Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model

Edward I. Altman, New York University, Stern School of Business

Salomon Center, Henry Kaufman Management Center, 44 West Fourth Street, New York, NY 10012, USA

Małgorzata Iwanicz-Drozdowska, Warsaw School of Economics

Institute of Finance, 02-513 Warsaw, Poland

Erkki K. Laitinen, University of Vaasa

P.O.Box 700, FI-65101 Vaasa, Finland

Arto Suvas, University of Vaasa

P.O.Box 700, FI-65101 Vaasa, Finland

Corresponding author: Arto Suvas, asuvas@uva.fi

ABSTRACT

The purpose of this paper is firstly to review the literature on the efficacy and importance of the Altman Z-Score bankruptcy prediction model globally and its applications in finance and related areas. This review is based on an analysis of 33 scientific papers published from the year 2000 in leading financial and accounting journals. Secondly, we use a large international sample of firms to assess the classification performance of the model in bankruptcy and distressed firm prediction. In all, we analyze its performance on firms from 31 European and three non-European countries. This kind of comprehensive international analysis has not been presented thus far. Except for the U.S. and China, the firms in the sample are primarily private and cover non-financial companies across all industrial sectors. Thus, the version of the Z-Score model developed by Altman (1983) for private manufacturing and non-manufacturing firms (Z''-Score Model) is used in our testing. The literature review shows that results for Z-Score Models have been somewhat uneven in that in some studies the model has performed very well, whereas in others it has been outperformed by competing models. None of the reviewed studies is based on a comprehensive international comparison, which makes the results difficult to generalize. The analysis in this study shows that while a general international model works reasonably well, for most countries, with prediction accuracy levels (AUC) of about 75%, and exceptionally well for some (above 90%), the classification accuracy may be considerably improved with country-specific estimation especially with the use of additional variables. In some country models, the information provided by additional variables helps boost the classification accuracy to a higher level.

JEL codes: G15, G32, G33

Keywords: Z-Score, bankruptcy, failure, default, financial distress

Acknowledgements : Erkki K. Laitinen and Arto Suvas are grateful for the financial support of Foundation for Economic Education (Liikesivistysrahasto) and of Jenny and Antti Wihuri Foundation.

Electronic copy available at: http://ssrn.com/abstract=2536340

Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model

1. Introduction

The first multivariate bankruptcy prediction model was developed by E.I. Altman (1968) from New York University in the late 1960's . After this pioneering work, the multivariate approach to failure prediction spread worldwide among researchers in finance, banking, and credit risk. Failure prediction models are important tools for bankers, investors, asset managers, rating agencies, and even for the distressed firms themselves. The banking industry as the main provider of financing in the economy is especially interested in minimizing the level of non-performing loans in order to maximize profit on the credit activity and reduce their own risk of default. Another issue of interest for bankers is capital adequacy and an internal ratings-based approach was encouraged by Basel 2 (first version in 1999, implemented in 2004). The Z-Score model has become a prototype for many of these internal-rate based models. Asset manager investors need to have reliable tools for the selection of companies into their portfolios. Financial distress of the companies is on the one hand detrimental to investor returns, but on the other hand, risk may give opportunities for high returns on short-sale strategies. Rating agencies assess the risk of the entities and securities issues, thus they need to have a tool to predict default. In addition, Altman (1983, 1993 and 2006) has suggested that the management of distressed firms can utilize the Z-Score model as a guide to a financial turnaround.

The approach used for bankruptcy prediction has been evolving over time. Beaver (1966, 1968) used univariate analysis for selected ratios and detected that some of them had a very good predictive power. Altman (1968) moved significantly forward since he developed a multiple discriminant analysis model (MDA) called the Z-Score Model with 5 ratios. The next two decades brought even more financial distress research (e.g. Ohlson 1980, who used the logit model[1], Taffler 1984, who developed a Z-score model for the UK) which was summarized by Zmijewski (1984)[2], who used a probit approach in his own model. Dimitras et al. (1996) reviewed 47 studies on business prediction models (of which 13 were from the US and nine from the UK). They summarized the methods used (discriminant analysis was prevailing) and the variety of ratios used.

The next summary of different approaches to credit risk analysis was given by Altman and Saunders (1998). Balcaen and Ooghe (2006) reviewed models of business failure prediction and classified 43 models presented in the literature into 4 categories (univariate model: 1, risk index models: 2, MDA models: 21, conditional probability models: 19). They omitted, however, the fast growing number of models based on the option pricing theory and contingent claims (e.g. Vassalou and Xing 2004, commercialized into the KMV model) and hazard models (e.g. Shumway 2001). Kumar and Ravi (2007) reviewed 128 statistical and artificial intelligence models for the bankruptcy prediction of banks and firms, with special attention paid to the technique used in different models, pointing out that neural networks were the most popular intelligence technique. Jackson and Wood (2013) presented in their review the frequency of the occurrence of the specific forecasting techniques in the prior literature. The top-five popular techniques were: (1) multiple discriminant analysis, (2) the logit model, (3) neural network, (4) contingent claims and (5) univariate analysis.

Recent valuable reviews on the efficacy of the models have been delivered by Agarwal and Taffler (2008), Das, Hanouna and Sarin (2009) and Bauer and Agarwal (2014), taking into account the

[1] Altman's z-score and Ohlson's o-score have been compared by Dichev (1998).
[2] Re-estimation of Ohlson's and Zmijewski's models was presented by Grice and Dugan (2003).

Electronic copy available at: http://ssrn.com/abstract=2536340

performance of accounting-based models, market-based models and hazard models. These three types of models prevail in the finance literature. According to Agarwal and Taffler (2008) there is little difference in the predictive accuracy of accounting-based and market-based models, however the usage of accounting-based models allows for a higher level of risk-adjusted return on the credit activity. In Das, Hanouna and Sarin (2009) it was shown that accounting-based models perform comparably to the Merton structural, market-based approach for CDS spread estimation. However, the comprehensive model which used both sources of variables outperformed both of them. In Bauer and Agarwal (2014) hazard models that use both accounting and market information (Shumway 2001 and Campbell et al. 2006) were compared with two other approaches: the original Taffler's (1984) accounting based z-score model that was tested in Agarwal and Taffler (2008), and a contingent claims-based model using the Bharath and Shumway (2008) approach. The hazard models were superior in UK data in bankruptcy prediction accuracy (their default probabilities were close to the observed default rates), ROC analysis, and information content.

In spite of the vast research on failure prediction, the original Z-Score Model introduced by Altman (1968) has been the dominant model applied all over the world. Thus, although the Z-Score Model has been in existence for more than 45 years, it is still used as a main or supporting tool for bankruptcy or financial distress prediction or analysis, both in research and pracitice. Our study is focused on this classic model. The purpose of the paper is twofold. Firstly, we review the literature on the Z-Score Model (or its versions Z'-Score for private manufacturing firms and Z''-Score for non-manufacturing and manufacturing firms) applications in order to check its vitality. This review is based on an analysis of 34 scientific papers published from the year 2000 on in leading financial and accounting journals, which have not - according to our knowledge - been presented so far. Secondly, we use a large international sample of firms to assess the classification performance of the Z''-Score model in bankruptcy prediction[3]. In all, we analyze its performance in firms from 31 European and 3 Non-European countries (China, Colombia and the U.S.). Except for the U.S. and two sub-samples (out of three) of Chinese firms, the firms in this study are primarily private. A large number of firms are from non-manufacturing industries. Therefore, we use the version of the model developed by Altman (1983 and also found in Altman & Hotchkiss 2006) for private, manufacturing and non-manufacturing firms (Z''-Score Model). Such an extensive international analysis of the performance of the model in a large number of countries has not been presented thus far. We regard our review and analysis as important contributions to the economic literature.

The paper is structured as follows. In the introduction we present a short summary of the development of failure prediction research starting from Beaver (1966) and Altman (1968). In the second section we summarize the original Z-Score Model (1968) and its extension for private non-manufacturing firms, Z''-Score Model (1983). Then, in the third section we present results and conclusions from the literature review on these models. The fourth section presents seven research hypotheses on the performance of the Z''-Score Model for empirical analysis. In the fifth section, the empirical data and statistical methods are discussed, while the sixth section presents empirical findings. Finally, the seventh section summarizes the study.

2. Classic Z-Score Models

2.1. Z-Score Model for public firms

Altman (1968) criticized prior studies on financial difficulties and said that the adaptation of their results for assessing the bankruptcy potential of firms, both theoretically and practically, is

[3] We use as equivalents: bankruptcy, failure, default and financial distress.

questionable. The dominant methodology was essentially univariate and emphasis was placed on individual signals of impending difficulties. This made the ratio analysis vulnerable to faulty interpretation and potentially confusing analytics. As an appropriate extension, Altman suggested building upon univariate findings and to combining several measures into a meaningful predictive model. Then the question arises which ratios are most important in detecting bankruptcy potential, what weights should be attached to those selected ratios, and how should the weights be objectively established. Altman suggested multiple discriminant analysis (MDA) as the appropriate statistical technique. MDA is a statistical technique used to classify an observation into one of several a priori groupings dependent upon the individual characteristics of observations. For the adaptation of the MDA model, it is crucial how the sample of firms for the two groups of interest, bankrupt and non-bankrupt, and the variables of the model were originally selected.

The initial sample was composed of sixty-six corporations with thirty-three firms in each of the two groups. The bankrupt group (Group 1) consisted of manufacturers that filed a bankruptcy petition under Chapter X of the National Bankruptcy Act during the period 1946 – 1965. The mean asset size of these firms was 6.4 million US dollars, with a range of between 0.7 – 25.9 million US dollars. Altman recognized that this group was not homogenous with respect to size and industry although all firms were relatively small and from manufacturing industries. Therefore, he attempted to make a careful selection of non-bankrupt firms (Group 2). Group 2 consisted of a paired sample of manufacturing firms chosen on a stratified random basis. These firms were stratified by industry and size with the asset size range restricted to 1 – 25 million US dollars. Altman eliminated both the small firms (less than 1 million US dollars in total assets) and the very large firms because of lack of data for small firms and of the rarity of bankrupcies in that period of large firms. He did not match the assets size of the two groups exactly and therefore firms in Group 2 are slightly larger than those in Group 1. The data collected for the firms in both groups were from the same years. For Group 1, the data was derived from financial statements one reporting period prior to bankruptcy. The average lead time of the financial statements was approximately seven and one-half months.

The financial ratios selected for model building were based on balance sheet and income statement data. In past studies, a very large number of variables were found to be significant indicators of financial difficulties. Therefore, Altman compiled a list of 22 potentially important financial ratios for evaluation. He classified these variables into five standard ratios categories: liquidity, profitability, leverage, solvency, and activity ratios. The ratios were chosen on the basis of their 1) popularity in the literature and 2) potential relevancy to the study. The list also included a few “new” ratios. In addition, Altman did not consider cash flow ratios because of the lack of consistent and precise depreciation data. From the original list of 22 financial ratios, Altman selected five ratios for the profile as doing the “best” overall job in the prediction of corporate bankruptcy. This profile did not contain all of the most significant variables measured independently. Instead, the contribution of the entire profile was evaluated. To arrive at a final profile of variables, Altman utilized the following procedures: 1) observation of the statistical significance of various alternative functions including determination of the relative contributions of each independent variable, 2) evaluation of inter-correlations between the relevant variables, 3) observation of the predictive accuracy of the various profiles, and 4) judgment of the analyst.

The final discriminant function estimated by Altman (1968) is as follows:

$Z = 0.012 \cdot X_1 + 0.014 \cdot X_2 + 0.033 \cdot X_3 + 0.006 \cdot X_4 + 0 .999 \cdot X_5$

or

$Z = 1.2X, + 14X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$ (when the first four variables are expressed in decimals, e.g. 0.20, rather than percentages, e.g., 20.0%)

where

X_1 = Working capital/Total assets
X_2 = Retained Earnings/Total assets
X_3 = Earnings before interest and taxes/Total assets
X_4 = Market value of equity/Book value of total liabilities
X_5 = Sales/Total assets
Z = Overall Index

1) The Working capital/Total assets ratio (X_1) is a measure of the net liquid assets of the firm relative to the total capitalization. Working capital is defined as the difference between current assets and current liabilities. When a firm is experiencing consistent operating losses, it will have shrinking current assets in relation to total assets. X1 proved to be the more valuable in analyses than the current ratio and the quick ratio. This ratio explicitly considers liquidity and size dimensions. 2) The Retained Earnings/Total assets ratio (X_2) refers to the earned surplus of a firm over its entire life. This measure of cumulative profitability over time is one of the two (the other is the use of the market value of equity, in X_4, instead of the book value) "new" ratios evaluated by Altman. It considers implicitly the age of the firm due to its cumulative nature and the use of leverage in the firm's financing of its asset growth. 3) The Earnings before interest and taxes/Total assets ratio (X_3) is a measure of the true productivity or profitability of the assets of a firm. It is not affected by any tax or leverage factors. It reflects the earning power of the assets that determines the value of assets. In a bankrupt sense, insolvency occurs when the total liabilities exceed this fair value. 4) The Market value equity/Book value of total liabilities ratio (X_4) shows how much the assets of a firm can decline in value (measured by market value of equity plus debt) before the liabilities exceed the assets and the firm becomes insolvent. This ratio adds a market value dimension to the model. The reciprocal of this ratio – the familiar Debt/Equity ratio – is used to measure financial leverage. Indeed, Altman's use of the market value of equity was the first study utilizing market measures and was in some ways, a predecessor to the so-called structural approach, championed by Merton (1974) and commercialized by KMV. 5) The Sales/Total Assets ratio is the standard capital-turnover ratio illustrating the sales generating ability of the assets of a firm. It refers to the capability of management in dealing with competitive conditions. This ratio was dropped in the Z"-Score model.

Altman evaluated the importance of the five ratios in several ways. Firstly, he used the F test to evaluate the univariate difference between the average values of the ratios in each group to the variability (or spread) of values of the ratios within each group. In this test, variables from X_1 to X_4 were all significant at the 0.001 p-level indicating significant difference in the variables between the groups. However, X_5 did not show any significant difference on a univariate basis. All five ratios indicated higher values for the non-bankrupt group, which is consistent with the positive signs of the discriminant function. Secondly, Altman determined the relative contribution of each variable to the total discriminating power of the function using the scaled vector. In this vector, the profitability measure (X_3) showed the highest contribution while the Sales/Total assets ratio (X_5) gave the second highest contribution although it was insignificant on a univariate basis. To explain this, Altman found a negative correlation (-0.78) between X_3 and X_5 in the bankrupt group. Usually, negative correlations are more helpful than positive correlations in adding new information to the function. Altman explains that this negative correlation will occur when bankrupting firms suffer losses and deteriorate toward failure, and their assets are not replaced as much as they were in better times. In addition, cumulative losses further reduce the asset size through debits to retained earnings. Thus, the asset size reduction apparently dominates any sales movements.

2.2. Z'-Score and Z''-Score Models for private firms

The original Z-Score Model was based on the market value of the firm and was thus applicable only to publicly traded companies. Altman (1983) emphasized that the Z-Score Model is a publicly traded firm model and *ad hoc* adjustments are not scientifically valid. Therefore, Altman (1983) advocated a complete re-estimation of the model substituting the book value of equity for the market value in X_4. Using the same data, Altman extracted the following revised Z'-Score Model:

$Z' = 0.717 \cdot X_1 + 0.847 \cdot X_2 + 3.107 \cdot X_3 + 0.420 \cdot X_4 + 0.998 \cdot X_5$

where

X_1 = Working capital/Total assets
X_2 = Retained Earnings/Total assets
X_3 = Earnings before interest and taxes/Total assets
X_4 = Book value of equity/Book value of total liabilities
X_5 = Sales/Total assets
Z' = Overall Index

Altman did not test the Z'-Score model on a secondary sample due to lack of a private firm data base. However, he analyzed the accuracy of a four-variable Z''-Score Model excluding the Sales/Total assets ratio X_5 from the revised model, because of a potential industry effect. The industry effect is more likely to take place when this kind of industry-sensitive variable (asset turnover) is included into the model. Thus, in order to minimize the potential industry effect, Altman estimated the following four-variable Z''-Score model (Altman 1983):

$Z'' = 3.25 + 6.56 \cdot X_1 + 3.26 \cdot X_2 + 6.72 \cdot X_3 + 1.05 \cdot X_4$

where

X_1 = Working capital/Total assets
X_2 = Retained Earnings/Total assets
X_3 = Earnings before interest and taxes/Total assets
X_4 = Book value of equity/Book value of total liabilities
Z'' = Overall Index

The EBIT/Total assets ratio X3 again made the highest contribution to discrimination power in this version of model. The classification results for the Z''-Score Model were identical to the revised five-variable (Z'-Score) model. In this study, our empirical analysis is focused on the performance of the Z''-Score model in bankruptcy prediction.

In concluding remarks, Altman (1983) regarded the general applicability of his Z-Score Model as debatable. He admited that the model did not scrutinize very large and very small firms, the observation period was quite long (almost two decades), and the analysis included only manufacturing companies. Altman concluded as follows: *"Ideally, we would like to develop a bankruptcy predicting model utilizing a homogenous group of bankrupt companies and data as near to the present as possible."* Therefore, he advised the analysts interested in practical utilization of the Z-Score Model to be careful. This advisement deals with the versions Z'-Score and Z''-Score models of the original Z-Score model as well.

3. Literature review of Z-Score Models

For the literature review, we searched for papers published from the year 2000 on[4] in prominent international journals from SciVerse ScienceDirect, JSTOR and Springer Link, Cambridge Journals and Oxford Journals (see table 1). We selected 33 articles from the below mentioned journals in which the Z-Score was used as a failure prediction proxy or Z-Score methodology was assessed, mostly in terms of predictive ability. The journals in alphabetical order and one handbook were the following: The British Accounting Review (1), Computational Economics (1), Empirical Economics (1), Journal of Accounting Research (2), Journal of Banking and Finance (3), Journal of Business Research (2), Journal of Contemporary Accounting and Economics (1), Journal of Empirical Finance (1), Journal of Finance (4), Journal of Financial Markets (1), Journal of Financial and Quantitative Analysis (2), Journal of Financial Stability (1), Mathematics and Financial Economics (1), Lecture Notes in Computer Science (2), Review of Accounting Studies (3), The Review of Financial Studies (2), Review of Finance (3), Review of Quantitative Finance and Accounting (2) and Handbook of Quantitative Finance and Risk Management (1).. Out of the 33 studies, in 16 cases Altman's Z-Score Model[5] was used as the measure of distress or financial strength (M), in 14 studies Altman's original model was verified and/or modified (V) and in 3 cases it was used for robustness check (R). We focused on this part of research that verified or modified Altman's model.

(Table 1 here)

The wide usage of the Z-Score Model as a measure of financial distress or financial strength in the economic and financial research points out that it is widely accepted as a reasonable, simple and consistent measure of the distressed firm at risk.

In case of the modification of the Z-Score Model, the most important changes were: (1) the use of firms' up-dated financial data in order to re-estimate coefficients and (2) the use of other estimation techniques in order to improve efficacy in comparison to the original model. The use of Altman's ratios combined with other than MDA modeling techniques has improved the prediction ability. Also, the application of new data improved model performance in the case of both US and non-US firms.

In comparison with market-based models or hazard models, Altman's Z-Score Model generally underperformed (4 studies) or provided similar results. In Reisz and Perlich (2007), it was assessed as a better measure for short-term bankruptcy prediction than the market-based models. The question of whether market data is better than acccounting data has been raised many times. The same applies to accounting-based models vs. market-based models (e.g. discussion in Das, Hanouna and Sarin 2009; Bauer and Agarwal 2014). Recall, however, that the original Z-Score model is not solely an accounting-based approach, since the market value of equity is utilized, as well. Our purpose is not to contribute to this strand of studies and we focus on an accounting-based approach, as almost all the estimated models in this study are based on primarily privately held firms and, by definition, there are no market-based data for these firms.

[4] Research devoted to the application of Z-Score Model before 2000 was reviewed by Grice andand Ingram (2001).
[5] Most of the models focused on stock exchange listed firms, thus Z'-Score Model was not used.

4. Research hypotheses

The literature survey shows that the Z-Score Model (publicly traded firms), Z'-Score Model (private firms), and Z''-Score Model (private manufacturing and non-manufacturing firms) have been very widely adapted in different contexts for different purposes. For these kinds of widely used studies, performance plays the key role. In this study, we are firstly interested in assessing the performance of the original Z''-Score model in classifying bankrupt and non-bankrupt firms in an international context. Mainly, our study is focusing on assessing performance in a European context. However, we also validate the results in a set of non-European countries to get a more global view. Secondly, we will re-estimate the model using extensive international data and then use the re-estimated Z''-Score Model as a benchmark assessing the effects of different factors on the performance in terms of classification accuracy. We will assess the effects of the following five factors on this performance: year of bankruptcy, size of firms, age of firms, industry, and country of origin. In all, we shall test a set of hypotheses based on the effects on performance of the model on two different levels. Firstly, we will test a set hypotheses on a pooled set of all firms and, secondly, on data from each country individually. In comparison to the previous research, our contribution consists of the focus on an international context, not just the model application or re-estimation for a given country data. The hypotheses of the study are as follows.

H_1: Obsolescence of the coefficients. The Z''-Score Model was estimated using the same sample of firms that was used to develop the original Z-Score Model. The bankruptcies in the estimation data occurred during the period 1946-1965. Thus, the oldest observations are from almost seventy years ago during the post-war period. Altman (1983) recommended utilizing data as near to the present as possible when developing a bankruptcy prediction model. It is obvious that the financial behaviour of firms and their business environment have significantly changed after that, potentially making the importance of the financial ratios to differ from their original importance reflected by the coefficients of the model. Therefore, we suggest as the first hypothesis (H_1) that the re-estimation of the coefficients of the four original variables of the Z''-Score model will improve the classification performance of the model in an international context. It is supported by the previous research (e.g. Grice and Ingram 2001) and practice. H_1 aims to support this evidence on the international level.

H_2: Method of estimation. The original Z''-Score Model has been estimated using MDA. However, MDA is based on ordinary least squares method (OLS) and thus requires assumptions of multinormality, homoscedasticity, and linearity which are not often met in empirical financial ratio analysis. Therefore, we re-estimate the Z''-Score Model using the logistic regression analysis (LRA) to assess the effect of the method of estimation. LRA does not require most of the restricting assumptions of MDA. In LRA, multivariate normality of the independent variables, homoscedasticity, and linearity are not required. For the sake of OLS, MDA can be more useful than LRA for small samples, such as the original sample of 66 firms used in the estimation of the Z''-Score Model. However, in a large sample LRA may potentially perform better. In this study we use large samples, which is advantageous for LRA. Therefore, our second hypothesis (H_2) suggests that the classification performance of the re-estimated Z''-Score Model will be improved when it is estimated using LRA instead of MDA. The model re-estimated for the original variables using LRA and all pooled data is called here the Z''-Score LR-model. The performance of this re-estimated model is used in this study as the benchmark for further analyses.

H_3: Bankruptcy year. The model based on the relationship between bankruptcy and financial ratios is likely to be affected by the macro-economic environment. These effects may significantly decrease the classification accuracy of the model. If the model is estimated using data from one year and it will be applied to data from another year, the validity of the model can be questioned. Business cycles in terms of economic growth, credit policy, and interest rates can have an effect on the boundary

between bankrupt and non-bankrupt firms. The original Z''-Score Model is estimated using data from the period 1946-1965 which includes several business cycles. Therefore, the model is not focused on any specific stage of cycle and does not explicitly take account of the bankruptcy year. Altman (1983) suggested gathering data from firms for the last couple of years when developing a prediction model. In this study, the benchmark Z''-Score LR-model will be estimated for a shorter period than in the original estimation which however covers several recent years on different stages of the business cycle in different countries. Therefore, the third hypothesis (H_3) assumes that the classification accuracy of the benchmark model can be increased by explicitly taking account of the year of bankruptcy in the estimation of the model.

H_4: Size of the firm. The boundary between bankrupt and non-bankrupt firms is different for small and larger firms which decreases the performance of the model estimated for data from one size category and applied for data from another size category. For the bankrupt and non-bankrupt firms in the original data for Z''-Score Model estimation, the range of asset size was about between 1 – 25 million US dollars. The data did not include very small and very large firms. Altman (1983) regarded the suitability of the original Z-Score Model (and in the same way Z''-Score Model) for all firms as debatable because it did not scrutinize very large or very small firms. In this study, the benchmark Z''-Score LR-model will be estimated for data from many size categories from very small firms to very large firms. The fourth hypothesis (H_4) assumes that the classification performance of the uniform benchmark LR-model based on the original four financial variables of the Z''-Score Model is increased when the size category of the firm is explicitly taken into account.

H_5: Age of the firm. International insolvency statistics generally shows that bankruptcy risk is a function of the age of the firm. Especially, very young firms typically show a very high risk. The original Z''-Score Model does not take explicitly account of the age. However, Altman (1983) noted that the age of a firm is implicitly considered in the Retained Earnings/Total Assets ratio (X_2) that was regarded as a new ratio in bankruptcy prediction context. A relatively young firm will probably show a low ratio because it has not has time to build up cumulative profits. Thus, a young firm is to some degree discriminated against in the model and its likelihood to be classified as bankrupt is relatively higher than that of an older firm. For this argument, Altman (1987) concluded: "*But, this is precisely the situation in the real world.*" The incidence of failure is much higher in the early years of a firm. In spite of the fact that the age of the firm in this way is implicitly taken into account in X_2, we expect that an explicit consideration of the age will improve the classification accuracy due to controlling for the age factor. Therefore, the fifth hypothesis (H_5) proposes that the performance of the uniform benchmark model based on the original four financial variables of the Z''-Score Model is increased when the age of the firm is explicitly taken into account.

H_6: Industry of the firm. The original Z'-Score Model is estimated only for manufacturing firms. Altman (1983) stated that ideally we would like to develop a bankruptcy prediction model utilizing a homogenous group of bankrupt firms. If we are interested in a particular industry grouping, we should gather data from bankrupt and non-bankrupt firms in these groupings. Previous studies show that the financial distress analysis is influenced by industry effect (Smith and Liou 2007). Firms in different industries tend to report different levels of the same financial ratios which may have an effect on the boundary between bankrupt and non-bankrupt firms. This industry effect may be present in the Z'-Score Model especially due to the Sales/Total Assets ratio (X_5). This ratio showed the least significance on a univariate basis but had a very significant contribution to the discriminant power of the multivariate model. Altman (1983) recognized the potential industry effect due to a wide variation among industries in asset turnover, and specified the Z''-Score Model without X_5 for private non-manufacturing firms. However, the Z''-Score Model is also estimated using the original sample of manufacturing firms. In our analysis, the uniform benchmark model based on the original four financial variables of the Z''-Score Model is estimated for a statistical sample representing different industries.

Therefore, the sixth hypothesis (H_6) assumes that an explicit consideration of industry will improve the classification accuracy of this benchmark model.

H_7: Country of origin. The original Z''-Score Model has been estimated only for U.S. firms. However, in previous studies, the Z''-Score model has been applied in countries all over the world. It can be expected that the international applicability of the model to other countries is affected by country-specific differences. Economic environment, legislation, culture, financial markets, and accounting practices in a country may affect the financial behaviour of firms and the boundary between bankrupt and non-bankrupt firms. These factors may potentially weaken the classification performance of the model in other countries outside the country in which the model is originally estimated (Ooghe and Balcaen 2007). Therefore, the seventh hypothesis (H_7) assumes that taking account explicitly of the country of origin of a firm will improve the classification accuracy of the benchmark model. In our empirical study, the country effect will be assessed by including a variable of country risk.

5. Empirical data and statistical methods

5.1. Sample of firms

The principal data of this study are extracted from the ORBIS databases of Bureau Van Dijk (BvD). The main data are from ORBIS Europe that is a commercial database which at the moment of sampling contained administrative information on over 50 million European firms. However, income statement and balance sheet information was available for about 8 million companies. More than 99% of the companies covered in this database are private companies from different industries, justifying the use of the Z''-Score Model (for private manufacturing and non-manufacturing firms) instead of the original Z-Score Model (for publicly traded manufacturing firms). The ORBIS database organizes the public data from administrative sources and filters them into various standard formats to facilitate searching and company comparisons. The ORBIS formats have been derived from the most common formats used for the presentation of business accounts in the world (Ribeiro et al., 2010). It is clear that international comparability may be a problem when administrative firm-level data are internationally pooled. For administrative data, the definition of variables is usually less harmonized. This is less of a problem in the ORBIS database because of the common international format of balance sheets. For example, although some discrepancies in profit/loss statements may arise because of differences in fiscal systems across countries, balance sheet variables largely adhere to international standards. Therefore, ORBIS provides us with a useful and extremely large database for our study.

For statistical sampling, several requirements are set for the empirical data. Firstly, we require that the firm to be selected must be an industrial (non-financial) company. Secondly, its owners must have limited liability (whereby partnerships and sole proprietors are left out of the study). Thirdly, we set a minimum requirement for the size of the firm. Because financial ratios in very small firms are generally too unstable for a failure prediction model, these firms are excluded (Balcaen and Ooghe 2006). We require that the Total Assets must have exceeded 100 thousand EUR at least once in the available time series for a firm. Fourthly, we include in our estimation sample firms from all European countries where the number of failed firms is more than 60. If the number of failed firms for a country is less than 60, the firms from this country are only included in the test sample. For qualifying European countries, the failed firms are randomly classified into the estimation and test samples so that the number of sample firms is about equal in both samples. In all, our estimation data include firms from 28 European and 3 Non-European countries. Fifthly, all failed firms that fulfill the requirements above are included into our samples. However, if the number of non-failed firms in a country is very high, a sample is randomly selected from that country. Finally, the time span of fiscal years potentially available for this study ranges from 2002 to 2010. Because the last financial statements for failed firms in the database are from a financial period within 2007 and 2010, earlier years are excluded, for comparability, also for non-failed firms. But all qualifying observations of non-failed firms from years 2007 to 2010 are included in the data sets. As for failed firms, we restrict the analyses to the last financial statements available before failure. The four independent variables of the Z''-Score Model were winzorized at 1% and 99% to minimize outliers.

Our data is not restricted to European countries only. The results are estimated and tested also for three non-European countries (United States, China, and Colombia) to get a more global view of the performance of the Z''-Score Model. Therefore, a sample of firms from these countries is randomly selected as for the estimation and test data from ORBIS World including middle-size (total assets over 1.5 Million Euro) and larger firms from all over the world. The main principles for selecting these data follow the description outlined above for European firms. Table 2 shows the resulting number of non-failed and failed firms in the estimation data and test data by country. In all, the estimation sample includes data from 2,602,563 non-failed and 38,215 failed firms from 28 European and 3 non-European countries. The test sample is slightly larger because it includes data from 31 European and 3 non-European countries. From the country of origin of the Z''-Score Model, the U.S., the estimation sample only includes 56 bankrupt firms. The U.S. data that was available to us consists only of listed (and delisted) firms. From China, there are three sub-samples. Public firms with Special Treatment (ST)[6] failure status are included in the estimation sample although there are only 32 such firms altogether.[7] The Chinese datasets of predominantly private firms (CN) and of public firms with delisted (DL) failure status are separately analysed only in the test data.[8] ST firms are listed firms suffering from serious financial difficulties. Delisted firms are firms delisted from the stock enchange. Excluding the special U.S. and the two non-private Chinese datasets, 99.4% of observations in the data are private firms.

(Table 2 here)

5.2. Status of failed firms

ORBIS has five classes for potentially active firms (active; default of payment; receivership; dormant; branch) and seven classes for inactive firms which do not carry out business activities anymore (bankruptcy; dissolved; dissolved (merger); dissolved (demerger); in liquidation; branch; no precision). From these classes, only active is selected to represent non-distressed firms. In selecting the failed firms, we try to avoid ambiguity as much as possible by considering (with exceptions described below) a firm failed if its status in ORBIS is stated as bankruptcy. However, because of the small number of bankrupt firms in some countries, we also consider receivership (active) firms failed although they are active. These firms generally suffer from serious financial distress. However, liquidation (inactive) firms are not included in the sample of failed firms (with one exception). Firms in liquidation may, depending on the country, contain firms that have ceased activities due to reasons other than failure (mergers, discontinuing the operations of a daughter company or of a foreign branch, etc.). Therefore, for most countries, we select only firms that are coded as being bankrupt or under receivership. However, there are a number of special countries where failed firms are coded under a different status heading. These special countries or samples are the following:

[6] See Zhang, Altman, and Yen (2010) for the rationale for using special-treatment firms as a proxy for bankruptcies. These are firms put on probation by the stock exchange for poor operating performance and/or negative equity.

[7]This is done because the results about predictability were good also for such a small sample.

[8]These firms are included only in the test data because the predictability of failure was exceptionally poor.

Country:	Status categories:
Bulgaria	In liquidation, Bankruptcy
Denmark	Inactive (no precision)
Greece	Active (receivership), In liquidation, Bankruptcy
Ireland	In liquidation, Active (receivership)
Norway	In liquidation
Slovenia	In liquidation
Spain	Active (receivership), In liquidation, Bankruptcy
Ukraine	In liquidation, Bankruptcy
United Kingdom, liquidation	Inactive (In liquidation)
China, ST	Active (Special Treatment)
China, delisted, DL	Active (delisted)

In case no such category for failed firms could be identified, that country was excluded from the study (for example, Switzerland). If there was only a very small number of failed observations, the country was dropped from the study (Luxembourg, Liechtenstein, Montenegro, typically small countries). It should also be noted that the status classes (including the bankruptcy category) are not completely homogenous within European countries due to different legislations, although there are obvious similarities in insolvency acts (Philippe et al., 2002). China is a special case including samples with three different criteria of failure (bankruptcy, Special Treatment, delisted). For the UK, there are two different samples (liquidation, receivership).

5.3. Statistical methods

In this study, seven research hypotheses are drawn for statistical testing. The statistical analysis will begin with calculating the original Z''-Score for the firms in the data. Following the original model, this Z''-Score will be calculated for all sample firms as follows:

$$Z'' = 3.25 + 6.56 \cdot X_1 + 3.26 \cdot X_2 + 6.72 \cdot X_3 + 1.05 \cdot X_4$$

where

X_1 = Working capital/Total assets
X_2 = Retained Earnings/Total assets
X_3 = Earnings before interest and taxes/Total assets
X_4 = Book value of equity/Book value of total liabilities

The classification performance of the original model is assessed by the AUC (Area Under Curve) measure extracted from the ROC (Receiver Operating Characteristic) curve. ROC curve is a plot of true positive rate against false positive rate for all different possible cut-off-points. These profiles show the trade-offs between Type I and Type II errors and represent statistically the cumulative probability distribution of failed events. AUC measures the accuracy of the estimated model in relation to the perfect model. With a perfect model AUC is 1, and with a random model 0.5. AUC has a close connection with the Accuracy Ratio (AR) since $AR = 2 \cdot AUC - 1$. AR equals 0 for a random model and 0.5 for a model with an average classification performance. In all statistical analyses, SAS software is used.

The first hypothesis (H_1) assumes that the coefficients of the original model are obsolete. H_1 is tested by re-estimating the coefficients of the Z''-Score Model using the original statistical method (the multiple discriminant analysis or MDA). In MDA, the discriminant function is determined by a parametric method (a measure of generalized squared distance) and the distribution of independent variables within both groups is assumed to be multivariate normal. The purpose is to estimate the new coefficients for the model to statistically represent the overall sample. The problem is that the estimation sample includes different numbers of failed and non-failed firms from 31 countries. In the original sample of Altman (1983), an equal number of bankrupt and non-bankrupt firms were selected from one country (U.S.). Following the characteristics of these data, we weight the firms so that the weights for the failed and non-failed firms are equal. In this way, the non-proportional sampling in different countries does not affect the re-estimated model. The number of firms from different countries however varies significantly which leads to greater weights for larger countries. To avoid this problem, the observations are also weighted so that each country has an equal weight in the analysis. Then, the coefficients of the Z''-Score Model are re-estimated using these weighted data and the resulting AUC is compared with the one based on the original model.

The second hypothesis (H_2) expects that the classification performance of the re-estimated Z''-Score Model improves when it is re-estimated by the logistic regression analysis (LR) that is based on less restrictive statistical assumptions than MDA. In this estimation, the dependent variable Y = 0 when the firm is non-failed and Y = 1 when it is failed. LRA does not require that independent variables are multivariate normal or that groups have equal covariance matrices which are basic assumptions in MDA (Hosmer and Lemeshow, 1989). LRA creates a score (logit) L for every firm. It is assumed that the independent variables are linearly related to L. This score or logit is used to determine the conditional probability of failure as follows:

$$p(Y=1|X)=\frac{1}{1+e^{-L}}=\frac{1}{1+e^{-(b_0+b_1X_1+..+b_4X_4)}}$$

where b_i (i =0,…, 4) are the coefficients and X_i (i =1,…, *4*) are the four independent variables of the original Z''-Model. The effect of the method on classification performance is assessed by testing the statistical significance of the difference between AUCs for this LR model and for the re-estimated MDA model. The resulting model is here called the Z''-Score LR-Model and it is used as a benchmark for further statistical AUC comparisons, because LR is applied as the principal method in testing of the rest of the research hypotheses.

The third hypothesis (H_3) is associated with the performance effect of taking account of the bankruptcy year in estimation. This hypothesis is tested estimating a LR model based on the following logit:

$$L=b_0+\sum_{i=1}^{4}b_iX_i+\sum_{j=1}^{3}c_jD_j$$

where b_0 is a constant, X_i (i =1,…, 4) are the four independent variables of the original Z''-Model, b_i (i =1,…, 4) are their coefficients, c_j (j =1,…, 3) are coefficients of the dummy variables and:

D_1 = 1 when year = 2008, 0 otherwise

D_2 = 1 when year = 2009, 0 otherwise

D_3 = 1 when year = 2010, 0 otherwise

The dummy variables do not directly refer to the bankruptcy year that is not given in the data base but to the last available year. For failed firms, there is about 1—2 years lead time to failure from this year. In this model, year 2007 is the base category. If AUC of this extended LR model statistically significantly exceeds AUC of the Z''-Score LR-Model (benchmark), evidence supports hypothesis H_3.

The research hypotheses H_4-H_7 are statistically tested using the same approach as for the third hypothesis above. However, for each hypothesis appropriate variables are used instead of the year dummies. Hypothesis H_4 deals with the performance effect of taking account of size and it will be tested using in the LR model two additional variables of size. In this LR model, size is measured by natural logarithm of total assets and its squared form. In this way the effect of logarithmic size can be reflected by a function following the second-order parabola. Hypothesis H_5 predicts that taking explicitly account of the age of the firm improves classification performance. When testing this hypothesis, the category 6-12 years is used as the base category and two dummy variables are incorporated in the LR model (D_1: 0-6 years, D_2: 12- years). Hypothesis H_6 is associated with the effect of the explicit consideration of industry on the classification performance. It is tested here using dummy variables for seven industries (D_1: restaurants and hotels, D_2: construction, D_3: whole sale and retailing, D_4: agriculture, D_5: manufacturing, D_6: energy and water production, D_7: information technology) all other industries acting as the base category.

Hypothesis H_7 predicts that the explicit consideration of the country of origin improves classification performance. This hypothesis is tested not using dummy variables for countries but using country risk measures instead. The country risk of each country is measured by Standard & Poor's Country Risk Rating per six months after annual closing of accounts. The rating is numerically recoded in the way that the best rating AAA = 1, the second best rating AA+ = 2, and et cetera. Finally, the lowest rating D = 22. Thus, H_7 is tested estimating a LR model based on the four financial ratios of the original Z''-Score Model and a 22-step variable referring to country risk. The five LR models with the orginal four financial ratios and the additional variables specified in the hypotheses are estimated for all data. In addition, a LR model including all additional variables is estimated for all data to assess the simultaneous effect of all variables. Finally, six of the seven hypotheses are tested for the data of each country separately. In this country level testing, hypothesis H_7 is not tested because the additional variable (country risk) is constant within the country.

6. Empirical results

6.1. All data: coefficients of the Z''-Score Models

Table 3 presents descriptive statistics of the four independent variables (X_1-X_4) of the Z''-Score Model for all test data. The variation in the ratios is significant as is shown by the standard deviation and the quartiles. For X_1 (WCTA), X_2 (RETA), and X_3 (EBITTA) the median and the mean for non-failed firms are close to each other referring to a symmetry of distributions. However, this is not the case for the failed firms. For the failed/distressed firms, the median exceeds the mean for these three ratios, referring to negatively skewed distributions. For X_4 (BVETL), the means significantly exceed the median for both failed and non-failed firms, indicating a positively skewed distribution. For each of the four variables both the mean and the median are higher for non-failed firms than for failed firms which is consistent with the expectations. The difference between the means of non-failed and failed firms is larger in the original U.S. data than in our all data for RETA and EBITTA but about the same size for WCTA and BVETL (Altman 1983). These characteristics of the data may indicate lower classification accuracy than in the original sample.

(Table 3 here)

Table 4 presents the coefficients of the different models estimated for all data. All LRA estimates (Model 2 to Model 9) are statistically significant at 0.001 due to their contributions and the large sample size. The first column presents the coefficients of the original Z''-Score model. Column "Model

1” shows the coefficients when they are re-estimated by the same statistical method or MDA. The coefficients are here negative, since the models are estimated using Y = 1 for the failed firms. In our all data, EBITTA has a significantly higher weight than in the original U.S. data while the weights of WCTA and BVETL have proportionally decreased. The re-estimated coeffient of BVETD is very small referring to a minor effect on the logit. Column “Model 2” presents the coefficients for the Z’’-Score LR-Model. These coefficients are directly comparable with those of the MDA model, which is expected for the sake of the exceptionally large sample. The differences in the coefficients of the original four variables between the eight LR models (Models 1-8) are small. For each model, the coefficient of BVETD is very close to zero. This indicates that the original four variables and the additional variables are quite independent of each other.

(Table 4 here)

Table 4 also shows the coefficients of the additional variables, beyond the original four, in the LR models. The negative coefficients of the dummy (year) variables of Model 3 indicate that after 2007 (the base category), this risk of failure has significantly decreased year by year. The base year of 2007 refers to a failure emerged during 2008-2009. This phenomenon is caused by the global financial crisis. The crisis played a significant role in the failure of key businesses and a downturn in economic activity leading to the 2008–2012 recession, especially in Europe. The coefficients of Model 4 for the size variables show that the contribution of size on the logit (risk measure) is at its maximum value when logarithmic total assets is 15 or when total assets is about 3.3 million Euro. Model 5 confirms the riskiness of young firms, since the risk to fail is very high for newly founded firms (less than 6 years old) as is shown by the coefficient of the first dummy variable. The coefficients of the industry dummies in Model 6 show that construction is an exceptionally risky industry followed by manufacturing. For Model 7, the coefficient of the country risk dummy is statistically significant (for the sake of the large sample) but negative and very close to zero. Finally, the coefficients of all variables in Model 8 are directly comparable with those in Models 3-7.

6.2. All data: performance of the Z’’-Score Models

Table 5 shows the resulted AUCs in the test data for the different “all data” models, by country. Model 1 refers to the original Z’’-Score Model. The classification performance of the score at the level of all countries is fair, since AUC = 0.743 refers to AR = 0.486 that is about average accuracy (0.5). However, the score gives relatively good results (AUC > 0.8) for China (ST firms), Poland, Finland, Estonia, Czech Republic, Croatia, Bosnia, Russia and Sweden. Its performance is quite low (AUC < 0.7) for Norway, Latvia, Iceland, Ireland, and Germany. The lower part of the table shows the resulting AUCs for the countries only included in the test data. The performance of the score is very low in the Chinese CN (primarily private) and DL (delisted) samples. It is also low for the liquidation firms in the United Kingdom.

Appendix 1 shows the medians of the four ratios (X_1 to X_4) by status and country. Table 6 presents the differences of these medians between the non-failed and failed firms by country. In this table, also the AUC of the Z’’-Score and its correlation to the difference of medians are presented. This correlation is high for each financial ratio showing that the effects of the ratios on AUC are well balanced. For China’s ST firms, the differences are not exceptionally large except for EBITTA. This implies, with the exceptionally high AUC, that the ST firms systematically differ from the non-ST firms although the differences are not extremely large. The differences between the medians are very large in Poland for each ratio justifying high AUC. They are also large in Finland, and also in Czech Republic where the difference in EBITTA is however average. In Germany, Latvia, China (CN and delisted) and UK (liquidation) the differences in all four ratios are below the average which is obviously associated with low AUC. In the Chinese delisted firms sample, the differences in RETA and EBITTA are even negative. In Iceland and Ireland, the differences only in EBITTA are exceptionally small.

(Table 5 here)
(Table 6 here)

Model 2 in Table 5 is the re-estimated Z''-Score Model where the coefficients are estimated by MDA for all data. Its AUC (0.745) is only slightly higher than for the original model (0.743), supporting only very weakly, if at all, H_1 (obsolescence of the coefficients). The classification accuracy AR (0.490) is about the average level. However, the re-estimation of the coefficients has led to improved classification accuracy in a number of countries, especially in Bosnia, China (ST), Norway and Greece. However, it has impaired classification accuracy in UK and China (delisted). Column "Benchmark" reports the results for the benchmark model (Z''-Score LR-model) showing the effect of the estimation method. For the benchmark model, AUC in all data is 0.748, that is higher than for Model 1 and 2. The differences between AUCs are very small supporting only weakly H_2 (estimation method). The LR-model (benchmark) and the MDA model (Model 2) give almost identical AUCs for each country. This result was expected, since the coefficients of the models are directly comparable. The similar results for the models may also indicate that the independent variables conform to multinormality. This last result confirms what most researchers in the field of default classification models have concluded, that accuracy levels are extremely similar, despite any statistical violations, like normality, between MDA and logistical regression models.

Model 3 (LR-model with year dummies) leads in all test data to higher AUC (0.752) than the benchmark model, supporting H_3 (bankruptcy year effect). However, the AUC effects are not positive for all countries. The effects are positive for example, for Russia, Estonia, Germany, Ireland, and Latvia, but these effects are not statistically significant. There are statistically significant negative effects for China (ST) and Serbia. There are notable negative effects especially for Bosnia, Croatia, Hungary, Norway, and Slovenia. These diverse results are due to exceptional annual distributions of failed firms in these countries. For the countries with negative effects, the percent of failed firms from 2010 ($D_3 = 1$) exceeds 50%, while for the countries with positive effects it is only a couple of percent. For each group, non-failed firms are quite equally distributed over years. When the coefficient of D_3 is very low (-0.666), it strongly decreases the risk estimates of most failed firms in the former countries but only in few failed firms in the latter countries. This leads to the observed effects.

Model 4 (LR-model with size variables) performs better than the benchmark model which gives support to H_4 (size effect). It leads to AUC = 0.760 referring to AR = 0.520. However, it leads to significant improvements in AUC for example in China (delisted) and Austria. For China (delisted), the increase in AUC is extremely strong. There are also improvements in AUC for instance for Estonia, Italia, Slovakia, Spain, and UK. Model 4 has also led to lower AUCs in few countries but the decrease is not significant. Model 4 is based on the four original variables and the size effect following a second-order parabola. This kind of size effect is very small for micro firms but increases approaching its maximum value in middle-sized firms. The countries with a positive change in AUC have typically data where the percent of failed micro firms (in all failed firms) is relatively low (40-60%) while there are a lot of middle-sized failed firms. Because the size effect is strongest for middle-sized firms, AUC will increase. On contrary, the countries with a negative change in AUC have data where the percent of failed micro firms is exceptionally high (70-80%).

Model 5 (LR-model with age category dummies) gives for all test data about the same AUC as the benchmark model (AUC = 0.748). Thus, empirical evidence does not support H_5 (age effect). For each country, the effect of the age on AUC is small. For Austria, the effect is however positive and notable. For this country, the percent of non-failed young firms (0-6 years) is only about 10% (of non-failed firms) whereas this percent for failed firms is more than 20% (of failed firms). Because Model 5 includes a strong positive age risk effect for young firms (D_2), it increases the risk of many failed firms but that of only few nonfailed firms. This obviously leads to an improvement in AUC.

Model 6 (LR-model with industry dummies) outperforms the benchmark model in AUC supporting H_6 (industry effect). It gives AUC = 0.751 leading to AR = 0.502. However, its AUC is notably higher than the benchmark AUC only in a couple of countries, for example in France, Latvia, Portugal, and Spain. Model 6 makes a negative effect on failure risk for example in restaurants and hotels and information technology industries but a positive effect in construction and manufacturing industries. For the countries with a positive effect on AUC, the percent of nonfailed firms in restaurants and hotels and information technology is high while that of failed firms is low. For the risky industries (construction and manufacturing), these distributions are reversed. Thus, Model 6 gives a positive (negative) risk effect for many failed (nonfailed) firms and a negative (positive) risk effect only for few failed (nonfailed) firms. Therefore, AUC will be increased. This situation is reversed for the countries with a negative effect on AUC (Austria and Slovenia). The samples of these countries include a high percent of nonfailed firms in the manufacturing industry, leading to a decrease in AUC.

Model 7 (LR-model with country risk measure) leads to a marginally higher classification performance (AUC = 0.749) than the benchmark model. This result gives only very weak support for H_7 (country of origin effect). However, for each country, the resulted AUC is almost identical with that given by the benchmark model. This result was expected due to the neglible coefficient (-0.003) of the country risk measure (SP Country rating rank) in Model 7. This result implies that country risk does not make any effect on the boundary between bankrupt and non-bankrupt firms.

Model 8 (LR-model with all variables) includes the four financial ratios and all additional variables. It leads in all test data to a sizeable increase in AUC (AUC = 0.771) in comparison to the benchmark AUC. However, the effect on AUCs largely varies and is either negative or positive in different countries. The effect is large in several countries, for example in Estonia, France**,** Iceland, Italy, Latvia, and China (delisted). However, it also makes a negative effect on AUC in several countries, such as Bosnia, Hungary, and Norway. These results show that the inclusion of additional variables into the original model will usually increase AUC, but not in every country.

The "all data" benchmark performs fairly well also for the U.S. and the Colombian samples (the U.S. firms being, unlike the the majority of other firms in this study, listed or delisted companies). The poor performance of Chinese predominantly private (CN) and delisted (DL) firm samples is associated with very small differences between the medians of the non-failed and failed groups, as is shown in Table 6. It is clear that the status "delisted" is not comparable with "bankruptcy". When the status is defined as "ST", the predictability of Chinese listed firms is very high. Prior studies based on Chinese ST-firms have also demonstrated good predictability (Zhang et al., 2010; Wang and Campbell, 2010), see footnote 6, earlier. Nevertheless, this puzzle calls for additional research and modeling work regarding unlisted and delisted Chinese firms.

6.3. Country-level data: performance of the Z''-Score Models

The heterogeneity of the firms and their distributions in "all data" makes it difficult for a uniform all data model to increase AUCs in different countries. Table 7 presents the test data AUCs for the different models estimated for each country separately (country-level models). In this table, the "all data" Z''-Score LR-model acts as the benchmark. When the models are estimated from the country data, this benchmark will be clearly outperformed by the resulting MDA (Model 1) and LR models (Model 2) only in a couple of countries (Bulgaria, France, Latvia, Spain, and Sweden). These results give, however, only weak support for H_1 (obsolescence of the coefficients) at the country level because the effects are not significant. In addition, the benchmark leads to higher AUCs than Models 1 and 2, at least in Austria, Bosnia, Ireland, Slovenia, and United States. The differences in AUCs given by Models 1 and 2 are generally small. In Romania only, Model 1 is clearly outperformed by Model 2. Thus, country-level evidence does not support H_2 (estimation method).

Model 3 (LR-model with year dummies) leads in most countries to clearly higher AUC than the benchmark model. This evidence gives support to H_3 (bankruptcy year effect). Model 3 leads to lower AUC than the benchmark model in Austria, Bosnia, Slovenia, and United States. Model 4 (LR-model with size variables) leads to improved performance almost in every country supporting H_4. This improvement is significant however only in the United States where AUC is now 0.816 when it is only 0.710 for the benchmark model. It is notable also in Bulgaria, France, Latvia, Spain, and United Kingdom. Model 5 (LR-model with age category dummies) gives in several countries higher AUCs than the benchmark model but none of the improvements are significant. The positive effect is strong especially in Bulgaria, France, and Iceland. Although there are also negative effects on AUC, this evidence weakly supports H_5 (age effect) since these negative effects are relatively small.

Model 6 (LR-model with industry dummies) shows both negative and positive effects on AUCs when compared with the benchmark. The only significant effect, however, is the negative effect found in Austria. In Bulgaria and Slovenia, AUC has also notably decreased due to the industry dummies. However, there are notable positive effects on AUC, for example in France, Iceland, Latvia, and Sweden. Thus, the effect is not systematic and gives only weak support to H_6 (industry effect). Model 7 (LR-model with all variables) leads to a remarkable increase in AUC as compared with the benchmark model in Colombia, Iceland, and Unites States. It also leads to notable improvements in AUC at least in Belgium, Denmark, Estonia, France, Germany, Hungary, Latvia, Norway, Spain, and United Kingdom. There are found negative effects on AUC only in Austria and Slovenia. Thus, as a conclusion we can state that the classification performance of "all test" data Z''-Score Model can in general remarkably increase in most countries when different effects are taken into account by additional variables. This increase is found in most European countries and also in Colombia and United States. In China, the AUC for the ST sample is extremely high for the Z''-Score LR-model and it can only slightly be improved by additional variables.

(Table 7 here)

7. Summary of the study and a suggested extension

The purpose of this study was to assess the classification performance of the Z''-Score Model originally introduced by Altman (1983). This model was originally applied for small and middle-sized US manufacturing bankrupt and non-bankrupt firms but is applied widely all over the world in different industries and for different size categories for different purposes. It is a modified version of the Z-Score Model (1968) developed for listed manufacturing firms. The Z''-Score Model is modified to apply for private and non-manufacturing firms as well as manufacturers.

The literature review section summarizes recent articles in prominent academic journals that have utilized Altman's Z-Score or Z''-Score models, or re-estimated versions of them, in empirical analyses. These models are typically used as benchmarks in failure prediction modeling studies, where one or several alternative methods or approaches (hazard models, contingent-claims, intelligent algorithms etc.) have been tested. However, in a considerable number of the reviewed studies failure prediction is not the primary focus. Instead, these models have been largely used as measures of financial strength. As to the failure prediction studies, the results have been somewhat uneven so that in some studies the models have performed well, whereas in others they have been outperformed by competing models. None of the reviewed studies were based on comprehensive international comparisons, which makes the results difficult to generalize.

In this study, the classification performance of the Z''-Score model is assessed using very large data sets in an international context. The purpose is to test how the original version of the Z''-Score Model performs in different countries, and how re-estimation, using another statistical method, and different additional variables, affect the classification performance when the data are very heterogeneous. For this kind of testing in an international context, seven research hypotheses on classification performance are extracted. These research hypotheses are tested for all data and also separately for country data (country-level analysis). The hypotheses are associated with the following effects on the classification performance of the Z''-Score Model: 1) re-estimation of coefficients, 2) estimation method, 3) year, 4) size, 5) age, 6) industry, and 7) country. The estimation data are from 29 countries and the results are validated for 34 countries. The countries are mainly from Europe, but also three non-European countries are included (China, Colombia, and United States). The status used in the classification is mainly bankruptcy/active firms but also receivership firms are considered failed. In Chinese data, also ST (special treatment) and delisted firms are separately analysed as failed firms.

The analyses at the level of all data show that the original Z''-Score model performs very satisfactorily in an international context. The effects of the four financial ratios on performance are well balanced although Book Value of Equity / Total Liabilities (BVETD) showed a very small contribution in re-estimation. The original model performs very well in several countries, such as Poland, Finland, and China (ST firms). The re-estimation of the coefficients using MDA only marginally improved classification performance supporting weakly the obsolescence hypothesis (H_1) or, to put it differently, shows that the original coefficients are extremely robust across countries and over time. This same conclusion holds for the re-estimation of the model using LRA, since the performance results are very similar as for MDA (H_2). The use of additional variables in the model generally improved classification accuracy of the original model but the results for countries are dependent on the distribution of failed and nonfailed firms. When the coefficients are estimated for all data, the effects on performance in a country depend on how the distributions in that country correspond to the distributions in all data. For all sets of additional variables, performance is generally improved but the improvement is not strong and the effects vary country by country. Thus, evidence gives weak support to effects of all additional variables. For the effects of bankruptcy year (H_3) and size (H_4), the effects are stronger but also the variations in the effects between countries are stronger. The effects of age (H_5), industry (H_6), and country (H_7) are marginal. When all additional variables are included in the same model, performance generally significantly increases, but at the same time variations between countries become stronger.

In summary, our evidence thus indicates that the original Z''-Score Model performs well in an international context. It is, however, possible to extract a more efficient country model for most European countries and also for non-European countries using the four original variables, accompanied with a set of additional background variables. Considering practical applications, it is obvious that while a general international model works reasonably well, for most countries the classification accuracy may be somewhat improved with country-specific estimation. In a country model, the information provided even by simple additional variables may help boost the classification accuracy to a much higher level.

In finance and accounting research, failure prediction models may be utilized as risk measures in many different contexts, as was revealed by the literature review section. Where failure prediction modeling is not the primary focus, it would be time-consuming, uneconomical, and superfluous to first estimate a failure prediction model (or models), and only then proceed to study the phenomenon of interest. In these kinds of instances, a well tested general model that works reliably and consistently across different countries is highly desirable. Based on our empirical tests in this study, the original Z''-Score

Model and its re-estimated version, containing the four Altman (1983) study variables with coefficients re-estimated using a large European data set, work consistently well internationally and are easy to implement and interpret.

An extension to our work could include a comprehensive and more complete analysis of the Altman and Rijken (2011) thesis that a country's risk of default on its sovereign debt can be assessed by an analysis of the default probabilities of its private sector. Our database permits extension of default estimates to non-publicly held firms, by far the dominant population in most countries of the world and, our analysis , can include small and medium size firms, as well as larger corporate entities.

References

Acosta-González E, Fernández-Rodríguez F (2014) Forecasting financial failure of firms via genetic algorithms. Computational Econ, 43:133–157

Agarwal V, Taffler R (2008) Comparing the performance of market-based and accounting-based bankruptcy prediction models. J. Bank. Financ. 32:1541–1551

Aktas N, de Bodt E, Lobez F, Statnik J-C (2012) The information content of trade credit. J. Bank. Financ. 36:1402–1413

Allayannis G, Brown GW, Klapper LF (2003) Capital structure and financial risk: Evidence from foreign debt use in East Asia. J. Financ 58:2667–2709

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Financ 23:589–609

Altman EI (1983) Corporate Financial Distress A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy. Wiley Interscience, John Wiley and Sons

Altman EI, Hotchkiss E (2006) Corporate Financial Distress & Bankruptcy , 3rd edition, John Wiley

Altman EI, Saunders A (1998) Credit risk measurement: Developments over the last 20 years. J. Bank. Financ. 21:1721–1742

Altman EI, Rijken, H (2011) Toward a Bottom-Up Approach for Assessing Sovereign Default Risk. J. of Appl. Corp. Finan. 23: 20-31

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

Bauer J, Agarwal V (2014) Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. J. Bank. Financ. 40:432–442

Beaver WH (1966) Financial ratios as predictors of failure. J. of Acc. Res. 4:71–111

Beaver WH (1968) Alternative accounting measures as predictors of failure. Acc. Rev. 43:113–122

Begley J, Ming J, Watts, S (1996) Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. Rev. Acc. Stud. 1:267–284

Bharath, ST Shumway, T (2008) Forecasting default with the Merton distance to default model. Rev. Finan. Stud. 21:1339-1369

Butler AW, Wan H (2010) Stock market liquidity and the long-run stock performance of debt issuers Rev. Finan. Stud. 23:3966–3995

Campbell J, Hilscher J, Szilagyi J (2008) In search of distress risk. J.Financ 63:2899–2939

Carling K, Jacobson T, Lindé J, Roszbach K (2007) Corporate credit risk modeling and the macroeconomy. J. Bank. Financ. 31:845–868

Chava S, Jarrow RA (2004) Bankruptcy prediction with industry effects. Rev. Finance 8:537–569

Chen H, Kacperczyk M, Ortiz-Molina H (2012) Do nonfinancial stakeholders affect the pricing of risky debt? Evidence from unionized workers. Rev. Finance 16:347–383

Clarke J, Ferris SP, Jayaraman N, Lee J (2006) Are analyst recommendations biased? Evidence from corporate bankruptcies. J. Finan. Quant. Anal. 41:169–196

Clayton MJ, Ravid SA (2002) The effect of leverage on bidding behavior: Theory and evidence from the FCC auctions. Rev. Finan. Stud. 15:723–750

Das S, Hanouna P, Sarin A (2009) Accounting-Based versus Market-Based cross-sectional models for CDS spreads. J. Bank. Financ. 33: 719-730

Dawkins MC, Bhattacharya N, Smith Bamber L (2007) Systematic share price fluctuations after bankruptcy filings and the investors who drive them. J. Finan. Quant. Anal. 42:399–419

DeLong, ER, DeLong, DM Clarke-Pearson, DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44: 837-845

Dichev I (1998) Is the Risk of Bankruptcy a Systematic Risk? J. Financ 53:1131-1147

Dimitras AI, Zanakis SH, Zopounidis C (1996) Theory and methodology A survey of business failures with an emphasis on prediction methods and industrial applications. Eur. J. Oper. Res. 90:487–513

Eisdorfer A (2008) Empirical evidence of risk shifting in financially distressed firms. J. Finance, 63:609–637

Fich EM, Slezak SL (2008) Can corporate governance save distressed firms from bankruptcy? An empirical analysis. Rev. Quant. Finance Acc. 30:225–251

Franzen LA, Rodgers KJ, Simin TT (2007) Measuring distress risk: The effect of RandD intensity. J. Financ 62:2931–2967

Grice JS, Dugan MT (2001) The limitations of bankruptcy prediction models: Some cautions for the researcher. Rev. Quant. Finance Acc. 17:151–166

Grice JS, Jr, Dugan MT (2003) Re-estimations of the Zmijewski and Ohlson bankruptcy prediction models. Adv. Acc. 20:77–93

Grice JS, Ingram RW (2001) Tests of the generalizability of Altman's bankruptcy prediction model. J.Bus. Res. 54:53–61

Griffin JM, Lemmon ML (2002) Book-to-market equity, distress risk, and stock returns. J. Finance 57:2317–2336

Hillegeist SA, Keating EK, Cram DP, Lundstedt KG (2004) Assessing the probability of bankruptcy Rev. Acc. Stud. vol 9:5–34

Hirsch, B T and Macpherson, D A (2003) Union membership and coverage database from the Current Population Survey: Note. Ind. Labor Rel. Rev. 56:349–354

Ho C-Y, McCarthy P, Yang Y, Ye X (2013) Bankruptcy in the pulp and paper industry: market's reaction and prediction. Empirical Economics 45:1205–1232

Holder-Webb LM, Wilkins MS (2000) The incremental information content of SAS No 59 Going-concern opinions. J. Acc. Res. 38:209–219

Hosmer D, Lemeshow S (1989) Applied Logistic Regression. Wiley Interscience, Prentice Hall International

Jackson RHG, Wood A (2013) The performance of insolvency prediction and credit risk models in the UK: A comparative study. British Acc. Rev. 45:183–202

Karels, G, Prakash, AJ (1987) Multivariate normality and forecasting of business bankruptcy. J. Bus. Finan. Account. 14:573-593

Kieschnick R, La Plante M, Moussawi R (2013) Working capital management and shareholders' wealth. Rev. Finance 17:1827–1852

Kumar PR, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. Eur. J. Oper. Res. 180:1–28

Kwak W, Shi Y, Cheh JJ, Lee H (2005) Multiple criteria linear programming data mining approach: An application for bankruptcy prediction. Lecture Notes Comput. Sci. 3327:164–173

Li M-YL, Miu P (2010) A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. J. Empir.Financ. 17:818–833

Lyandres E, Zhdanov A (2013) Investment opportunities and bankruptcy prediction. J. Financ.Mark. 16:439–476

Merkevicius E, Garšva G, Girdzijauskas S (2006) A hybrid SOM-Altman model for bankruptcy prediction. Lecture Notes Comput. Sci. 3994:364–371

Ohlson J (1980) Financial ratios and probabilistic prediction of bankruptcy. J. Acc. Res. 18:109-131

Ooghe H, Balcaen S (2007) Are failure prediction models widely usable? An empirical study using a Belgian dataset. Multinatl. Financ. J. 11: 33-76

Philippe and Partners and Deloitte and Touche, Corporate Finance (2002) Bankruptcy and a Fresh Start: Stigma on Failure and Legal Consequences of Bankruptcy. International Insolvency Institute, Brussels, Belgium

Philosophov LV, Batten JA, Philosophov VL (2008) Predicting the event and time horizon of bankruptcy using financial ratios and the maturity schedule of long-term debt. Math. Finan. Econ. 1:181–212

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

Piotroski JD (2000) Value investing: The use of historical financial statement information to separate winners from losers. Supplement: Studies on accounting information and the economics of the firm. J. Acc. Res. 38:1–41

Reisz AS, Perlich C (2007) A market-based framework for bankruptcy prediction. J.Finan. Stab. 3:85–131

Shen C-H, Chen Y-K, Huang B-Y (2010) The prediction of default with outliers: Robust logistic regression. Handbook of Quantitative Finance and Risk Management

Shumway, T (2001) Forecasting bankruptcy more accurately: A simple hazard model. J. Bus. 74:101-124

Singhal R, Zhu Y (2013) Bankruptcy risk, costs and corporate diversification. J. Bank. Financ. 37:1475–1489

Smith, M, Liou, D-K (2007) Industrial sector and financial distress. Managerial Auditing Journal 22:376 - 391

Sun L (2007) A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. Rev. Quant. Finance Acc. 28:55–78

Taffler RJ (1984) Empirical models for the monitoring of UK corporations J. Bank. Financ. 8:199-227

Uhde A, Heimeshoff U (2009) Consolidation in banking and financial stability in Europe: Empirical evidence. J. Bank. Financ. 33:1299–1311

Vassalou M, Xing Y (2004) Default Risk in Equity Returns. J. Finance 59:831-868

Wang Y, Campbell M (2010) Business failure prediction for publicly listed companies in China. J. Bus. Manag. 16:75-88

Wu Y, Gaunt C, Gray S (2010) A comparison of alternative bankruptcy prediction models. J. Contemp. Acc. Econ. 6:34–45

Xu M, Zhang C (2009) Bankruptcy prediction: the case of Japanese listed companies. Rev. Acc.Stud. 14:534–558

Zhang L, Altman EI, Yen J (2010) Corporate financial distress diagnosis model and application in credit rating for listing firms in China. Frontiers Comput. Sci. China 4:220-236

Zmijewski ME (1984) Methodological Issues Related to the Estimation of Financial Distress Prediction Models. J. Acc. Res. 22:59-82

Tables

Table 1. Summary of the literature review of Z-Score Models from year 2000.

Authors and year of publication	Period and population or sample	Goal, methodology and Z-Score application	Main findings
Holder-Webb L.M., Wilkins M.S. (2000)	217 firms filing for bankruptcy identified by NAARS or Wall Street Journal between 1975-1996	(M) Evaluation of the impact of implementation of auditing standard on going concern on information value for investors (bankruptcy surprise and audit opinion were assessed by investors' reactions to bankruptcy news in the event window). Authors implemented BKPRO variable based on z-score in the model as a ratio-based measure of financial distress.	Bankruptcy surprise was greater in case of a clean going concern opinion. The results hold after controlling for e.g. firm's level of financial distress.
Piotroski J.D. (2000)	14,043 firms with high book-to-market (BM) values across 21 years (1976 -1996) in Compustat database	(M) The application of a simple accounting-based fundamental analysis strategy as a tool for broad portfolio management of high BM firms. The author used z-score as financial distress measure in order to justify his own model.	The author concluded that firms with lower level of financial distress earned significantly stronger future returns than the highly distressed firms. Mean return for the investor involved in the high BM assets could be increased by selection of firms that are financially strong.
Grice J.S., Ingram R.W. (2001)	972 companies for 1985–1987 in the estimation sample (148 distressed and 824 non-distressed) and 1002 companies for 1988–1991 in the prediction sample (148 distressed and 854 non-distressed) in Compustat database.	(V) Verification of z-score model with 3 questions: Is the model as useful as in the previous periods? Is the model useful for predicting bankruptcy of non-manufacturing firms as it was for manufacturing firms? Is the model useful for predicting financial stress other than bankruptcy?	Prediction accuracy of the model declined over time and coefficients of the model significantly changed, which means that the relation between financial ratios and financial distress has changed over time. Model was sensitive to industry classification (better for manufacturing firms) and not sensitive to the type of financial distress. Similar conclusions were drawn in case of Ohlson's (1980) and

			Zmijewski's (1984) model by Grice J.S., Dugan M.T. (2001).
Clayton M.J., Ravid S.A. (2002)	Federal Communication Commission (FCC) spectrum auctions for 1994-1995 together with firm-specific and region-specific variables	(M) The authors examined the link between capital structure of firms and their bidding behaviour. The authors used z-score as a proxy for financial strength and alternative measure of default risk.	Empirical analysis showed that as the firm's and the competitor's debt level increase, firms tend to submit lower bids, which supported authors' theoretical model.
Griffin J.M., Lemmon M.L. (2002)	Sample similar to Fama and French (1992). Non-financial NYSE, Nasdaq, Amex stocks monthly returns for stocks with nonnegative book value of equity from July 1965 to June 1996	(R) The authors investigated relationship between book-to-market equity, distress risk and stock returns. They used Ohlson's o-score, but additionally run robustness check with the use of z-score as measure of distress risk.	In case of firms with the highest risk of distress the difference in return between high and low book-to-market equity securities was more than twice higher than in case of other groups.
Allayannis G. et al.(2003)	SBC Warburg Dillon Read database with local, foreign and synthetic local (hedged foreign) currency debt for 327 largest East Asian non-financial companies from 1996-1998.	(M) The first research issue was to check determinants of firm's choice of debt currency: local, foreign or synthetic local. The second research issue was related to the relation of the type of debt and financial performance during the Asian crisis. Asian financial crisis was treated as natural experiment. They used modified z-score (2000) as one of the financial performance measures.	The authors found no evidence that unhedged foreign debt was the cause of poor performance. However, the use of synthetic currency caused poor stock market performance because of the derivatives market illiquidity during the crisis.
Hillegeist S.A. et al. (2004)	Sample covered 78,100 firm-year observations (representing 14,303 individual industrial firms) for 1980-2000 period with 756 initial bankruptcies.	(V) The authors confronted Altman's z-score and Ohlson's o-score (with original and up-dated coefficients) with the model based on BSM option pricing model (BSM-Prob). They used relative information content tests to compare the out-of-sample performance.	BSM-Prob outperformed accounting based models. Conclusions were robust to various modifications of both accounting based models. BSM-Prob was not dependent on time horizon and difference in accounting standards in various countries. Thus it allowed for international comparisons.

Chava S., Jarrow R.A. (2004)	US publicly listed companies' monthly and yearly data for 1962-1999 period with 1461 bankruptcies.	(V) The authors employed extended bankruptcy database to prove the superiority of Shumway's model (2001) over Altman's (1968) and Zmijewski's (1984) models. Three other goals were included into their research. The authors re-estimated models over the 1962–1990 period and forecasted bankruptcies over 1991–1999.	In case of Shumway's model 74.4 % (in the first decile) of the bankruptcies were correctly identified, for Altman's model – it was 63.2% and for Zmijewski's model - 43.2 %. Market-based model outperformed also in case of ROC curve. Market variables were better in predicting bankruptcy than accounting variables and the use of monthly data improved prediction.
Kwak et.al (2005)	Data from 1992 to 1998 for bankrupt firms and 5.84 times more of control firms for the same period. Standard and Poor's database was used for US companies from major 3 stock exchanges.	(V) Multiple Criteria Linear Programming (MCLP) was used to model 5 Altman's and 9 Ohlson's variables in order to apply it for bankruptcy prediction.	In case of both models overall predication rate of Ohlson's model is similar to Altman's model (86.76% vs. 88.56%) with more control firms. The MCLP approach performed better than Altman's model and give similar or better results than Ohlson's model.
Merkevicius E. et al. (2006)	Data from two sources: EDGAR PRO Online Database ("traindata" with 1108 records) and database form Lithuanian bank ("testdata" with 742 records) for year 2004.	(V) The hybrid model was developed with the use of Altman's z-score model with the changed weights for variables and self-organizing map, which was an unsupervised learning artificial neural network (called SOM-Altman model).	Hybrid SOM-Altman's model reached the prediction rate of 92.35%. It is easy to adopt for different data sets, also from developed and developing countries.
Clarke J. et al. (2006)	The initial sample of 995 firms that filed for bankruptcy from 1995 to 2011. Analysts recommendations were taken from IBES database. In the final sample the authors used 289 bankrupt firms and their non-bankrupt matches.	(M) The research focus was to test if there is a bias in analysts' recommendations for firms that filed for bankruptcy (chapter 11). Altman's z-score was calculated two years preceding the bankruptcy year.	Analysts actively revised their recommendations downward in case of firms that faced bankruptcy. The recommendation declined from "buy" about 2 years prior to bankruptcy. The authors found no evidence of analysts' bias and conflict of interest.
Reisz A.S., Perlich C. (2007)	The authors used 5784 industrial firms for 1988-2002, which gave	(V) The authors used barrier option model for bankruptcy prediction (DOC) and compared its discriminatory power	The authors proved domination of Altman's z-score and z"-score for short term bankruptcy prediction. For medium- and

	33,238 non-bankrupt and 799 bankrupt firm-years	with other market-based (BSM and KMV) and accounting-based models (Altman's z-score and z"-score, developed in 1993 for private firms).	long-term bankruptcy prediction DOC outperformed other models.
Sun L. (2007)	US public traded firms on the major stock exchanges from 1991 to 2002. The final bankruptcy training sample consists in 344 firms (1,657 firm-years) and the final non-bankruptcy training sample – 3,183 (20,918 firm-years). The test samples used respectively 243 and 1,165 firms.	(M) The main purpose of the research was to compare the performance of auditors' going concern opinions with the results of bankruptcy prediction models. The author improved previous research by using a composite score of financial distress, stock-market variables, industry failure rate and applying hazard model. Altman's z-score was one of the 4 criteria (besides auditor's opinion, Zmijewski's probability and stock return) used as a "stress" variable. In the final model the author used Zmijewski's probability as the best.	The hazard model outperformed auditor's going concern opinions, so statistical models could be of any help for auditors' opinions. Market variables had significant explanatory power for auditor's opinion, while industry failure rate did not have it. Auditors' going concern opinions had incremental prediction ability beyond traditional financial ratios, a composite measure of financial distress, market variables and industry failure rate.
Franzen L.A. et. al (2007)	US firms listed on the major stock exchanges for the period 1981-2004 with sufficient data to compute accounting-based measures of financial distress, R&D intensity and portfolio returns. The number of firms filing for bankruptcy was 929.	(M) The authors focused on the impact of R&D spending on financial distress prediction. They used accounting based models with the special focus on Ohlson's o-score since it included only accounting based measures. Altman's z-score was used for supporting purposes, since it included also market data.	Accounting-based measures less accurately classify firms with higher R&D spending. In case of adjustment for conservative accounting treatment of R&D and tax effects the classification was better. Z-score was used as a benchmark. Z-score was more robust to the changes that distort the o-score due to the use of market value. Adjustment of z-score for the conservative treatment of R&D improved the classification.
Dawkins M.C. et.al (2007)	272 US firms that filed for bankruptcy between 1993 and 2003, traded on the major stock exchanges. Investors'	(M) The authors investigated pattern of returns and investors trades around and shortly after Chapter 11 bankruptcy announcement. They divided the	Large traders were dominating transactions around and after bankruptcy filing. The authors found a modest negative correlation between the price plunge and

	intraday transactions were taken from Trade and Quote database (TAQ).	investigated period into bull and bear markets. The authors used five event windows. Altman's z-score was used as a measure of firm's financial condition.	post-filing period returns. Investors' reactions were anomalously optimistic in the period of a bull market.
Agarwal V. and Taffler R. (2008)	Non-finance UK firms listed on the LSE at any time during the period 1985-2001. The final sample covered 2006 firms (15,384 firm-years) and 103 failures.	(V) In the paper the authors compared accounting-based models and market based models in terms of predictive ability, information content and bank credit portfolio profitability. Accounting-based models were represented by Taffler's (1984) UK z-score, based on Altman's (1968).	In the case of the predictive ability there was a little difference between those two types of models, however, both carried unique information about firm failure. The use of the z-score model generated much higher risk-adjusted return, profit and return on risk-weighted assets.
Philosophov L.V. et al. (2008)	No sample of non-bankrupt firms. 100 US non-financial firms that filed for bankruptcy between 1997 and 2002.	(V) The authors applied multi-alternative decision rules for bankruptcy prediction (BMP). Bayesian-type forecasting rules used debt repayment schedule and traditional financial ratios. Results of the new approach were compared with Altman's z-score.	The authors have identified 4 factors important for multiperiod approach. Two of them reflect quantity and quality of debt and the two others the ability to pay the debt back. Those factors are similar to as those proposed in the z-score. The set of factors did not change, however the predictive ability decreased over time. BPM outperformed z-score in total probability of correct predictions, while "false alarm" was comparable.
Fich E.M., Slezak S.L. (2008)	The sample includes 781 companies with 34 that filed for bankruptcy. The period covers 1992-2000. Except the availability of financial and market data, the availability of corporate governance data from Edgar data retrieval system was a prerequisite.	(M) The authors explored how governance characteristics affect firm's ability to avoid bankruptcy and power of accounting/financial information to predict bankruptcy with the use of hazard models. Altman's z-score was applied as a measure of firm's financial distress together with interest coverage ratio (ICR).	Corporate governance characteristics are important for bankruptcy prediction. About 25-30% of the variation in the occurrence of bankruptcy was explained by governance characteristics. Smaller and independent boards with a higher ratio of non-inside directors and with larger ownership stakes of inside directors are more effective at avoiding bankruptcy if the firm became distressed.

Eisdorfer A. (2008)	US companies traded on the major stock exchanges for period from 1963 to 2002, representing 7,114 firms (52,112 firm-years).	(M) The author tested two hypotheses: volatility has a positive effect on distressed firms' investment and investments of these firms generate less value during high uncertainty periods with the use of real option approach. Altman's z-score and KMV model were used as a measure of firm's financial distress.	In case of first hypothesis z-score results were stronger than those based on KMV, but in case of the second hypothesis KMV results were stronger. Empirical tests indicated that risk-shifting considerations are taken into account by distressed firms.
Pindado J. et al. (2008)	Sample covered 1583 US companies (15,702 observations) and 2250 companies (18,160 observations) for the other G-7 countries for the period from 1990 to 2002 from Compustat. The data panel was unbalanced.	(R)The authors' developed ex-ante model for estimation of financial distress likehood (FDL) and presented financially (not legally) based definition of distress. Re-estimated Altman's z-score was used for robustness check.	The model is stable in terms of magnitude, sign and significance of the coefficients and it provides a measure of FDL more robust to time and countries than other models. In case of the re-estimated z-score model only profitability and retained earnings maintained their significance for different years and countries.
Xu M., Zhang C. (2009)	Non-financial Japanese listed companies from 1992 to 2005. The number of firms was 3,510 with 76 bankruptcies.	(V) The authors applied Altman's z-score, Ohlson's o-score and Merton's distance-to-default (d-score) to Japanese firms in order to check if the models are useful for bankruptcy prediction in Japan. They also "merged" the models into c-score. They introduced Japan-unique variables in order to check if corporate structure variables have impact on the probability of bankruptcy (x-score).	The models were also useful for Japanese firms in bankruptcy prediction, but the market-based model was more successful. The developed c-score and x-score models improved bankruptcy prediction.
Li M.-Y.L., Miu P. (2010)	Non-financial firms from Compustat database from 2nd Quarter 1996 to 4th Quarter 2006. The sample covered 73 failed firms and 138 non-failed firms with at least rating B during the whole period.	(M) Altman's z-score and Merton's distance-to-default (DD) were used as variables in the model. The authors developed hybrid bankruptcy prediction model with constant and dynamic	The results of the research pointed out the superiority of the model with dynamic loadings. In the case of firms with relatively poor credit quality the prediction accuracy might be improved by putting more attention to DD and less to the z-score. In

		loadings with the use of binary quintile regression (BQR).	case of firms with relatively good credit quality the recommendations are the opposite, e.g. more attention should be put to the z-score.
Wu Y. et al. (2010)	NYSE- and AMEX-listed Compustat non-financial firms for the period from 1980 to 2006. The final sample covered 50,611 firm-years, with 887 bankruptcies and 49,724 non-bankrupt firm-years.	(V) Five models (Altman 1968, Ohlson 1980, Zmijewski 1984, Shumway 2001, Hillegeist et al. 2004) from the literature are used by the authors in order to evaluate their performance with an up-to-date data set. Based on them, the authors build their own integrated model (multi-period logit model with expanded set of variables).	The integrated model using accounting and market data as well as firms' characteristics outperformed the other models. Altman's z-score performed poorly in comparison to the other 4 models. Shumway's model outperformed, model of Hillegeist et al. performed adequately and Ohlson's and Zmijewski's models performed adequately, but their performance deteriorated over time.
Shen C.-H. et al. (2010)	Non-financial companies listed on the Taiwan Stock Exchange, covering the period from 1999 to 2004. There were 52 default companies and 156 non-default companies in the sample.	(V) The authors applied a robust logit method to build up an empirical model based on Altman's z-score (one variable was omitted: retained earnings to total assets). They took into account outliers which so far have been omitted in the models. The results of robust logit method were compared with logit method.	In the case of in-the-sample forecasts, robust logit models preformed better than the logit model. However in the case of out-the-sample forecast superiority of robust logit model disappeared. Robust logit model was more aggressive in assigning firms as default.
Butler A.W., Wan H. (2010)	New debt offerings from 1975-1999 (4,293 offerings, within 632 convertible debt issue). Sample covered non-regulated utility and non-financial companies listed on major US stock exchanges. The first offering in the sample period for a particular firm was used and the other offerings from the same firms	(M) The aim of the paper was to assess long-term stock performance of debt and convertible debt issuers. The authors revised the results of the previous research on the stock returns of bond issuing firms. According to the authors, previous models omitted liquidity factor. Altman's z-score was used as a proxy for financial distress, besides momentum as another matching criterion.	Introduction of the liquidity factor into the model erased long-run underperformance of debt and convertible debt issuers. Momentum and Altman's z-score did not erase this underperformance.

	within the following five years were excluded from the sample.		
Aktas N. et al. (2012)	US listed non-financial firms with z-score higher than the median of the population from 1992 to 2007. The sample consisted in 15,928 firm-years from 2096 firms.	(M) The article investigated the relationship between the use of trade credit and the quality of firms' investments. Altman's z-score was used as one of three proxies for measuring the quality of firm's investment projects (additionally return on assets and long-term abnormal returns). The authors developed a theoretical model for trade credit decision with the use of strategic game with imperfect information, as well as empirical study.	According to the empirical study, the change in the trade credit signaled the future positive change in firm's performance. Empirical results supported conclusions from the theoretical model. In the case of the use of next period change of the z-score, the authors confirmed both CEO incentive and entrenchment, however, in the case of two other proxies, only CEO incentive.
Chen H. et al. (2012)	US non-financial companies issuing debt securities from Lehman Brothers Bond Database, data on unionization described in Hirsch and Macpherson (2003) supported by financial and demographic data. The period covered years from 1984 to 1998. The sample consisted in 996 firms and 5,557 firm-years observations.	(M) The authors focused on the impact of "unionized workers" on the pricing of corporate debt. Altman's z-score was used as one of 5 measures of distress (accompanied by interest coverage, book equity, Ohlson's o-score and default measure developed by Bharath and Shumway in 2008).	The authors drew 4 conclusions: In the case of more "unionized" industries, the yield on bonds is lower; the impact is stronger in firms with weak financial standing; highly "unionized" firms invested less in R&D than in physical assets; they were less likely to be the acquisition target. According to the authors "unionized workers" protected bondholders' and other fixed claimants wealth.
Jackson R.H.G., Wood A. (2013)	LSE-listed non-financial UK firms. The sample consisted in 101 failed firms, of which failure was registered from 30 Sept. 2000 to 31 Dec. 2009 and 6494 non-failed firm-years from 1 Jan. 2000 to 31 Dec. 2009	(V) The authors tested 13 different models of bankruptcy prediction and assessed their efficacy with the use of ROC curve. They selected 3 single variable models; 3 accounting-based models (with Altman's z-score) in two versions: with up-dated coefficients and neural network approach; 4 contingent claims models.	Contingent claims models outperformed the other models. The best four models are contingent claims models based on European call and barrier options. In the case of application of the neural network to accounting based models the predictive ability improved, however was still lower than in case of market-based models.

Singhal R., Zhu Y. (2013)	US non-financial firms. Final sample covered 769 bankruptcy filings between 1 Jan. 1991 to 31 Dec. 2007. Failed firms were divided to two groups: focused (622) and diversified (147).	(M) The focus of the study was to check bankruptcy risk and bankruptcy costs in the case of focused vs. diversified firms. Altman's z-score was used as a measure of financial distress.	The mean and median z-score for diversified groups were higher than for focused groups. The Z-score was highly correlated with the leverage in the authors' model. Diversified firms represented a lower risk of bankruptcy, however they underperformed focused firms in other areas (like inefficient segment investment and longer time in the bankruptcy procedure, what incurred higher bankruptcy costs).
Lyandres E., Zhdanov A. (2013)	US companies (financial and non-financial) listed on major stock-exchanges. The sample covered period from 1985 to 2005. The number of bankruptcies was 948 (compiled from different sources, like Thomson Financial and T. Shumway's database) and the number of firm-years was 146,836.	(V) The authors posed the question if the inclusion of variables related to investment opportunities improved the predictive power of three models (Altman's z-score, Zmijewski's and Shumway's models). They used 3 proxies for investment opportunities (market-to-book, value-to-book, R&D-to-assets).	The measures of investment opportunities were related to the likehood of default. Inclusion of either of three investment opportunities measures improved models' out-of-sample forecasting ability.
Ho C.-Y. et al. (2013)	North America publicly listed pulp and papers firms. Sample includes 122 firms, within 12 that filed for bankruptcy. The period from 1990 to 2009. Data were compiled from different sources (like Compustat, CRSP, WRDS).	(R) The paper investigated bankruptcy prediction with the use of financial ratios and investors' reactions to bankruptcy filing. The authors used Ohlson's model (original and re-estimated) to predict bankruptcy in the pulp and paper industry. Altman's z-score (original and re-estimated) was used for robustness check.	Investors' suffered significant losses during the month a bankruptcy occurred. O-scores of failed firms were higher than in case of non-failed firms 1 to 2 year before the bankruptcy. Re-estimated Ohlson's model performed better than the original. In the case of Altman's z-score, the re-estimated model performed better due to the reduction of Type II errors that occurred in the original one. In both cases models were useful for bankruptcy prediction.
Kieschnick R. et al. (2013)	Non-financial US companies from 1990 to 2006. There were 3,786	(M) The paper focused on the relationship between working capital management and shareholders' wealth	The authors concluded that: "incremental dollar invested in net operating capital is worth less than an incremental dollar held

	companies in the sample (yearly average).	as well as their determinants. Altman's original (1968) and revised (2000) z-score were used as a measure of the financial distress.	in cash for the average firm". The impact on the shareholder value was determined by sales expectations, access to external capital, bankruptcy risk and debt load. The higher the z-score was, the more valuable its investment in working capital.
Acosta-González E., Fernández-Rodríguez F. (2014)	Spanish building industry firms. For model estimation the sample covered 93 firms that failed in 2004 and 257 randomly selected non-failed firms for the period from 2000 to 2004. For ex-ante forecast: accounting information from 2005 for randomly selected 400, out of which 80 failed in 2007.	(V) The authors used genetic algorithms with Schwarz information criterion (GASIC) for variables selection (out of 32 preselected ratios) combined with the logit model for bankruptcy prediction. Altman's z-score model was used as one of two benchmarks for authors' model evaluation.	One step ahead forecast Altman's model was better in predicting failed firms but the type II error was high; two and three steps ahead forecasts were similar in terms of failed firms, but in the case of non-failed Altman's model prediction was worse; four steps ahead forecasts GASIC model outperformed in the case of failed firms and was comparable in the case of non-failed firms.

Notes: M - Z-Score use as a measure of distress or financial strength; V – Z-Score verification or modification; R – Z-Score use used for robustness check

Table 2. Number of observations by country.

	Estimation data:		Test data:	
Country	Non-failed	Failed	Non-failed	Failed
Austria (AT)	7430	55	7526	44
Belgium (BE)	179979	2994	179818	2944
Bosnia (BA)	29391	35	29139	32
Bulgaria (BG)	50041	48	42351	44
China (CN)			39315	198
China, delisted dataset			16291	29
China, ST data	846	16	1020	16
Colombia (CO)	8366	139	6982	125
Greece (GR)			51763	28
Croatia (HR)	59541	249	58478	275
Czech Republic (CZ)	92835	556	92562	564
Denmark (DK)	167934	1334	168538	1398
Estonia (EE)	34313	234	34196	242
Finland (FI)	90878	481	91227	459
France (FR)	160749	6124	161653	6318
Germany (DE)	98814	910	99496	921
Hungary (HU)	19421	303	20155	313
Iceland (IS)	17399	248	17624	243
Ireland (IE)	6665	121	6406	139
Italy (IT)	167113	8101	166258	8124
Latvia (LV)	8064	433	8241	477
Lithuania (LT)			10000	56
Netherlands (NL)	20885	154	15854	147
Norway (NO)	172467	1294	170985	1206
Poland (PL)	87200	291	86233	264
Portugal (PT)	180114	3390	178646	3422
Romania (RO)	161992	97	164259	93
Russian Federation (RU)	116903	2534	115711	2481
Serbia (RS)			100100	68
Slovakia (SK)	7856	120	7788	124
Slovenia (SI)	14419	59	14081	41
Spain (ES)	156746	3036	158122	2991
Sweden (SE)	169810	2256	169999	2314
Ukraine (UA)	133342	1787	133980	1765
United Kingdom (GB)	171493	760	170930	716
UK (GB), liquidation dataset			342423	4990
United States (US)	9557	56	9929	53
In total	2602563	38215	3148079	43664

Table 3. Descriptive statistics (all data).

	WCTA		RETA		EBITTA		BVETD	
Statistic	Non-failed	Failed	Non-failed	Failed	Non-failed	Failed	Non-failed	Failed
Median	0,152	-0,059	0,189	-0,024	0,041	-0,020	0,451	0,025
Mean	0,147	-0,213	0,188	-0,317	0,055	-0,108	3,594	0,703
Standard deviation	0,442	0,604	0,509	0,767	0,227	0,296	11,499	5,712
Upper quartile	0,420	0,142	0,476	0,087	0,131	0,042	1,548	0,215
Lower quartile	-0,040	-0,440	0,011	-0,450	-0,008	-0,208	0,100	-0,240
Maximum	0,956	0,956	0,958	0,958	0,785	0,785	68,606	68,606
Minimum	-1,637	-1,637	-2,453	-2,453	-0,828	-0,828	-0,649	-0,649

Legend:
WCTA = Working Capital/Total Assets
RETA = Retained Earnings/Total Assets
EBITTA = EBIT/Total Assets
BVETD = Book Value of Equity/Total Liabilities
SALTA = Sales/Total Assets

Table 4. The coefficients of the different models estimated for all data.

Variable	Z''-Score	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
		Coefficients for different statistical models:							
Constant	3,25	-0,042	0,035	0,207	-13.466	0,007	0,048	0,049	-13,302
WCTA	6,56	-0,561	-0,495	-0,483	-0,441	-0,487	-0,540	-0,496	-0,459
RETA	3.26	-0,724	-0,862	-0,891	-1,146	-0,846	-0,859	-0,863	-1,160
EBITTA	6.72	-1,791	-1,721	-1,790	-1,619	-1,757	-1,695	-1,717	-1,682
BVETD	1.05	-0,021	-0,017	-0,016	-0,012	-0,017	-0,016	-0,017	-0,013
Year dummies:									
Year 2008				-0,055					-0,034
Year 2009				-0,179					-0,150
Year 2010				-0,666					-0,631
Size variables:									
Total assets (log)					1,830				1,837
Total assets squared (log)					-0,061				-0,061
Age dummies:									
Less than 6 years						0,135			0,186
Over 12 years						-0,058			-0,099
Country risk:									
SP Country rating rank								-0,003	-0,014
Industry dummies:									
Restaurants and hotels							-0,653		-0,628
Construction							0,445		0,365
Whole sale and retailing							-0,112		-0,157
Agriculture							-0,180		-0,176
Manufacturing							0,139		0,095
Energy and water production							-0,454		-0,472
Information technology							-0,913		-0,915

Significance:

Coefficients are statistically significant at 0.0001.

Models:

Z''-Score = Original Altman (1983) Z''-Score Model coefficients

Model 1 = The MDA model

Model 2 = The LR model

Model 3 = The LR model estimated for all data with year dummies

Model 4 = The LR model estimated for all data with size variables

Model 5 = The LR model estimated for all data with age category dummies

Model 6 = The LR model estimated for all data with industry dummies

Model 7 = The LR model estimated for all data with country risk rank

Model 8 = The LR model estimated for all data with all variables

Table 5. Test data AUCs for different countries, based on all data model versions. Comparisons are with AUC of the LR model estimated for all data (benchmark).

	Test data AUC for different models:								
Country	Benchmark	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
All Data	0,748	0,743----	0,745----	0,752++++	0,760++++	0,748+++	0,751++++	0,749++++	0,771++++
Countries with estimation data:									
Austria (AT)	0,800	0,788	0,797	0,805	0,818+	0,814+	0,782	0,800	0,819
Belgium (BE)	0,772	0,760	0,770	0,777	0,747	0,777	0,765	0,772	0,758
Bosnia (BA)	0,862	0,805	0,857	0,776	0,863	0,847	0,855	0,862	0,784
Bulgaria (BG)	0,684	0,630	0,680	0,654	0,691	0,680	0,659	0,684	0,632
China, ST companies	0,985	0,911----	0,983	0,958-	0,977	0,978	0,987	0,985	0,968
Colombia (CO)	0,726	0,724	0,727	0,715	0,758	0,728	0,726	0,726	0,757
Croatia (HR)	0,844	0,812	0,839	0,803	0,837	0,835	0,832	0,844	0,801
Czech Republic (CZ)	0,811	0,813	0,811	0,819	0,828	0,807	0,820	0,811	0,838
Denmark (DK)	0,803	0,798	0,800	0,781	0,813	0,802	0,801	0,803	0,796
Estonia (EE)	0,823	0,827	0,823	0,847	0,866	0,826	0,833	0,823	0,890
Finland (FI)	0,867	0,864	0,866	0,835	0,862	0,870	0,878	0,867	0,853
France (FR)	0,739	0,723	0,735	0,749	0,771	0,741	0,762	0,739	0,799
Germany (DE)	0,673	0,658	0,666	0,695	0,656	0,684	0,677	0,673	0,688
Hungary (HU)	0,742	0,746	0,740	0,660	0,738	0,755	0,735	0,742	0,696
Iceland (IS)	0,664	0,674	0,666	0,694	0,678	0,673	0,672	0,664	0,716
Ireland (IE)	0,679	0,672	0,676	0,708	0,677	0,681	0,688	0,679	0,712
Italy (IT)	0,806	0,799	0,804	0,833	0,835	0,806	0,799	0,806	0,849
Latvia (LV)	0,678	0,691	0,678	0,704	0,676	0,686	0,698	0,678	0,724
Netherlands (NL)	0,752	0,754	0,750	0,775	0,769	0,754	0,746	0,752	0,787
Norway (NO)	0,716	0,694	0,713	0,658	0,682	0,720	0,715	0,716	0,645

Poland (PL)	0,903	0,904	0,904	0,908	0,902	0,903	0,899	0,903	0,904
Portugal (PT)	0,741	0,724	0,736	0,749	0,773	0,738	0,755	0,741	0,785
Romania (RO)	0,758	0,740	0,754	0,709	0,749	0,755	0,748	0,758	0,703
Russian Federation (RU)	0,811	0,802	0,812	0,843	0,799	0,807	0,799	0,811	0,814
Slovakia (SK)	0,777	0,774	0,776	0,780	0,811	0,769	0,786	0,777	0,808
Slovenia (SI)	0,737	0,725	0,733	0,674	0,740	0,747	0,718	0,737	0,721
Spain (ES)	0,734	0,713	0,732	0,707	0,793	0,728	0,753	0,734	0,774
Sweden (SE)	0,813	0,801	0,809	0,799	0,784	0,817	0,823	0,814	0,800
Ukraine (UA)	0,708	0,714	0,710	0,715	0,721	0,708	0,702	0,708	0,722
United Kingdom (GB)	0,699	0,719	0,699	0,686	0,736	0,695	0,706	0,699	0,729
United States (US)	0,710	0,701	0,711	0,709	0,722	0,705	0,716	0,710	0,723
Countries only in test data:									
European:									
Greece (GR)	0,715	0,670	0,702	0,725	0,711	0,713	0,717	0,715	0,718
Lithuania (LT)	0,767	0,782	0,767	0,764	0,768	0,769	0,775	0,767	0,778
Serbia (RS)	0,736	0,713	0,730	0,603-	0,826	0,720	0,753	0,736	0,738
United Kingdom, Liquidation set	0,606	0,621	0,603	0,620	0,618	0,610	0,607	0,607	0,635
Non-European:									
China (CN)	0,558	0,570	0,557	0,572	0,543	0,554	0,567	0,558	0,556
China, Delisted companies	0,529	0,546	0,519	0,563	0,740++++	0,542	0,520	0,529	0,707+++

Significance:

AUC better than benchmark: 0.0001 = ++++, 0.001 = +++, 0.01 = ++, 0.1 = +

AUC worse than benchmark: 0.0001 = ----, 0.001 = ---, 0.01 = --, 0.1 = -

Models:

Benchmark = The LR model estimated for all data with Z’’-model (1983) variables

Model 1 = The original Altman (1983) Z’’-Score Model

Model 2 = The MDA model estimated for all data

Model 3 = The LR model estimated for all data with year dummies

Model 4 = The LR model estimated for all data with size variables
Model 5 = The LR model estimated for all data with age category dummies
Model 6 = The LR model estimated for all data with industry dummies
Model 7 = The LR model estimated for all data with country risk rank
Model 8 = The LR model estimated for all data with all variables

Table 6. Differences of medians between non-failed and failed groups.

Country	WCTA	RETA	EBITTA	BVETD	AUC of Z''-Score
Austria (AT)	0,448	0,405	0,126	0,487	0,788
Belgium (BE)	0,223	0,264	0,077	0,431	0,760
Bosnia (BA)	0,106	0,249	0,088	0,580	0,805
Bulgaria (BG)	0,114	0,291	0,094	0,436	0,630
China (CN)	0,068	0,032	0,016	0,280	0,570
China, delisted data	0,089	-0,061	-0,032	0,083	0,546
China, ST data	0,298	0,293	0,139	0,468	0,911
Colombia (CO)	0,226	0,235	0,099	0,705	0,724
Croatia (HR)	0,275	0,335	0,044	0,274	0,812
Czech Republic (CZ)	0,404	0,397	0,069	0,667	0,813
Denmark (DK)	0,267	0,356	0,046	0,733	0,798
Estonia (EE)	0,331	0,388	0,113	1,033	0,827
Finland (FI)	0,402	0,606	0,207	0,907	0,864
France (FR)	0,140	0,213	0,065	0,424	0,723
Germany (DE)	0,131	0,136	0,032	0,262	0,658
Greece (GR)	0,171	0,262	0,049	0,349	0,670
Hungary (HU)	0,175	0,224	0,052	0,587	0,746
Iceland (IS)	0,246	0,261	0,051	0,337	0,674
Ireland (IE)	0,181	0,266	0,039	0,560	0,672
Italy (IT)	0,277	0,164	0,073	0,207	0,799
Latvia (LV)	0,117	0,120	0,042	0,254	0,691
Lithuania (LT)	0,246	0,218	0,051	0,569	0,782
Netherlands (NL)	0,204	0,253	0,077	0,432	0,754
Norway (NO)	0,157	0,219	0,115	0,329	0,694
Poland (PL)	1,340	0,920	0,124	1,351	0,904
Portugal (PT)	0,215	0,200	0,052	0,318	0,724
Romania (RO)	0,222	0,271	0,056	0,298	0,740
Russian Federation (RU)	0,350	0,242	0,069	0,245	0,802
Serbia (RS)	0,120	0,148	0,045	0,389	0,713
Slovakia (SK)	0,256	0,184	0,061	0,431	0,774
Slovenia (SI)	0,111	0,172	0,035	0,326	0,725
Spain (ES)	0,143	0,143	0,076	0,285	0,713
Sweden (SE)	0,255	0,346	0,099	0,663	0,801
Ukraine (UA)	0,204	0,200	0,031	0,449	0,714
United Kingdom (GB)	0,211	0,245	0,033	0,472	0,719
UK, Liquidation set	0,156	0,234	0,031	0,447	0,621
United States (US)	0,195	0,378	0,248	0,722	0,701

Average of column items	0,245	0,265	0,073	0,481	0,740
Correlation with Z′′-Score AUC	0,611	0,681	0,516	0,574	1,000

Legend:

WCTA = Working Capital/Total Assets

RETA = Retained Earnings/Total Assets

EBITTA = EBIT/Total Assets

BVETD = Book Value of Equity/Total Liabilities

SALTA = Sales/Total Assets

AUC = Area under the ROC curve

Z''-Score = Altman (1983) Z''-Score in the test data

Table 7. Test data AUCs for different estimated country models. Comparisons are with AUC of the LR model estimated for all data (benchmark).

	Test data AUC for different models:							
Country	Benchmark	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Austria (AT)	0,800	0,782	0,770-	0,774	0,829	0,787	0,707-	0,764
Belgium (BE)	0,772	0,779	0,780	0,826	0,776	0,783	0,781	0,834
Bosnia (BA)	0,862	0,850	0,832	0,833	0,838	0,838	0,820	0,853
Bulgaria (BG)	0,684	0,785	0,785	0,785	0,796	0,794	0,677	0,707
China (CN), ST companies	0,985	0,984	0,982	0,980	0,986	0,986	0,986	0,989
Colombia (CO)	0,726	0,755	0,754	0,818+	0,789	0,749	0,774	0,824+
Croatia (HR)	0,844	0,862	0,858	0,865	0,853	0,863	0,847	0,863
Czech Republic (CZ)	0,811	0,812	0,812	0,823	0,830	0,813	0,818	0,842
Denmark (DK)	0,803	0,806	0,809	0,833	0,818	0,804	0,813	0,843
Estonia (EE)	0,823	0,824	0,839	0,866	0,874	0,839	0,847	0,899
Finland (FI)	0,867	0,867	0,870	0,885	0,868	0,873	0,879	0,894
France (FR)	0,739	0,773	0,799	0,805	0,826	0,800	0,810	0,845
Germany (DE)	0,673	0,684	0,685	0,734	0,704	0,701	0,690	0,762
Hungary (HU)	0,742	0,736	0,743	0,788	0,767	0,765	0,747	0,820
Iceland (IS)	0,664	0,670	0,671	0,784+	0,683	0,714	0,708	0,824+
Ireland (IE)	0,679	0,663	0,666	0,718	0,661	0,665	0,680	0,724
Italy (IT)	0,806	0,802	0,810	0,835	0,839	0,809	0,816	0,868
Latvia (LV)	0,678	0,690	0,707	0,745	0,720	0,711	0,730	0,771
Netherlands (NL)	0,752	0,752	0,756	0,768	0,784	0,760	0,759	0,807
Norway (NO)	0,716	0,741	0,741	0,803	0,749	0,744	0,743	0,814
Poland (PL)	0,903	0,891	0,897	0,915	0,899	0,896	0,895	0,913
Portugal (PT)	0,741	0,753	0,757	0,761	0,780	0,758	0,765	0,797
Romania (RO)	0,758	0,718	0,773	0,805	0,757	0,755	0,753	0,774
Russian Federation (RU)	0,811	0,808	0,810	0,854	0,811	0,811	0,818	0,862
Slovakia (SK)	0,777	0,773	0,775	0,782	0,810	0,761	0,773	0,807

Slovenia (SI)	0,737	0,703	0,719	0,704	0,711	0,739	0,686	0,698
Spain (ES)	0,734	0,755	0,767	0,768	0,849	0,766	0,776	0,858
Sweden (SE)	0,813	0,826	0,829	0,843	0,831	0,831	0,892	0,901
Ukraine (UA)	0,708	0,712	0,712	0,720	0,742	0,718	0,718	0,758
United Kingdom (GB)	0,699	0,708	0,713	0,718	0,764	0,714	0,731	0,785
United States (US)	0,710	0,689	0,687	0,686	0,816+	0,689	0,714	0,816+
Average of AUC	0,768	0,773	0,778	0,801	0,799	0,782	0,779	0,823

Significance:

AUC better than benchmark: 0.0001 = ++++, 0.001 = +++, 0.01 = ++, 0.1 = +

AUC worse than benchmark: 0.0001 = ----, 0.001 = ---, 0.01 = --, 0.1 = -

Models:

Benchmark = The LR model estimated for all data with Z''-model (1983) variables

Model 1 = The MDA model estimated for country data

Model 2 = The LR model estimated for country data

Model 3 = The LR model estimated for country data with year dummies

Model 4 = The LR model estimated for country data with size variables

Model 5 = The LR model estimated for country data with age category dummies

Model 6 = The LR model estimated for country data with industry dummies

Model 7 = The LR model estimated for country data with all variables

Appendices

Appendix 1. Medians of the Altman (1983) model (Z'') variables, by status and country.

	Non-failed	Failed	Non-failed	Failed	Non-failed	Failed	Non-failed	Failed
Country	WCTA		RETA		EBITTA		BVETD	
Austria (AT)	0,170	-0,278	0,260	-0,145	0,055	-0,071	0,465	-0,022
Belgium (BE)	0,136	-0,087	0,157	-0,107	0,052	-0,025	0,460	0,029
Bosnia (BA)	0,087	-0,019	0,148	-0,101	0,032	-0,056	0,580	0,000
Bulgaria (BG)	0,216	0,102	0,272	-0,019	0,075	-0,019	0,504	0,068
China (CN)	0,105	0,037	0,064	0,032	0,064	0,048	0,870	0,590
China, delisted data	0,158	0,069	0,286	0,347	0,059	0,091	0,923	0,840
China, ST data	0,106	-0,192	0,281	-0,012	0,052	-0,087	0,776	0,308
Colombia (CO)	0,244	0,018	0,282	0,047	0,104	0,005	0,940	0,235
Croatia (HR)	0,093	-0,182	0,130	-0,205	0,030	-0,014	0,274	0,000
Czech Republic (CZ)	0,196	-0,208	0,230	-0,167	0,050	-0,019	0,591	-0,076
Denmark (DK)	0,128	-0,139	0,306	-0,050	0,005	-0,041	0,758	0,025
Estonia (EE)	0,222	-0,109	0,455	0,067	0,050	-0,063	1,169	0,136
Finland (FI)	0,233	-0,169	0,359	-0,247	0,082	-0,125	0,744	-0,163
France (FR)	0,146	0,006	0,223	0,010	0,059	-0,006	0,501	0,077
Germany (DE)	0,307	0,176	0,150	0,014	0,069	0,037	0,335	0,073
Greece (GR)	0,127	-0,044	0,049	-0,213	0,039	-0,010	0,460	0,111
Hungary (HU)	0,135	-0,040	0,271	0,047	0,054	0,002	0,717	0,130
Iceland (IS)	0,051	-0,195	0,046	-0,215	0,051	0,000	0,196	-0,141
Ireland (IE)	0,198	0,017	0,383	0,117	0,029	-0,010	0,728	0,168
Italy (IT)	0,107	-0,170	0,077	-0,087	0,032	-0,041	0,178	-0,029
Latvia (LV)	0,102	-0,015	0,150	0,030	0,069	0,027	0,342	0,088

Lithuania (LT)	0,205	-0,041	0,273	0,055	0,054	0,003	0,686	0,117
Netherlands (NL)	0,204	0,000	0,303	0,050	0,065	-0,012	0,511	0,079
Norway (NO)	0,185	0,028	0,139	-0,080	0,068	-0,047	0,398	0,069
Poland (PL)	0,198	-1,142	0,258	-0,662	0,076	-0,048	0,850	-0,501
Portugal (PT)	0,164	-0,051	0,133	-0,067	0,028	-0,024	0,343	0,025
Romania (RO)	0,071	-0,151	0,158	-0,113	0,041	-0,015	0,250	-0,048
Russian Federation (RU)	0,088	-0,262	0,106	-0,136	0,043	-0,026	0,194	-0,051
Serbia (RS)	0,093	-0,027	0,129	-0,019	0,027	-0,018	0,389	0,000
Slovakia (SK)	0,137	-0,119	0,185	0,001	0,060	-0,001	0,496	0,065
Slovenia (SI)	0,109	-0,002	0,215	0,043	0,039	0,004	0,434	0,108
Spain (ES)	0,117	-0,026	0,139	-0,004	0,029	-0,047	0,336	0,051
Sweden (SE)	0,266	0,011	0,357	0,011	0,058	-0,041	0,749	0,086
Ukraine (UA)	0,053	-0,151	0,034	-0,166	0,003	-0,028	0,442	-0,007
United Kingdom (GB)	0,179	-0,032	0,294	0,049	0,031	-0,002	0,579	0,107
UK, Liquidation set	0,179	0,023	0,294	0,060	0,031	0,000	0,579	0,132
United States (US)	0,164	-0,031	0,374	-0,004	0,003	-0,245	0,800	0,078
Average of medians§	0,153	-0,092	0,213	-0,050	0,048	-0,025	0,555	0,075

Legend:

WCTA = Working Capital/Total Assets

RETA = Retained Earnings/Total Assets

EBITTA = EBIT/Total assets

BVETD = Book Value of Equity/Total Liabilities

§ = Because the two datasets for non-failed UK firms are identical, the non-failed medians are used only once in the calculation of the averages.