that executing the hazard modelling procedure requires a baseline hazard approximator (representing the firm-invariant default risk), included in the pooled logit model as an additional predictor variable. Shumway (2001), Bauer and Agarwal (2014) among others use the annual default rate, which is computed as follows:

$$ADR_t = \frac{Bankrupt\ firms_t}{Bankrupt\ firms_t + Non-bankrupt\ firms_t} * 100\% \qquad \text{(Shumway, 2001)} \quad \text{Eq. (C.1)}$$

Replicating this procedure would require me to collect all listed firms on the ORBIS database in Europe and manually compute the default rate. There is however a better option: I will use the freely available Moody's Investors Service European Corporate Default Rates. Graph 1 shows the annual corporate default rate in Europe.



Graph 1: Annual default rates across the EU

The Moody's Investors Relations database provides data on corporate default rates for registered users for free. For the period of 1985-2009 the used default rates can be found in exhibit 19 at page 19, in the column "all":

https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_123911

For the subsequent period in my sample (i.e. 2011-2013), the used default rates can be found in this excel file in exhibit 31 and column X (all rated):

https://www.moodys.com/research/Annual-Default-Study-Corporate-Default-and-Recovery-Rates-1920-2012--PBC_151956

As the annual default rate of 2013 is not available as of yet, the 2013 annual default rate is taken to be the weighted average of the default rates from 2010-2012 (Annual Issuer-Weighted Corporate Default Rates by Alphanumeric Rating, 1983-2012).

Appendix D: Ratio's used for Factor Analysis

I conducted a factor analysis on the following set of variables. I included all variables from Altman (1968), Ohlson (1980), Zmijewski (1984) as well as all variables from Pompe and Bilderbeek (2005) that were available in my dataset. All variables from the latter were divided into four categories that reflect firm performance: solvability, liquidity, activity and profitability. All factors up that explained up to at least 70% of the cumulative variance were retained. Then, the most important contributor to that factor (i.e. the variable that correlates the highest with that factor) was included in the prediction model instead of that factor.

Table D: Variables included in the factor analysis

Variable definition:	Retained as predictor:
1. Working capital/total assets	No
2. Retained earnings/total assets	No
3. EBIT/total assets	No
4. Market value of assets/total liabilities	No
5. Sales/total assets	No
6.Net income/total assets	Yes
7.Total liabilities/total assets	No
8. Current assets/current liabilities	Yes
9. Natural log of total assets	Yes
10. Dummy, 1 if liabilities exceed total assets, 0 otherwise	No
11. EBITDA/total liabilities	No
12. Dummy, 1 if net income was negative the last two years, 0 otherwise	No
12. Change in net income	No
13. Current assets/total assets	Yes
14. Working capital/sales	No
15. Working capital/operating revenue	Yes
16. Market capitalization/Total assets	Yes
17. Long term debt/total assets	No
18. current liabilities/total assets	No
19. EBIT/total assets	No
20. Net income/total equity	No
21. EBIT/total equity	No
22. Equity/turnover	No
23. Turnover/current assets	Yes
24. Working capital/no. of employees	No
25. Inflation	No
26. GDP Growth	Yes

Note: Some variables are featured in more than one model. For convenience sake, I removed double entries.

Appendix E: Additional descriptive statistics

Table E1: Descriptive statistics (Balanced) (Data used in hypothesis 1)

Panel A: Estimation sample (2005)

	Altman (1968) variables					**Ohlson (1980) variables**								
	A1	**A2**	**A3**	**A4**	**A5**	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**
Mean	0.234	0.006	0.028	1.000	1.684	7.475	0.645	0.227	1.232	0.072	0.005	0.093	0.147	-0.001
Median	0.226	0.019	0.037	0.165	1.433	7.513	0.644	0.216	0.806	0.000	0.019	0.057	0.000	0.028
SD	0.267	0.175	0.173	4.129	1.341	1.544	0.356	0.266	2.263	0.259	0.178	0.462	0.355	0.393
Min	-0.546	-0.950	-0.778	0.004	0.000	3.488	0.010	-0.546	0.005	0.000	-0.950	-1.892	0.000	-1.000
Max	0.904	0.447	0.553	34.712	7.583	10.977	2.537	0.904	18.066	1.000	0.447	2.505	1.000	1.000
N	2097	2097	2097	2097	2097	2362	2362	2362	2362	2362	2362	2362	2362	2362
	Zmijewski (1984) variables					**Factor analysis variables**								
	Z1	**Z2**	**Z3**			**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	
Mean	0.005	0.646	2.257			0.008	0.253	0.702	1.943	0.004	7.643	2.104	0.174	
Median	0.018	0.644	1.239			0.021	0.155	0.773	1.241	0.001	7.702	1.608	0.098	
SD	0.177	0.359	4.435			0.172	0.537	0.255	3.215	0.011	1.498	1.371	0.223	
Min	-0.950	0.010	0.054			-0.950	-0.971	0.019	0.054	0.000	3.488	0.707	0.003	
Max	0.447	2.537	41.048			0.447	4.341	1.000	41.048	0.085	10.977	9.471	1.696	
N	2395	2395	2395			2116	2116	2116	2116	2116	2116	2116	2116	

	Macroeconomic predictors		**Dependent variable**
	M1	**M2**	**D1**
Mean	2.080	2.414	0.144
Median	1.607	1.999	0.000
SD	1.285	1.050	0.352
Min	0.000	0.000	0.000
Max	9.471	8.989	1.000
N	2588	2588	2588

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents list wise (i.e. balanced) data. Variable definitions are as follows: A1. Working capital/total assets, A2. Retained earnings/total assets, A3. EBIT/total assets, A4. Market cap/total liabilities, A5. Sales/total assets, O1. Natural log of total assets, O2. Total liabilities/total assets, O3. Working capital/total assets, O4. Current liabilities/current assets, O5. 1 If total liabilities exceed total assets, 0 otherwise, O6. Net income/total assets, O7. Income from operations after depreciation/total liabilities, O8. 1 If net income was negative for the last 2 years, 0 otherwise, O9. Change in net income for the most recent period, Z1.Net income/total assets, Z2.Total liabilities/total assets, Z3. Current assets/current liabilities, F1.Net income/total assets, F2. Working capital/operating revenue, F3. Current assets/total assets, F4. Current assets/current liabilities, F5. Turnover/current assets, F6. Natural log of total assets, F7. GDP growth (%), F8. Market capitalization/total assets, B1.Working capital/operating revenue, B2.EBIT/market cap, B3. Current liabilities/total assets, B4. Market cap/turnover, C1.Working capital/sales, C2.Net income/market cap, C3.Long term debt/total assets, C4.Working capital/number of employees, M1. GDP growth rate (%), M2. Inflation rate (%), D1. 1=Bankruptcy, 0 otherwise.

Panel B: Hold-out sample (2006-2007)

	Altman (1968) variables					Ohlson (1980) variables								
	A1	A2	A3	A4	A5	O1	O2	O3	O4	O5	O6	O7	O8	O9
Mean	0.214	0.014	0.038	0.908	1.635	7.681	0.649	0.211	1.179	0.082	0.017	0.137	0.146	0.013
Median	0.205	0.028	0.044	0.145	1.403	7.698	0.632	0.198	0.775	0.000	0.027	0.072	0.000	0.042
SD	0.262	0.185	0.182	3.781	1.324	2.576	0.373	0.262	2.059	0.275	0.179	0.470	0.353	0.350
Min	-0.546	-0.950	-0.778	0.004	0.000	3.488	0.010	-0.546	0.005	0.000	-0.950	-1.892	0.000	-1.000
Max	0.904	0.447	0.553	34.712	7.583	10.977	2.537	0.904	18.066	1.000	0.447	2.505	1.000	1.000
N	4615	4615	4615	4615	4615	5129	5129	5129	5129	5129	5129	5129	5129	5129

	Zmijewski (1984) variables			Factor analysis variables							
	Z1	Z2	Z3	F1	F2	F3	F4	F5	F6	F7	F8
Mean	0.015	0.650	2.394	0.017	0.252	0.700	2.169	0.004	7.798	2.990	0.168
Median	0.027	0.633	1.284	0.028	0.152	0.774	1.294	0.001	7.828	2.375	0.088
SD	0.181	0.376	5.042	0.179	0.563	0.259	4.144	0.011	1.562	1.413	0.231
Min	-0.950	0.010	0.054	-0.950	-0.971	0.019	0.054	0.000	3.488	0.824	0.003
Max	0.457	2.537	41.048	0.447	4.431	1.000	41.048	0.085	10.977	11.087	1.696
N	5262	5262	5262	4904	4904	4904	4904	4904	4904	4904	4904

	Macroeconomic predictors		Dependent variable
	M1	M2	D1
Mean	3.017	2.136	0.144
Median	2.375	1.791	0.000
SD	1.357	0.916	0.351
Min	0.000	0.000	0.000
Max	11.087	10.107	1.000
N	5945	5945	5945

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents listwise (i.e. balanced) data. Variable definitions are as follows: A1. Working capital/total assets, A2. Retained earnings/total assets, A3. EBIT/total assets, A4. Market cap/total liabilities, A5. Sales/total assets, O1. Natural log of total assets, O2. Total liabilities/total assets, O3. Working capital/total assets, O4. Current liabilities/current assets, O5. 1 If total liabilities exceed total assets, 0 otherwise, O6. Net income/total assets, O7. Income from operations after depreciation/total liabilities, O8. 1 If net income was negative for the last 2 years, 0 otherwise, O9. Change in net income for the most recent period, Z1.Net income/total assets, Z2.Total liabilities/total assets, Z3. Current assets/current liabilities, F1.Net income/total assets, F2. Working capital/operating revenue, F3. Current assets/total assets, F4. Current assets/current liabilities, F5. Turnover/current assets, F6. Natural log of total assets, F7. GDP growth (%), F8. Market capitalization/total assets, B1.Working capital/operating revenue, B2.EBIT/market cap, B3. Current liabilities/total assets, B4. Market cap/turnover, C1.Working capital/sales, C2.Net income/market cap, C3.Long term debt/total assets, C4.Working capital/number of employees, M1. GDP growth rate (%), M2. Inflation rate (%), D1. 1=Bankruptcy, 0 otherwise.

Table E2: Descriptive statistics (Balanced) (Data used in hypotheses 2,3, 4)

Panel A: Estimation sample (2005-2007)

	Altman (1968) variables					Ohlson (1980) variables								
	A1	**A2**	**A3**	**A4**	**A5**	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**
Mean	0.220	0.011	0.035	0.937	1.650	7.616	0.647	0.216	1.195	0.079	0.014	0.123	0.146	0.009
Median	0.212	0.025	0.042	0.151	1.413	7.632	0.635	0.203	0.785	0.000	0.025	0.067	0.000	0.038
SD	0.263	0.182	0.179	3.893	1.330	1.589	0.368	0.263	2.125	0.270	0.179	0.468	0.354	0.603
Min	-0.546	-0.950	-0.778	0.004	0.000	3.488	0.010	-0.546	0.005	0.000	-0.950	-1.892	0.000	-1.000
Max	0.904	0.447	0.553	34.712	7.583	10.977	2.537	0.904	18.066	1.000	0.447	2.505	1.000	1.000
N	6712	6712	6712	6712	6712	7491	7491	7491	7491	7491	7491	7491	7491	7491

	Zmijewski (1984) variables			Factor analysis variables							
	Z1	**Z2**	**Z3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**
Mean	0.012	0.649	2.351	0.014	0.252	0.700	2.101	0.004	7.751	2.723	0.170
Median	0.024	0.637	1.270	0.027	0.153	0.773	1.278	0.001	7.783	2.362	0.091
SD	0.181	0.371	4.861	0.177	0.555	0.258	3.888	0.011	1.544	1.458	0.228
Min	-0.950	0.010	0.054	-0.950	-0.971	0.019	0.054	0.000	3.488	0.707	0.003
Max	0.447	2.537	41.048	0.447	4.341	1.000	41.048	0.085	10.977	11.087	1.696
N	7657	7657	7657	7020	7020	7020	7020	7020	7020	7020	7020

	Alternative set of predictors (1)				Alternative set of predictors (2)				Macroeconomic predictors		Dependent variable
	B1	**B2**	**B3**	**B4**	**C1**	**C2**	**C3**	**C4**	**M1**	**M2**	**D1**
Mean	0.251	1.744	0.562	2774.826	0.261	0.690	0.089	94.565	2.733	2.220	0.145
Median	0.154	0.430	0.542	130.892	0.156	0.163	0.018	6.207	2.375	1.821	0.000
SD	0.529	5.949	0.327	11860.428	0.567	4.503	0.147	977.909	1.403	0.967	0.352
Min	-0.971	-13.104	-0.210	1.692	-0.777	-16.129	0.000	-2196.489	0.000	0.000	0.000
Max	4.341	36.220	2.179	97895.781	4.599	26.756	0.807	43230.320	11.087	10.107	1.000
N	6971	6971	6971	6971	4936	4936	4936	4936	8528	8528	8528

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents list wise (i.e. balanced) data. Variable definitions are as follows: A1. Working capital/total assets, A2. Retained earnings/total assets, A3. EBIT/total assets, A4. Market cap/total liabilities, A5. Sales/total assets, O1. Natural log of total assets, O2. Total liabilities/total assets, O3. Working capital/total assets, O4. Current liabilities/current assets, O5. 1 If total liabilities exceed total assets, 0 otherwise, O6. Net income/total assets, O7. Income from operations after depreciation/total liabilities, O8. 1 If net income was negative for the last 2 years, 0 otherwise, O9. Change in net income for the most recent period, Z1.Net income/total assets, Z2.Total liabilities/total assets, Z3. Current assets/current liabilities, F1.Net income/total assets, F2. Working capital/operating revenue, F3. Current assets/total assets, F4. Current assets/current liabilities, F5. Turnover/current assets, F6. Natural log of total assets, F7. GDP growth (%), F8. Market capitalization/total assets, B1.Working capital/operating revenue, B2.EBIT/market cap, B3. Current liabilities/total assets, B4. Market cap/turnover, C1.Working capital/sales, C2.Net income/market cap, C3.Long term debt/total assets, C4.Working capital/number of employees, M1. GDP growth rate (%), M2. Inflation rate (%), D1. 1=Bankruptcy, 0 otherwise.

Panel B: Hold-out sample (2011-2013)

	Altman (1968) variables					Ohlson (1980) variables								
	A1	**A2**	**A3**	**A4**	**A5**	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**
Mean	0.213	-0.199	0.003	1.150	1.460	7.818	0.625	0.208	1.302	0.091	-0.017	0.066	0.210	-0.067
Median	0.198	0.013	0.026	0.163	1.206	7.838	0.588	0.187	0.730	0.000	0.013	0.000	0.000	-0.034
SD	0.260	0.224	0.204	5.400	1.237	1.624	0.424	0.262	2.807	0.287	0.220	0.877	0.407	0.612
Min	-0.604	-1.310	-1.086	0.003	0.000	3.587	0.006	-0.604	0.003	0.000	-1.311	-1.390	0.000	-1.000
Max	0.860	0.416	0.456	53.217	6.850	11.288	2.988	0.861	24.131	1.000	0.417	50.110	1.000	1.000
N	6132	6132	6132	6132	6132	6818	6818	6818	6818	6818	6818	6818	6818	6818

	Zmijewski (1984) variables			Factor analysis variables							
	Z1	**Z2**	**Z3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**
Mean	-0.017	0.625	3.099	-0.021	0.295	0.678	2.712	0.003	7.950	0.437	0.182
Median	0.013	0.587	1.368	0.013	0.156	0.749	1.378	0.001	7.955	1.618	0.092
SD	0.220	0.427	8.242	0.224	0.821	0.269	6.502	0.008	1.592	2.615	0.269
Min	-1.311	0.006	0.039	-1.311	-1.623	0.014	0.039	0.000	3.587	-8.864	0.003
Max	0.417	2.988	73.868	0.417	6.877	1.000	73.868	0.065	11.288	8.264	2.102
N	7141	7141	7141	6246	6246	6246	6246	6246	6246	6246	6246

	Alternative set of predictors (1)				Alternative set of predictors (2)				Macroeconomic predictors		Dependent variable
	B1	**B2**	**B3**	**B4**	**C1**	**C2**	**C3**	**C4**	**M1**	**M2**	**D1**
Mean	0.299	1.023	0.524	2820.863	0.341	0.103	0.110	96.528	0.677	2.314	0.140
Median	0.157	0.243	0.477	192.525	0.178	0.070	0.038	29.543	1.711	2.117	0.000
SD	0.800	5.599	0.362	7758.678	0.870	5.027	0.160	591.77	2.451	0.865	0.347
Min	-1.623	-19.943	0.001	2.127	-1.323	-22.529	0.000	-5985.883	-8.864	-1.094	0.000
Max	6.877	30.801	2.305	39734.737	7.487	45.742	0.828	21236.341	8.264	6.094	1.000
N	6199	6199	6199	6199	4203	4203	4203	4203	8129	8129	8129

Note: descriptive statistics of all predictor variables used in this study to forecast bankruptcy, and the dependent variable. This table presents listwise (i.e. balanced) data. Variable definitions are as follows: A1. Working capital/total assets, A2. Retained earnings/total assets, A3. EBIT/total assets, A4. Market cap/total liabilities, A5. Sales/total assets, O1. Natural log of total assets, O2. Total liabilities/total assets, O3. Working capital/total assets, O4. Current liabilities/current assets, O5. 1 If total liabilities exceed total assets, 0 otherwise, O6. Net income/total assets, O7. Income from operations after depreciation/total liabilities, O8. 1 If net income was negative for the last 2 years, 0 otherwise, O9. Change in net income for the most recent period, Z1.Net income/total assets, Z2.Total liabilities/total assets, Z3. Current assets/current liabilities, F1.Net income/total assets, F2. Working capital/operating revenue, F3. Current assets/total assets, F4. Current assets/current liabilities, F5. Turnover/current assets, F6. Natural log of total assets, F7. GDP growth (%), F8. Market capitalization/total assets, B1.Working capital/operating revenue, B2.EBIT/market cap, B3. Current liabilities/total assets, B4. Market cap/turnover, C1.Working capital/sales, C2.Net income/market cap, C3.Long term debt/total assets, C4.Working capital/number of employees, M1. GDP growth rate (%), M2. Inflation rate (%), D1. 1=Bankruptcy, 0 otherwise.

Appendix F: Complete parameter estimates – hypothesis 1

In this appendix, I show parameter estimates for every bankruptcy prediction model used in this study. Each panel shows a different set of predictor variables included in the bankruptcy prediction model. Each column represents a different bankruptcy prediction model: column 2 and column 3 are models estimated with MDA, column 4 and 5 are estimated with logistic regression. Column 3 and column 5 include the two macroeconomic variables (GDP growth rate and inflation rate) as additional predictors in the model. In Panel D, no macroeconomic variables are added, because factor analysis determined the set of predictors. This appendix shows estimates used to evaluate hypothesis 1.

Table F: 2005 parameter estimates

Estimation technique:	**Multiple discriminant analysis:**		**Logistic regression:**	
Panel A: Altman (1968) variables				
Constant	0.327	-0.304	-2.121	-0.190
			(0.000)	(0.529)
A1.	0.106	-0.068	0.118	0.579
			(0.644)	(0.035)
A2.	2.727	2.538	-2.294	-2.140
			(0.002)	(0.003)
A3.	3.426	3.334	-5.545	-5.624
			(0.000)	(0.000)
A4.	0.036	0.030	-0.163	-0.119
			(0.014)	(0.033)
A5.	-0.298	-0.274	0.303	0.296
			(0.000)	(0.000)
M1.		0.177		-0.571
				(0.000)
M2.		0.121		-0.501
				(0.000)
-2 LL			1509.833	1426.872
Cox and Snell R^2			0.192	0.223
Nagelkerke R^2			0.317	0.368
Hosmer-Lemeshow Test			40.043	45.210
			(0.000)	(0.000)
Box's M	910.564	1630.030		
	(0.000)	(0.000)		
Wilks' lambda	0.777	0.759		
	(0.000)	(0.000)		
N	2097	2097	2097	2097
Panel B: Ohlson (1980) variables				
Constant	2.944	2.095	-8.524	-5.849
			(0.000)	(0.000)
O1.	-0.320	-0.299	0.631	0.610
			(0.000)	(0.000)
O2.	-0.992	-0.848	2.562	2.540
			(0.000)	(0.000)
O3.	0.152	-0.020	-0.024	0.301
			(0.932)	(0.321)
O4.	0.099	0.089	-0.379	-0.331
			(0.000)	(0.000)
O5.	3.350	3.271	-0.321	0.178
			(0.326)	(0.605)
O6.	0.031	0.017	-3.343	-2.707
			(0.000)	(0.000)
O7.	-0.218	-0.318	-1.229	-1.442
			(0.000)	(0.000)
O8.	-0.514	-0.504	0.569	0.590
			(0.002)	(0.002)
O9.	0.751	0.728	-1.074	-1.029
			(0.000)	(0.000)
M1.		0.121		-0.264
				(0.008)
M2.		0.176		-1.006
				(0.000)

-2LL			1294.634	1207.168
Cox and Snell R^2			0.258	0.285
Nagelkerke R^2			0.452	0.499
Hosmer-Lemeshow Test			8.939	58.222
			(0.348)	(0.000)
Box's M	2343.402	2932.104		
	(0.000)	(0.000)		
Wilks' lambda	0.713	0.696		
	(0.000)	(0.000)		
N	2362	2362	2362	2362
Panel C: Zmijewski (1984) variables				
Constant	0.534	-0.531	-1.815	1.112
			(0.000)	(0.004)
Z1.	4.966	4.591	-4.487	-4.371
			(0.000)	(0.000)
Z2.	-0.945	-0.796	0.567	0.650
			(0.010)	(0.008)
Z3.	0.023	0.021	-0.282	-0.280
			(0.000)	(0.001)
M1.		0.142		-0.492
				(0.000)
M2.		0.289		-1.014
				(0.000)
-2LL			1721.135	1575.299
Cox and Snell R^2			0.134	0.185
Nagelkerke R^2			0.232	0.320
Hosmer-Lemeshow Test			66.744	63.481
			(0.000)	(0.000)
Box's M	826.422	1377.010		
	(0.000)	(0.000)		
Wilks' lamba	0.836	0.810		
	(0.000)	(0.000)		
N	2395	2395	2395	2395
Panel D: Factor analysis variables:				
Constant	2.817		-5.811	
			(0.000)	
F1.	5.190		-6.925	
			(0.000)	
F2.	0.322		-0.728	
			(0.001)	
F3.	-1.216		2.599	
			(0.000)	
F4.	0.044		-0.644	
			(0.000)	
F5.	-7.534		-28.526	
			(0.070)	
F6.	-0.328		-0.505	
			(0.000)	
F7.	0.176		-0.450	
			(0.000)	
F8.	-0.012		0.358	
			(0.412)	
-2LL			1310.560	
Cox and Snell R^2			0.266	
Nagelkerke R^2			0.440	
Hosmer-Lemeshow Test			21.844	
			(0.005)	
Box's M	1912.132			
	(0.000)			
Wilks' lambda	0.730			
	(0.000)			
N	2116		2116	

Note: Parameters estimated on a 2005 sample. These bankruptcy prediction models are used to test hypothesis 1. Statistical significance indicated in parentheses. All sets of variables have been checked for the presence of multicollinearity via the VIF method. No variable within any selection had a VIF>10. Furthermore, in conventional bankruptcy prediction research, it is common to use (1=non-bankrupt, 0=bankrupt as dependent variable setup for discriminant analysis, and 1=bankrupt, 0=non-bankrupt in logistic regression. I have followed this tradition in the estimation procedures, but not in the information content evaluation. In a good model, MDA and logistic regression coefficients should thus in general have opposite signs, though not necessarily.

Appendix G: Complete parameter estimates – hypothesis 2, 3, 4

Each panel shows a different set of predictor variables included in the model. Each column represents a different bankruptcy prediction model: column 2 and column 3 are models estimated with MDA, column 4 and 5 are estimated with logistic regression. Column 3 and column 5 include the two macroeconomic variables (GDP growth rate and inflation rate) as additional predictors in the model. In Panel D, no macroeconomic variables are added, because factor analysis determined the set of predictors. This appendix shows estimates used to evaluate hypothesis 2, 3 and 4.

Table G: 2005-2007 parameter estimates

Estimation technique:	**Multiple discriminant analysis**		**Logistic regression**	
Panel A: Altman (1968) variables				
Constant	0.205	-0.389	-1.833 (0.000)	-0.468 (0.000)
A1.	0.082	0.012	0.103 (0.473)	0.287 (0.058)
A2.	3.072	3.017	-3.208 (0.000)	-3.350 (0.000)
A3.	3.017	2.877	-4.836 (0.000)	-4.638 (0.000)
A4.	0.046	0.038	-0.538 (0.000)	-0.462 (0.000)
A5.	-0.247	-0.226	0.239 (0.000)	0.216 (0.000)
M1.		0.123		-0.254 (0.000)
M2.		0.122		-0.397 (0.000)
-2 LL			4766.382	4628.778
Cox and Snell R^2			0.200	0.217
Nagelkerke R^2			0.330	0.357
Hosmer-Lemeshow Test			114.885 (0.000)	153.447 (0.000)
Box's M	4367.151 (0.000)	5003.362 (0.000)		
Wilks' lambda	0.774 (0.000)	0.762 (0.000)		
N	6712	6712	6712	6712
Panel B: Ohlson (1980) variables				
Constant	-3.069	2.375	-8.636 (0.000)	-6.798 (0.000)
O1.	0.310	-0.303	0.611 (0.000)	0.623 (0.000)
O2.	1.248	-1.143	3.067 (0.000)	3.034 (0.000)
O3.	-0.244	0.155	-0.263 (0.097)	0.000 (0.999)
O4.	-0.118	0.111	-0.402 (0.000)	-0.356 (0.000)
O5.	0.515	-0.537	-0.358 (0.038)	-0.331 (0.062)
O6.	-2.661	2.641	-2.484 (0.000)	-2.334 (0.000)
O7.	0.036	-0.041	-1.299 (0.000)	-1.407 (0.000)
O8.	0.611	-0.612	0.560 (0.000)	0.564 (0.000)
O9.	-0.439	0.423	-0.626 (0.000)	-0.633 (0.000)
M1.		0.093		-0.177 (0.000)
M2.		0.168		-0.808 (0.000)
-2LL			4217.030	4017.680
Cox and Snell R^2			0.248	0.268

Nagelkerke R^2			0.434	0.469
Hosmer-Lemeshow Test			9.507	40.014
			(0.301)	(0.000)
Box's M	6212.125	6797.032		
	(0.000)	(0.000)		
Wilks' lambda	0.724	0.711		
	(0.000)	(0.000)		
N	7491	7491	7491	7491
Panel C: Zmijewski (1984) variables				
Constant	0.752	-0.018	-2.151	-0.129
			(0.000)	(0.522)
Z1.	4.412	4.291	-4.683	-4.764
			(0.000)	(0.000)
Z2.	-1.284	-1.182	1.091	1.065
			(0.000)	(0.000)
Z3.	0.012	0.010	-0.303	-0.312
			(0.000)	(0.000)
M1.		0.104		-0.280
				(0.000)
M2.		0.202		-0.672
				(0.000)
-2LL				5028.314
Cox and Snell R^2			0.164	0.189
Nagelkerke R^2			0.282	0.326
Hosmer-Lemeshow Test			200.501	234.372
			(0.000)	(0.000)
Box's M	4041.497	4536.093		
	(0.000)	(0.000)		
Wilks' lamba	0.801	0.786		
	(0.000)	(0.000)		
N	7657	7657	7657	7657
Panel D: Factor analysis variables				
Constant	2.447		-4.956	
			(0.000)	
F1.	5.264		-6.934	
			(0.000)	
F2.	0.275		-0.418	
			(0.000)	
F3.	-1.128		2.689	
			(0.000)	
F4.	0.035		-0.966	
			(0.000)	
F5.	-4.599		-42.460	
			(0.000)	
F6.	-0.294		0.427	
			(0.000)	
F7.	0.157		-0.318	
			(0.000)	
F8.	0.022		0.366	
			(0.104)	
-2LL			4220.796	
Cox and Snell R^2			0.266	
Nagelkerke R^2			0.445	
Hosmer-Lemeshow Test			72.245	
			(0.000)	
Box's M	5615.794			
	(0.000)			
Wilks' lambda	0.735			
	(0.000)			
N	7020		7020	
Panel E: Alternative prof/solv/liq/act variables (set 1)				
Constant	-1.624	-0.874	-3.695	-1.818
			(0.000)	(0.000)
B1.	-0.187	-0.116	-0.292	-0.103
			(0.000)	(0.223)
B2.	-0.059	-0.058	-0.154	-0.156
			(0.000)	(0.000)
B3.	3.081	2.886	3.440	3.331
			(0.000)	(0.000)
B4.	0.000	0.000	0.000	0.000
			(0.000)	(0.000)
M1.		-0.086		-0.221
				(0.000)

M2.		-0.192		-0.655
				(0.000)
-2LL			4876.752	4656.581
Cox and Snell R^2			0.192	0.217
Nagelkerke R^2			0.321	0.363
Hosmer-Lemeshow Test			54.469	143.007
			(0.000)	(0.000)
Box's M	572.398	1113.106		
	(0.000)	(0.000)		
Wilks' lambda	0.790	0.777		
	(0.000)	(0.000)		
N	6971	6971	6971	6971
Panel F: Alternative prof/solv/liq/act variables (set 2)				
Constant	-0.376	-1.579	-1.041	1. 097
			(0.000)	(0.000)
C1.	0.472	0.301	-0.481	-0.267
			(0.000)	(0.006)
C2,	0.225	0.191	-0.364	-0.356
			(0.000)	(0.000)
C3.	1.286	0.717	-1.414	-0.723
			(0.000)	(0.016)
C4.	0.000	0.000	0.000	0.000
			(0.003)	(0.008)
M1.		0.026		-0.045
				(0.242)
M2.		0.560		-1.048
				(0.000)
-2LL			4474.983	4190.018
Cox and Snell R^2			0.143	0.192
Nagelkerke R^2			0.219	0.293
Hosmer-Lemeshow Test			245.862	220.416
			(0.000)	(0.000)
Box's M	2625.302	3101.934		
	(0.000)	(0.000)		
Wilks' lambda	0.893	0.858		
	(0.000)	(0.000)		
N	4936	4936	4936	4936

Note: Parameters estimated on a 2005-2007 sample. These bankruptcy prediction models are used in the testing of hypothesis 2, 3 and 4. Statistical significance indicated in parentheses. All sets of variables have been checked for the presence of multicollinearity via the VIF method. No variable within any selection had a VIF>10. Furthermore, in conventional bankruptcy prediction research, it is common to use (1=non-bankrupt, 0=bankrupt as dependent variable setup for discriminant analysis, and 1=bankrupt, 0=non-bankrupt in logistic regression. I have followed this tradition in the estimation procedures, but not in the information content evaluation. In a good model, MDA and logistic regression coefficients should thus in general have opposite signs, though not necessarily.

Appendix H: Goodness of fit measures for MDA and logistic regression

In MDA, the key statistic indicating whether or not there is a relationship between the independent and dependent variables is the significance test for Wilks' lambda. Wilks' lambda is the proportion of the total variance in the discriminant scores not explained by differences between the two groups. The smaller the value of Wilks' lambda, the better.

If the means of the independent variables are equal for all groups, the means will not be a useful basis for predicting the group to which a case belongs, and thus there is no relationship between the independent variables and the dependent variable. If the chi-square statistic corresponding to Wilks' lambda is statistically significant we can conclude that there is a relationship between the dependent groups and the independent variables (Hair et al., 2006).

The Box' M statistic test is used to determine whether two or more covariance matrices are equal. If the test statistic shows significance, it means that the null hypothesis of equal covariance matrices is rejected (Hair et al, 2006). That means that the assumption of equal covariances in MDA is violated.

An often-used measure for logistic regression is the Hosmer–Lemeshow (1982) test. It is a statistical test for goodness of fit for logistic regression models and can also be applied to probit analysis, which is similar to logistic regression (see section 2.1.2).

The test statistic is given by

$$H = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{N_g \pi_g (1 - \pi_g)} \qquad \text{Eq. (H.1)}$$

O_g, E_g, N_g, and π_g respectively denote the observed events, expected events, observations, predicted risk for the gth risk group, and G is the number of groups. The test statistic asymptotically follows a chi-squared distribution with $G - 2$ degrees of freedom. The standard test is conducted using deciles, so with $G=10$.

The interpretation is that models for which expected (according to the model) and observed event rates in the subgroups are similar are called well calibrated. Hence, if the goodness-of-fit test is significant, it means that there is a significant difference between the expected values and the data and that the model is not well calibrated.

Furthermore, several R-squared statistics (similar, but not identical to R-squared statistics in OLS regression) are presented as goodness-of-fit values, mentioned in both Hair et al. (2006) and Tinoco and Wilson (2013). Cox and Snell (1968)'s R-squared is a measure based on the log-likelihood of the model, the log-likelihood of the original (baseline) model and the sample size, and Nagelkerke (1991)'s max-rescaled R-squared is a refinement of the former, adjusted for the fact that the Cox and Snell R-squared statistic can never reach one. These are generally thought to be more correct equivalents of the OLS coefficient of determination than a standard pseudo R-squared (Hair et al., 2006).

References

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.

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.

Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, *18*(2), 65-74.

Allen, L., & Saunders, A. (2002). A survey of cyclical effects in credit risk measurement models.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589-609.

Altman, E. I., & McGough, T. P. (1974). Evaluation of a company as a going concern. *Journal of Accountancy*, *138*(6), 50-57.

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

Altman, E. I. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models. *Stern School of Business, New York University*, 9-12.

Ak, B. K., Dechow, P. M., Sun, Y., & Wang, A. Y. (2013). The use of financial ratio models to help investors predict and interpret significant corporate events. *Australian Journal of Management*, *38*(3), 553-598.

Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press.

Åstebro, T., & Winter, J. K. (2012). More than a dummy: The probability of failure, survival and acquisition of firms in financial distress. *European Management Review*, *9*(1), 1-17.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *Neural Networks, IEEE Transactions on*, *12*(4), 929-935.

Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2013). Anomalies and financial distress. *Journal of Financial Economics*, *108*(1), 139-159.

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

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 71-111.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 67-92.

Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, *10*(1), 93-122.

Bell, T., Ribar, G. S., & Verchio, J. (1990). Neural nets versus logistic regression: a comparison of each model's ability to predict commercial bank failures. Auditing Symposium X: Proceedings of the 1990 Deloitte and Touche/University of Kansas Symposium on Auditing Problems. Lawrence, KS: University of Kansas Press.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of Political Economy*, 637-654.

Begley, J., Ming, J., & Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, *1*(4), 267-284.

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.

Blöchlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, *30*(3), 851-873.

Bryant, S. M. (1997). A case‐based reasoning approach to bankruptcy prediction modeling. *Intelligent Systems in Accounting, Finance and Management*, *6*(3), 195-214.

Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: the Turkish case. *European Journal of Operational Research, 166*(2), 528-546.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, *63*(6), 2899-2939.

Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, *13*(3), 465-497.

Charitou, A., Dionysiou, D., Lambertides, N., & Trigeorgis, L. (2013). Alternative bankruptcy prediction models using option-pricing theory. *Journal of Banking & Finance*, *37*(7), 2329-2341.

Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance, 8*(4), 537-569.

Christidis, A. C. Y., & Gregory, A. (2010). Some new models for financial distress prediction in the UK.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data* (Vol. 21). CRC Press.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248-275.

Das, S. R., Hanouna, P., & Sarin, A. (2009). Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance*, *33*(4), 719-730.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 167-179.

Demirovic, A., & Thomas, D. C. (2007). The relevance of accounting data in the measurement of credit risk. *European Journal of Finance*, *13*(3), 253-268.

Dimitras, A. I., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). Business failure prediction using rough sets. *European Journal of Operational Research*,*114*(2), 263-280.

Dionysiou, D., Lambertides, N., Charitou, A., & Trigeorgis, L. (2008). *An alternative model to forecast default based on Black-Scholes-Merton model and a liquidity proxy*. Working paper, University of Cyprus, Department of Public and Business Administration.

Doumpos, M., Niklis, D., Zopounidis, C., & Andriosopoulos, K. (2015). Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from European listed firms. *Journal of Banking & Finance*, *50*, 599-607.

Du, Y., & Suo, W. (2007). Assessing credit quality from the equity market: can a structural approach forecast credit ratings?. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *24*(3), 212-228.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, *83*(3), 635-665.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283-306.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179-188.

Francis, J., & Schipper, K. (1999). Have financial statements lost their relevance?. *Journal of accounting Research*, 319-352.

Frydman, H., Altman, E. I., & KAO, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance*, *40*(1), 269-291.

Grice, J. S., & Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, *54*(1), 53-61.

Grice Jr, J. S., & Dugan, M. T. (2003). Re-estimations of the Zmijewski and Ohlson bankruptcy prediction models. *Advances in Accounting*, *20*, 77-93.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

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.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: a mixed logit model. *The Accounting Review*, *79*(4), 1011-1038.

Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European Journal of Operational Research*, *180*(1), 1-28.

Lacher, R. C., Coats, P. K., Sharma, S. C., & Fant, L. F. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research*, *85*(1), 53-65.

Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*,*10*(4), 511-531.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *Journal of finance*, 1131-1150.

Lee, T. S., & Yeh, Y. H. (2004). Corporate governance and financial distress: evidence from Taiwan. *Corporate Governance: An International Review*, *12*(3), 378-388.

Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, *40*(8), 2941-2946.

Lemeshow, S., & Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, *115*(1), 92-106.

Lyandres, E., & Zhdanov, A. (2013). Investment opportunities and bankruptcy prediction. *Journal of Financial Markets*, *16*(3), 439-476.

Lo, A. W. (1986). Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of econometrics*, *31*(2), 151-178.

Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, *1*(3), 249-276.

McKee, T. E. (2003). Rough sets bankruptcy prediction models versus auditor signalling rates. *Journal of Forecasting*, *22*(8), 569-586.

Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study. *Journal of Accounting Research*, 380-395.

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

Mossman, C. E., Bell, G. G., Swartz, L. M., & Turtle, H. (1998). An empirical comparison of bankruptcy models. *Financial Review*, *33*(2), 35-54.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691-692.

Nam, C. W., Kim, T. S., Park, N. J., & Lee, H. K. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, *27*(6), 493-506.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research, 18*(1), 109-131.

Partington, G., & Kim, M. (2008). Modeling bankruptcy prediction using Cox regression model with time-varying covariates. *Available at SSRN 1101876*.

Peel, M. J., Peel, D. A., & Pope, P. F. (1986). Predicting corporate failure—some results for the UK corporate sector. *Omega*, *14*(1), 5-12.

Pindado, J., Rodrigues, L., & de la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, *61*(9), 995-1003.

Platt, H. D., & Platt, M. B. (1991). A note on the use of industry-relative ratios in bankruptcy prediction. *Journal of Banking & Finance, 15*(6), 1183-1194.

Pompe, P. P., & Bilderbeek, J. (2005). The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business Venturing, 20*(6), 847-868.

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

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model*. The Journal of Business, 74(1), 101-124.

Shumway, T. & Bharath, S., (2004). Forecasting default with the KMV-Merton model. *Unpublished paper, University of Michigan*.

Shumway, T. & Bharath, S., (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies*, *21*(3), 1339-1369.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?. *Accounting Review*, 289-315.

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.

Sun, L. (2007). A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. *Review of Quantitative Finance and Accounting*, *28*(1), 55-78.

Taffler, R. J. (1983). The assessment of company solvency and performance using a statistical model. *Accounting and Business Research*, *13*(52), 295-308.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science*, *38*(7), 926-947.

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394-419.

Trujillo-Ponce, A., Samaniego-Medina, R., & Cardone-Riportella, C. (2013). Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management*, (ahead-of-print), 1-24.

Tsai, C. F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, *22*(2), 120-127.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *The Journal of Finance*, *59*(2), 831-868.

Westgaard, S. (2007). What Can Modern Statistical and Mathematical Techniques Add to the Analysis and Prediction of Bankruptcy?. *Available at SSRN 975295*.

Westgaard, S., & Van der Wijst, N. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European journal of Operational Research*, *135*(2), 338-349.

Wilson, R. L., & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, *11*(5), 545-557.

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

Zhang, G., Y Hu, M., Eddy Patuwo, B., & C Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European journal of operational research*, *116*(1), 16-32.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 59-82.