Journal of International Financial Management & Accounting 28:2 2017

Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's *Z*-Score Model

doi:10.1111/jifm.12053

Edward I. Altman

NYU Salomon Center, Henry Kaufman Management Center, New York University, Stern School of Business, 44 West Fourth Street, New York, NY 10012, USA

Małgorzata Iwanicz-Drozdowska

Institute of Finance, Warsaw School of Economics, al. Niepodleglosci 162, 02-513 Warsaw, Poland
e-mail: miwani@sgh.waw.pl

Erkki K. Laitinen

University of Vaasa, P.O.Box 700, FI-65101 Vaasa, Finland
e-mail: erkki.k.laitinen@uwasa.fi

Arto Suvas

University of Vaasa, P.O.Box 700, FI-65101 Vaasa, Finland
e-mail: asuvas@welho.com

Abstract

This paper assesses the classification performance of the *Z*-Score model in predicting bankruptcy and other types of firm distress, with the goal of examining the model's usefulness for all parties, especially banks that operate internationally and need to assess the failure risk of firms. We analyze the performance of the *Z*-Score model for firms from 31 European and three non-European countries using different modifications of the original model. This study is the first to offer such a comprehensive international analysis. Except for the United States and China, the firms in the sample are primarily private, and include non-financial companies across all industrial sectors. We use the original Z''-Score model developed by Altman, *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy* (1983) for private and public manufacturing and non-manufacturing firms. While there is some evidence that *Z*-Score models of bankruptcy prediction have been outperformed by competing market-based or hazard models, in other studies, *Z*-Score models perform very well. With-

The authors are grateful to the Editor, Richard Levich, and to the anonymous referees for many helpful comments and suggestions. We also wish to thank participants at the 2014 International Risk Management Conference (IRMC) in Warsaw for useful discussions. Laitinen and Suvas thank the Foundation for Economic Education, and Jenny and Antti Wihuri Foundation for financial support. Laitinen also thanks OP-Pohjola Group Research Foundation for support.

out a comprehensive international comparison, however, the results of competing models are difficult to generalize. This study offers evidence that the general Z-Score model works reasonably well for most countries (the prediction accuracy is approximately 0.75) and classification accuracy can be improved further (above 0.90) by using country-specific estimation that incorporates additional variables.

1. Introduction

The first multivariate bankruptcy prediction model was developed by Altman (1968) in the late 1960s. 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 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 credit activity, and banks seek to reduce their own risk of default. Another issue of interest for bankers is capital adequacy and the internal ratings-based approach encouraged by the Basel Accords. The Z-Score model has become a prototype for many of these models. Asset managers and investors need reliable tools that can help them select appropriate companies for their portfolios. Financial distress is detrimental to investor returns, but risk may provide opportunities for high returns on short-sale strategies. Rating agencies assess the risk of the entities and of securities issues, and thus, they need a tool to predict default. Altman (1983) suggested that the management of distressed firms can utilize the Z-Score model as a guide to financial turnaround.

The approach used for bankruptcy prediction has evolved over time. Beaver (1966, 1968) used univariate analysis for selected ratios and found that some had very good predictive power. Altman (1968) made strides by developing a multiple discriminant analysis model (MDA) called the Z-Score model. The next two decades saw additional contributions to financial distress research.[1] For example, Ohlson (1980) proposed a logit model,[2] Taffler (1984) offered a Z-Score model for the United Kingdom, and Zmijewski (1984)[3] used a probit approach. Dimitras et al. (1996) reviewed 47 studies on business prediction models, summarizing the methods employed and the variety of ratios used. Discriminant analysis was the prevailing method, and the most important financial ratios came from the solvency category, with profitability ratios also being important.

Balcaen and Ooghe (2006) reviewed 43 models of business failure prediction which they classified into four categories: univariate models (1); risk index models (2); MDA models (21); and conditional probability models (19). However, their review omitted the rapidly growing type of models based on option pricing theory and contingent claims (e.g., Vassalou and Xing, 2004; commercialized into Kealhofer, McQuown and Vasicek's model, known as the KMV model), as well as hazard models (e.g., Shumway, 2001). Kumar and Ravi (2007) reviewed 128 statistical and artificial intelligence models for bank and firm bankruptcy predictions, paying special attention to the techniques used in the different models. These authors noted that neural networks were the most popular intelligence technique. In their review, Jackson and Wood (2013) presented the frequency of occurrence of specific forecasting techniques in the prior literature. The five most popular techniques were as follows: (1) multiple discriminant analysis, (2) logit models, (3) neural networks, (4) contingent claims, and (5) univariate analysis.

Recent reviews of the efficacy of these models have been offered by Agarwal and Taffler (2008), Das et al. (2009), and Bauer and Agarwal (2014). These reviews take into account the performance of accounting-based, market-based, and hazard models. These three model types prevail in the literature. According to Agarwal and Taffler (2008), there is little difference in the predictive accuracy of accounting-based and market-based models; however, the use of accounting-based models allows for a higher level of risk-adjusted return on credit activity. Das et al. (2009) showed that accounting-based models perform comparably to the Merton structural, market-based approach for credit default spread (CDS) estimation. However, a comprehensive model, which used both sources of variables, outperformed the other models. In Bauer and Agarwal (2014), hazard models using accounting and market information (Shumway, 2001; Campbell et al., 2008) were compared with two other approaches: the original Taffler (1984) Z-score model, which was tested by Agarwal and Taffler (2008), and a contingent claims model using Bharath and Shumway's (2008) approach. Using U.K. data, the hazard models were superior in bankruptcy prediction accuracy, ROC (Receiver Operating Characteristic) analysis, and information content.

Even though the Z-Score model was developed more than 45 years ago and many alternative failure prediction models exist, the Z-Score model continues to be used worldwide as a main or supporting tool for bankruptcy or financial distress prediction and analysis both in research and in practice. We focus on accounting-based versions of the *Z*-Score models, which even though they are occasionally outperformed by other models, do not rely on market data. Most firms operating in business are privately held; hence, only accounting data and no market data (e.g., stock prices) are available. Private firms are usually financed by banks, which are obligated to assess their creditworthiness and monitor their performance. In the case of internationally active banks, from a regulatory perspective it is especially important to use a single model for distress prediction, provisioning, and economic capital calculation. According to current Basel regulatory requirements, banks need to validate their distress prediction models and document their efficacy. Thus, it is important to analyze the performance of accounting-based models in an international context.

In our study, we use a large international sample of firms to assess the classification performance of the *Z*-Score model in bankruptcy prediction.[4] We analyze the model's performance for firms from 31 European and three non-European countries (China, Colombia and the United States). These firms are mostly privately held, and a large number are from non-manufacturing industries. We use the version of the model developed by Altman (1983) for private and public manufacturing and non-manufacturing firms (the *Z*"-Score model). Such an extensive international analysis of the *Z*-Score model's performance has not been presented to date. We regard our review and analysis as important contributions to the economic literature.

The remainder of the paper is structured as follows. In the next section, we summarize the original *Z*-Score model (Altman, 1968) and its extension for private firms, that is, the *Z*'-Score and *Z*"-Score models (Altman, 1983). In the third section, we present the results and conclusions from the literature review on these models. The fourth section presents seven hypotheses on the performance of the *Z*"-Score model that we will subject to empirical analysis. In the fifth section, we discuss the empirical data and statistical methods, 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's (1968) initial sample was composed of 66 corporations, with 33 firms in each of two groups. The bankrupt group (Group 1) consisted of manufacturers that filed bankruptcy petitions under Chapter X of the National Bankruptcy Act during the 1946–1965 period. The mean asset size of these firms was 6.4 million USD, ranging between 0.7 and 25.9 million USD. Altman recognized that this group was not homogenous with respect to size and industry, although all firms were relatively small and from manufacturing industries. He attempted to carefully select 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 USD. Altman eliminated small firms (less than 1 million U.S.A. dollars in total assets) because of a lack of data and very large firms because of the rarity of bankruptcies among these firms in that period. He did not match the asset size of the two groups exactly, and therefore, the firms in Group 2 were 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 were derived from financial statements one reporting period prior to bankruptcy.

Using financial statements, Altman compiled a list of 22 potentially important financial ratios for evaluation. He classified these variables into five standard ratio categories: liquidity, profitability, leverage, solvency, and activity. These ratios were chosen based on their popularity in the literature and their potential relevance to the study. The final discriminant function estimated by Altman (1968) is as follows:

$$Z = 0.012 \cdot X_1 + 0.014 \cdot X_2 + 0.033 \cdot X_3 + 0.006 \cdot X_4 + 0.999 \cdot X_5 \qquad (1)$$

where X_1 = Working Capital/Total Assets; X_2 = Retained Earnings/Total Assets; X_3 = Earnings before Interest and Taxes/Total Assets; X_4 = Market Value of Equity/Book Value of Total Liabilities; X_5 = Sales/Total Assets; Z = Overall Index.

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 intended for publicly traded firms and that *ad hoc* adjustments are not scientifically valid. 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:

$$Z' = 0.717 \cdot X_1 + 0.847 \cdot X_2 + 3.107 \cdot X_3 + 0.420 \cdot X_4 + 0.998 \cdot X_5 \qquad (2)$$

where X_4 = Book value of equity/Book value of total liabilities, with the other variables the same as those in the original (1968) Z-Score model.

Due to the lack of a private firm database, Altman did not test the Z'-Score model on a secondary sample. However, he analyzed the accuracy of a four-variable Z"-Score model that excluded the Sales/Total assets ratio, X_5, from the revised model because of a potential industry effect that is more likely to take place when this kind of industry-sensitive variable (asset turnover) is included in the model. Altman then estimated the following four-variable Z"-Score model (Altman, 1983):

$$Z'' = 3.25 + 6.56 \cdot X_1 + 3.26 \cdot X_2 + 6.72 \cdot X_3 + 1.05 \cdot X_4 \qquad (3)$$

The EBIT/Total assets ratio, X_3, contributed most to the discrimination power in this version of the model. The classification results for the Z"-Score model were identical to the revised five-variable Z'-Score model. In the current study, our empirical analysis focuses on the performance of the Z"-Score model version in predicting bankruptcy, where it has its widest scope, as it is intended for both privately held and publicly listed firms and for both manufacturing and non-manufacturing firms.

3. Survey of Literature Related to the Altman *Z*-Score Model

We focus on papers published after 2000 in prominent international journals and books.[5] Of the many articles and books identified, we selected 31 articles in which the Z-Score was either used as a failure prediction proxy or assessed mostly in terms of predictive ability. Of the 31 studies, Altman's Z-Score model was used in 16 cases as the measure of distress or of financial strength.[6] In 13 studies, Altman's original model was modified and (or) verified, including re-estimation, and in two cases, it was used solely for the robustness check. As Pindado et al. (2008) noted, the Z-Score was also used for other purposes, such as the evaluation of the costs and benefits of covenants in bonds, the choice of debt type (bank versus non-bank, private or public), and the relationship between investment and internal funds. We focused on this part of the literature that verified and (or) modified Altman's original model. The broad use of the *Z*-Score model for measuring financial distress and performing robustness checks indicates its acceptability as a reasonable, simple, and consistent measure of distressed firms.

The most common modification to the *Z*-Score model was the use of other estimation techniques or country-specific data. The use of Altman's ratios in combination with techniques other than MDA improved its prediction capability. The application of new data on both United States and non-US firms also improved model performance. Kwak et al. (2005) used Multiple Criteria Linear Programming (MCLP) to model 5 Altman and 9 Ohlson variables with data on bankrupt US firms from 1992 to 1998 and nearly six times more matched U.S.A. control firms. The MCLP approach performed better than Altman's original model and gave results similar to or better than those of Ohlson's original model. The original models were not recalculated, and the authors referred solely to their original prediction rates.

Merkevicius et al. (2006), using data on United States and Lithuanian firms, developed a hybrid artificial discriminant model combining MDA and an unsupervised learning artificial neural network. This hybrid SOM-Altman model reached a high prediction rate of 92.35 per cent. Xu and Zhang (2009) applied Altman's *Z*-score, Ohlson's *O*-score, and Merton's distance-to-default (*D*-score) models to Japanese firms to check whether these models are useful for bankruptcy prediction in Japan. They also "merged" these models into a new *C*-score model. They then introduced variables unique to Japan to check whether corporate structure variables have any impact on the probability of bankruptcy; they called this the *X*-score model. These two models were useful for Japanese firms in predicting bankruptcy, but the market-based model was the most successful. In summary, the *C*-score and *X*-score (with country-specific variables) models improved bankruptcy prediction.

Tinoco and Wilson (2013) used the original *Z*-Score as one of the benchmarks to assess the performance of their model developed for U.K. listed companies with combined accounting, market, and macroeconomic data. Altman's *Z*-Score presented very good classification accuracy in the case of financially distressed firms (81 versus 87 per cent for the new model); however, it was less correct for non-distressed firm's prediction. Another modification was introduced by Lyandres and Zhdanov (2013), who posed the question of whether the inclusion of variables related to investment opportunities improved the predictive power of three models (Altman's *Z*-score model and Zmijewski's and Shumway's models). They used three proxies for investment opportunities (market-to-book, value-to-book, and R&D-to-assets). The measures of investment opportunities were linked to the likelihood of default. The inclusion of either of these measures improved the out-of-sample forecasting ability of all three.

The verification of Altman's model concentrated on its efficacy or on how it compares with other accounting-based, market-based, or hazard models. However, although the original *Z*-Score model was not solely based on accounting data because the market value of equity was utilized, we classify it here as accounting-based. Grice and Ingram (2001) used a novel dataset of US firms and posed three questions about the efficacy of Altman's model, concluding that the prediction accuracy of Altman's model had declined over time and that the coefficients of the model had significantly changed, which means that the relation between the financial ratios and the signs of financial distress had changed over time. The model was sensitive to industry classification (more efficient for manufacturing firms than for non-manufacturing firms) but was not sensitive to the type of financial distress. Similar conclusions were drawn by Grice and Dugan (2003) regarding Ohlson's (1980) and Zmijewski's (1984) models.

Hillegeist et al. (2004) compared Altman's *Z*-score and Ohlson's *O*-score (with original and updated coefficients) with a model based on Black-Scholes-Merton (BSM) option pricing (a so-called BSM-Prob model). Hillegeist et al. used relative information content tests to compare the out-of-sample performance of these various models and determined that BSM-Prob outperformed the alternative accounting-based models. The conclusions were robust to various modifications of accounting-based models, such as updated coefficients, industry effects, and the separation of variables. Chava and Jarrow (2004) employed an extended bankruptcy database of U.S.A. listed firms to test the superiority of Shumway's model (2001) over Altman's (1968) and Zmijewski's (1984) models. The authors re-estimated the models over the 1962–1990 period and forecasted bankruptcies over the 1991–1999 period. In the case of Shumway's model, 74.4 per cent (in the first decile) of the bankruptcies were correctly identified; with Altman's model, 63.2 per cent; and with Zmijewski's model, 43.2 per cent. Shumway's market-based model also outperformed accounting-based models in terms of the ROC curve (0.91).

Reisz and Perlich (2007) developed a model incorporating barrier options for bankruptcy prediction and compared its discriminatory power with other market-based models and Altman's *Z*-Score and *Z*"-Score. The dataset covered nearly 6000 industrial firms over the 1988–2002 period. The authors documented the superiority of Altman's *Z*-Score and *Z*"-Score models for short-term (up to 1 year) bankruptcy prediction. For medium- and long-term bankruptcy prediction, their barrier option model outperformed the other models.

Pindado et al. (2008) developed an *ex ante* model for the estimation of financial distress likelihood (FDL) using a panel data methodology and presented a financially (not legally) based definition of distress. Their sample covered 1,583 U.S.A. companies and 2,250 companies from other G-7 countries for the 1990–2002 period. They used a re-estimated *Z*-Score as a benchmark. The FDL model outperformed the *Z*-Score model in terms of stability and classification power for different countries and periods. In the case of the re-estimated *Z*-Score model, only profitability and retained earnings maintained their significance for different years and countries.

Wu et al. (2010) evaluated the performance of five models (Altman, 1968; Ohlson, 1980 Zmijewski, 1984; Shumway, 2001; Hillegeist et al., 2004) using an up-to-date dataset for U.S.A. listed firms. Based on these models, the authors built their own integrated model, that is, a multi-period logit model with an expanded set of variables. The integrated model, which combined accounting and market data, as well as firms' characteristics, outperformed the other models. Altman's *Z*-score performed poorly compared with the other four models. Shumway's model performed best, Hillegeist et al.'s model performed adequately, and Ohlson's and Zmijewski's models performed adequately, although their performance deteriorated over time.

Jackson and Wood (2013) tested 13 different models of bankruptcy prediction and assessed their efficacy using the ROC curve. They selected three single-variable models, three accounting-based models (including Altman's *Z*-Score) in two versions (with updated coefficients and a neural network approach), and four contingent claims models, and the latter group outperformed the other models. The four best models were contingent claims models based on European call and barrier options. Although the predictive performance improved with the application of the neural network to accounting-based models, it was still lower than with the market-based models.

Acosta-González and Fernández-Rodríguez (2014) used genetic algorithms with the Schwarz information criterion (GASIC) for variable selection combined with the logit model for bankruptcy prediction. Altman's Z-Score model was used as one of two benchmarks for the authors' model evaluation. For one-step-ahead forecasting, Altman's model was better at predicting failed firms, but the type II error was high. For two- and three-steps-ahead forecasts, the performance of the models was similar for failed firms, but for non-failed firms, and the prediction accuracy of Altman's model was worse. For four-steps-ahead forecasts, the GASIC model outperformed the other models for failed firms, but it performed comparably for non-failed firms.

In general, for the 31 articles we reviewed, Altman's Z-Score model underperformed compared with market-based models, but evidence indicated that it performed well for short-term distress prediction. The question of whether market-based models perform better than accounting-based models has been raised many times (e.g., discussion in Das et al., 2009; Bauer and Agarwal, 2014). Our purpose is not to contribute to this strand of research but rather to focus on the accounting-based approach. In this study, we primarily analyze privately held firms; by definition, there are no market data for these firms. In this case, an accounting-based approach is the only solution applicable by banks as lenders, or by investors holding debt securities of firms not listed on the stock exchange. Thus far, most studies have concentrated on the U.S.A. market; only a few of them have used data from other countries, such as Japan, the U.K., Lithuania, or the G-7 countries. Our analysis is a significant extension of the previous research.

4. Methodology and Research Hypotheses

The literature survey shows that the Z-Score model (publicly traded firms), the Z'-Score model (private manufacturing firms), and the Z''-Score model (private and publicly traded manufacturing and non-manufacturing firms) have been adapted for different purposes. In this study, we are interested first in assessing the performance of the original Z''-Score model in classifying bankrupt and non-bankrupt firms in an international context, with special focus on the European market.

However, we also validate the results in a set of non-European countries to generalize the results for circumstances outside Europe. Second, we re-estimate the model using extensive international data and then use the re-estimated *Z*"-Score model as a benchmark for assessing the effects of different factors on the model's performance in terms of classification accuracy. We 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. We test a set of hypotheses based on the effects of the model on performance on two different levels. First, we test a set of hypotheses on a pooled set of all firms and, second, on data from each country individually. Compared with previous research, this study's contribution is its focus on an international context, not simply model application or re-estimation of a given country's data. Because we focus on the performance of the *Z*"-Score model when using a large body of international data, the research hypotheses are of a technical nature and are given as follows.

4.1. H1: Obsolescence of the Coefficients

The *Z*"-Score model was originally estimated using the same sample of firms used to develop the *Z*-Score model. The bankruptcies in the estimation data occurred during the 1946–1965 period. Thus, the oldest observations are from nearly 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 firms' financial behavior and their business environment have significantly changed since then. The importance of the financial ratios, as reflected by the coefficients of the model, may differ from their original importance. Therefore, we suggest, as the first hypothesis (*H1*), 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. This hypothesis is supported by the previous research (e.g., Grice and Ingram, 2001) and by practice. *H1* aims to support this evidence on the international level and is expressed as follows:

> ***H1***: Re-estimating the coefficients of the *Z*"-Score model improves its classification accuracy.

4.2. H2: Method of Estimation

The original *Z*”-Score model has been estimated using MDA. However, MDA is based on the ordinary least squares (OLS) method and thus requires assumptions of multinormality, homoscedasticity, and linearity, which are not often met in empirical financial ratio analysis. We re-estimate the *Z*”-Score model using logistic regression analysis (LRA) to assess the effect of the estimation method. LRA does not require most of the restricting assumptions of MDA. In LRA, the multivariate normality of the independent variables is not required, nor are homoscedasticity and linearity. 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. Our second hypothesis (*H2*) is that the classification performance of the re-estimated *Z*”-Score model will improve when it is estimated using LRA instead of MDA. The model re-estimated for the original variables using LRA and all pooled data arecalled the *Z*”-Score LR model. The performance of this re-estimated model is used as the benchmark for further analyses. Thus, the second hypothesis is as follows:

> ***H2***: The prediction accuracy of the logistic regression version of the *Z*”-Score model is higher than that of the multiple discriminant analysis version.

4.3. H3: Bankruptcy Year

The model based on the relationship between bankruptcy and financial ratios is likely to be affected by the macroeconomic environment. These effects may significantly decrease the classification accuracy of the model. If the model is estimated using data from 1 year and will be applied to data from another year, the validity of the model can be questioned. In terms of economic growth, credit policy, and interest rates, business cycles can affect the boundary between bankrupt and non-bankrupt firms. The original *Z*”-Score model is estimated using data from the 1946–1965 period, which includes several business cycles. Therefore, the model is not focused on any specific stage of a cycle and does not explicitly take into account the bankruptcy year. Altman (1983) suggested gathering data from firms for the most recent few years when developing a prediction model. In this study, the benchmark *Z*"-Score LR model is estimated for a shorter period than in the original estimation, although it covers several recent years for different stages of the business cycle in different countries. The third hypothesis (*H3*) is that the classification accuracy of the benchmark model can be increased by explicitly taking into account the year of bankruptcy in the model estimation. The third hypothesis is as follows:

> ***H3*****:** The model's prediction accuracy is higher when the effect of the year of bankruptcy is included.

4.4. H4: Size of the Firm

The boundary between bankrupt and non-bankrupt firms is different for small and large firms, which decreases the performance of the model estimation when data from one size category are applied to another size category. For the bankrupt and non-bankrupt firms in the original data for *Z*"-Score model estimation, asset sizes ranged between approximately 1 and 25 million U.S.A. dollars. The data did not include very small or very large firms. Altman (1983) regarded the suitability of the original *Z*-Score model (and, likewise, the *Z*"-Score model) for all firms as debatable because of this omission. In the current study, the benchmark *Z*"-Score LR model is estimated for data from many size categories, from very small to very large firms. The fourth hypothesis (*H4*) assumes that the classification performance of the uniform benchmark LR model based on the original four financial variables of the *Z*"-Score model is improved when the size category of the firm is explicitly taken into account. Thus, we present the fourth hypothesis:

> ***H4*****:** The model's prediction accuracy is higher when the effect of size is included.

4.5. H5: Age of the Firm

International insolvency statistics generally show that bankruptcy risk is a function of the age of the firm. Very young firms typically show very high risk. The original Z"-Score model does not explicitly take age into account. However, Altman (1983) noted that the age of a firm is implicitly considered in the Retained Earnings/Total Assets ratio (X_2), which was regarded as a new ratio in the bankruptcy prediction context. A relatively young firm will probably show a low ratio because it has not had time to build up cumulative profits. Thus, a young firm is, to some degree, discriminated against in the model, and its likelihood of being classified as bankrupt is relatively higher than that of an older firm. The incidence of failure is much higher in the early years of a firm. Although the age of the firm is implicitly taken into account in X_2, we expect that an explicit consideration of age will improve the classification accuracy by controlling for the age factor. The fifth hypothesis (*H5*) proposes that the performance of the uniform benchmark model based on the original four financial variables of the Z"-Score model increases when the age of the firm is explicitly taken into account. The fifth hypothesis is as follows:

> ***H5*:** The model's prediction accuracy is higher when the effect of firm age is included.

4.6. H6: Industry of the Firm

The original Z'-Score model was estimated only for manufacturing firms. Altman (1983) stated that it would be ideal 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 that grouping. Previous studies show that financial distress analysis is influenced by the 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), which showed the least significance on a univariate basis while making a very significant contribution to the discriminant power of the multivariate model. Altman (1983) recognized the potential industry effect due to the wide variation among industries in asset turnover and specified the Z"-Score model without X_5. However, the Z"-Score model was 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. The sixth hypothesis (*H6*) assumes that an explicit consideration of industry will improve the classification accuracy of this benchmark model. *H6* can be expressed in the following form:

> ***H6*:** The model's prediction accuracy is higher when the effect of industry is included.

4.7. H7: Country of Origin

The original *Z*"-Score model has been estimated only for U.S.A. firms. It can be expected that the international applicability of the model to other countries is affected by country-specific differences. The economic environment, legislation, culture, financial markets, and accounting practices in a country may affect the financial behavior of firms and the boundary between bankrupt and non-bankrupt firms. These factors may weaken the classification performance of the model for countries other than that for which the model was originally estimated (Ooghe and Balcaen, 2007). The seventh hypothesis (*H7*) assumes that explicitly taking the country of origin of a firm into account will improve the classification accuracy of the benchmark model. In our empirical study, the country effect is assessed by including a variable for country risk. The seventh hypothesis is as follows:

> ***H7*:** The model's prediction accuracy is higher when the effect of country risk is included.

5. Empirical Data and Statistical Methods

5.1. Sample of Firms

The principal data for this study were extracted from the ORBIS databases of Bureau Van Dijk (BvD). ORBIS Europe is a commercial database that, at the moment of sampling, contained administrative information on more than 50 million European firms. However, income statement and balance sheet information was available for approximately 8 million companies. More than 99 per cent of the companies covered in this database are private companies from various industries, justifying the use of the *Z*"-Score model instead of the original *Z*-Score model. The *Z*"-Score model was originally made robust across all industrial groupings and for both private and public entities (Altman, 1983, 2014; Altman and Hotchkiss, 2006). Because we do not want to limit the scope of the *Z*"-Score model in this study, we retain both private and public firms from all industrial groupings.

The ORBIS formats have been derived from the world's most commonly used formats for the presentation of business accounts (Ribeiro et al., 2010). International comparability may be a problem when administrative firm-level data are pooled across countries. Although the definition of variables is usually less harmonized for administrative data, 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.

A number of factors influence the international applicability of bankruptcy prediction models: accounting legislation and practice, creditor rights and investor protection, judicial efficiency, corporate governance, bankruptcy protection and insolvency management, and firm risk-taking. These factors strongly differ between European and non-European countries. Therefore, we aim to test the performance of the *Z*"-Score model outside Europe. First, it is particularly important to include the US because it is the country of origin for the *Z*"-Score model and because it has the largest market capitalization in the world. Second, a central motivation in developing the modified *Z*"-Score model was to make it applicable to emerging market companies. We include firms from China and Colombia, which represent two very culturally and institutionally dissimilar emerging market countries. For other non-European countries, sufficient bankruptcy data (more than 60 bankrupt firms, i.e., the limit we set for European countries) from ORBIS World were not available. Thus, the results are also estimated and tested for three non-European countries (the United States, China, and Colombia) to gain a more global view of the *Z*"-Score model's performance. The samples of firms from these countries were extracted from ORBIS World, which contains middle-sized (total assets over 1.5 million EUR) and larger firms from around the world.

Several requirements are set for the statistical sampling of the empirical data. First, we require that the firm to be selected be an industrial (non-financial) company. Second, its owners must have limited liability (so partnerships and sole proprietors are left out of the study). Third, we set a minimum requirement for the size of the firm. Because the financial ratios of very small firms are generally too unstable for a failure prediction model, these firms are excluded (see Balcaen and Ooghe, 2006). We require that a firm's total assets must have exceeded 100 thousand EUR at least once in the available time series. Fourth, we include in our estimation sample firms from all European countries and three pre-selected non-European countries where the number of failed firms is greater than 60. If the number of failed firms for any European country is less than 60, the firms from this country are included only in the test sample. For qualifying European countries, firms are randomly classified in the estimation and test samples so that the number of firms is approximately equal in both samples. Thus, our estimation data include firms from 28 European and three non-European countries. Fifth, all failed firms that fulfill the above requirements are included in 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 most recent financial statements for failed firms in the database are from a financial period within 2007 and 2010, earlier years are also excluded for non-failed firms for comparability. All qualifying observations of non-failed firms from 2007 to 2010 are included in the datasets. We restrict the analyses of failed firms to the most recent financial statements available before failure. The four independent variables of the *Z*''-Score model were winsorized at 1 and 99 per cent to minimize outliers.

Table 1 shows the resulting number of non-failed and failed firms in the estimation data and test data by country. The estimation sample includes data from 2,602,563 non-failed and 38,215 failed firms from 28 European and three non-European countries. The test sample is slightly larger because it includes data from 31 European and three non-European countries. For the country of origin of the *Z*''-Score model, the United States, the estimation sample includes only 56 bankrupt firms. The available U.S.A. data consist only of listed (and delisted) firms. From China, there are three sub-samples. Only 32 public firms with special treatment (ST)[7] status are included in the estima-

Table 1. *Number of Observations by Country*

	Estimation data		*Test data*	
Country	*Non-failed*	*Failed*	*Non-failed*	*Failed*
Austria (AT)	7,430	55	7,526	44
Belgium (BE)	179,979	2,994	179,818	2,944
Bosnia (BA)	29,391	35	29,139	32
Bulgaria (BG)	50,041	48	42,351	44
China (CN)			39,315	198
China, delisted dataset, DL			16,291	29
China, ST data	846	16	1,020	16
Colombia (CO)	8,366	139	6,982	125
Greece (GR)			51,763	28
Croatia (HR)	59,541	249	58,478	275
Czech Republic (CZ)	92,835	556	92,562	564
Denmark (DK)	167,934	1,334	168,538	1,398
Estonia (EE)	34,313	234	34,196	242
Finland (FI)	90,878	481	91,227	459
France (FR)	160,749	6,124	161,653	6,318
Germany (DE)	98,814	910	99,496	921
Hungary (HU)	19,421	303	20,155	313
Iceland (IS)	17,399	248	17,624	243
Ireland (IE)	6,665	121	6,406	139
Italy (IT)	167,113	8,101	166,258	8,124
Latvia (LV)	8,064	433	8,241	477
Lithuania (LT)			10,000	56
Netherlands (NL)	20,885	154	15,854	147
Norway (NO)	172,467	1,294	170,985	1,206
Poland (PL)	87,200	291	86,233	264
Portugal (PT)	180,114	3,390	178,646	3,422
Romania (RO)	161,992	97	164,259	93
Russian Federation (RU)	116,903	2,534	115,711	2,481
Serbia (RS)			100,100	68
Slovakia (SK)	7,856	120	7,788	124
Slovenia (SI)	14,419	59	14,081	41
Spain (ES)	156,746	3,036	158,122	2,991
Sweden (SE)	169,810	2,256	169,999	2,314
Ukraine (UA)	133,342	1,787	133,980	1,765
United Kingdom (GB)	171,493	760	170,930	716
U.K. (GB), liquidation dataset			342,423	4,990
United States (US)	9,557	56	9,929	53
Total	2,602,563	38,215	3,148,079	43,664

tion sample.[8] The Chinese datasets of predominantly private firms (CN) and of public firms with delisted (DL) failure status are analyzed separately only for the test data.[9] ST firms are listed firms suffering from serious financial difficulties. Excluding the special U.S.A. data and the two non-private Chinese datasets, 99.4 per cent of the observations in the data are private firms.

5.2. Status of Failed Firms

ORBIS has five classes for potentially active firms (active, default of payment, receivership, dormant, and branch) and seven classes for inactive firms that no longer carry out business activities (bankruptcy, dissolved, dissolved-merger, dissolved-demerger, in liquidation, branch, and no precision). Among these classes, only active is selected to represent non-distressed firms. In selecting the failed firms, we try to avoid ambiguity 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 as failed even if they are active. These firms generally suffer from serious financial distress. However, firms in liquidation are generally not included in the sample of failed firms. 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 cases where failed firms are coded under a different status heading. These countries or samples are the following:

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
U.K., liquidation set	In liquidation
China, ST	Active (special treatment)
China, delisted, DL	Active (delisted)

If no such category of failed firms could be identified, that country was excluded from the study (for example, Switzerland). If a country had only a very small number of failed firms, it was dropped from the study (typically small countries, including Luxembourg, Liechtenstein, and Montenegro). 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 because it includes samples with three different criteria of failure (bankruptcy, special treatment, and delisted). Additionally, for the United Kingdom, there are two different samples (liquidation and receivership).

5.3. Statistical Methods

In this study, seven research hypotheses are drawn for statistical testing. The statistical analysis begins with calculating the original *Z*"-Score for the firms in the data, as in equation (3). The classification performance of the original model is assessed by the AUC (Area Under Curve) measure extracted from the ROC curve. AUC has a close connection with the Accuracy Ratio (AR) because AR = 2 · AUC – 1. AR equals 0 for a random model, 1 for a perfect model, and 0.5 for a model with an average classification performance. SAS software (SAS Institute Inc., Cary, NC, USA) is used for all statistical analyses.

The first hypothesis (*H1*) assumes that the coefficients of the original model are obsolete. *H1* is tested by re-estimating the coefficients of the *Z*"-Score model using the original statistical method (multiple discriminant analysis, or MDA). The problem is that the estimation sample includes different numbers of failed and non-failed firms from 31 countries. In the original *Z*-Score" sample (1983), equal numbers of bankrupt and non-bankrupt firms were selected from the US Following the characteristics of these data. Therefore, we weight the failed and non-failed firms equally. In so doing, the non-proportional sampling from different countries will not affect the re-estimated model. The number of firms from different countries, however, varies significantly, leading 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 that based on the original model.

The second hypothesis (*H2*) tests whether the classification performance of the re-estimated *Z*"-Score model improves when it is re-estimated using logistic regression analysis (LRA), which is based on less-restrictive statistical assumptions than MDA. In this estimation, the dependent variable $Y = 0$ for non-failed firms and $Y = 1$ for failed firms. LRA does not require that independent variables be multivariate normal or that groups have equal covariance matrices, which are basic assumptions in MDA (Hosmer and Lemeshow, 1989). LRA creates a (logit) score 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:

$$p(Y=1|X)=\frac{1}{1+e^{-L}}=\frac{1}{1+e^{-(b_0+b_1X_1+\ldots+b_4X_4)}} \tag{4}$$

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.[10] The effect of this 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 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 the remaining research hypotheses.

The third hypothesis (*H3*) is associated with the performance effect of taking account of the bankruptcy year in the estimation. This hypothesis is tested by estimating an LR model based on the following logit:

$$L=b_0+\sum_{i=1}^{4}b_iX_i+\sum_{j=1}^{3}c_jD_j \tag{5}$$

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, which is not given in the database, but rather to the last available year before bankruptcy. For failed firms, there is an approximately 1–2-year lead time to failure from this year. In this model, the year 2007 is the base category. If the AUC of this extended LR model statistically significantly exceeds the AUC of the *Z*"-Score LR model (benchmark), the evidence supports hypothesis *H3*.

Research hypotheses *H4–H7* are tested using the same approach as the third hypothesis above. However, for each hypothesis, appropriate variables are used instead of the year dummies. Hypothesis *H4* addresses the performance effect of taking size into account, and it is tested by adding two additional variables measuring firm size to the LR model. In this LR model, size is measured by the natural logarithm of total assets and its squared form. In this way, the effect of logarithmic size can be reflected in a function following the second-order parabola. Hypothesis *H5* tests whether the classification performance improves when the age of the firm is explicitly taken into account. When testing this hypothesis, the 6–14 years category is used as the base category, and two dummy variables are incorporated in the LR model (D_1: less than 6 years, D_2: 15 years or more). Hypothesis *H6* looks at whether classification performance is affected by the explicit consideration of industry effects. The hypothesis is tested here using dummy variables for seven industries (D_1: restaurants and hotels; D_2: construction; D_3: wholesale and retailing; D_4: agriculture; D_5: manufacturing; D_6: energy and water production; D_7: information technology), with all other industries acting as the base category.

Hypothesis *H7* tests whether the explicit consideration of the country of origin will improve classification performance. This hypothesis is tested by using country risk measures instead of dummy variables for countries. The country risk of each country is measured by Standard & Poor's Country Risk Rating per 6 months after the annual closing of accounts. The rating is numerically recoded such that the best rating, AAA, equals 1, the second-best rating, AA+, equals 2, and so on. Finally, the lowest rating, *D*, equals 22. Thus, *H7* is tested by estimating an 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 original four financial ratios and the additional variables specified in the hypotheses are estimated for all data. In addition, an 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. Hypothesis *H7* is not included in this country-level testing 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 2 presents descriptive statistics of the four independent variables (X_1–X_4) of the *Z*"-Score model for all data. The variation in the ratios

Table 2. *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

WCTA, Working Capital/Total Assets; RETA, Retained Earnings/Total Assets; EBITTA, EBIT/Total Assets; BVETD, Book Value of Equity/Total Liabilities.

is significant, as 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, indicating a symmetry of distributions. However, this is not the case for failed firms. For failed/distressed firms, the median exceeds the mean for these three ratios, indicating negatively skewed distributions. For X_4 (BVETD), 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, which is consistent with our expectations. The difference between the means of non-failed and failed firms is larger in the original U.S.A. data than in our all data for Retained Earnings/Total Assets (RETA) and EBIT/Total Assets (EBITTA) but is approximately the same size for Working Capital/Total Assets (WCTA) and Book Value of Equity/Total Liabilities (BVETD) (Altman, 1983). These characteristics of the data may indicate lower classification accuracy than in the original sample.

Table 3 presents the coefficients of the different models estimated for all data. All LRA estimates (Model 2 to Model 9) are statistically significant at 0.0001 due to their contributions and to the large sample size. The first column presents the coefficients of the original Z"-Score model. The "Model 1" column shows the coefficients when they are re-estimated by the same statistical method, specifically MDA. The coefficients here are negative because the models are estimated using

Table 3. *The Coefficients of the Different Models Estimated for all Data*

Variable	*Z"-Score*	*Coefficients for different statistical models*							
		Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*	*Model 8*
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
15 years or more						−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
Wholesale 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 all 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 rankings; Model 8 = The LR model estimated for all data with all variables.

$Y = 1$ for the failed firms. In our all data, EBITTA has a significantly higher relative weight than in the original U.S.A. data, while the weights of WCTA and BVETD have proportionally decreased. The re-estimated coefficient of BVETD is very small, indicating a minor effect on the logit. The “Model 2” column presents the coefficients for the Z”-Score LR model. These coefficients are directly comparable with those of the MDA model, as expected for this exceptionally large sample. For each model, the coefficient of BVETD is very close to zero. The differences in the coefficients of the original four variables among the eight LR models (Models 1–8) are small, indicating that the original four variables and the additional variables are quite independent of each other.

Table 3 also shows the coefficients of the additional variables 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 significantly decreased year by year. The base year of 2007 indicates that a failure emerged during 2008–2009 as a result of the global financial crisis. The crisis played a significant role in the failure of key businesses and caused a downturn in economic activity, leading to the 2008–2012 recession. The effects were especially felt in Europe. The coefficients of Model 4 for the size variables show that the contribution of size to the logit (risk measure) reaches its maximum value when logarithmic total assets are 15 or when total assets are approximately 3.3 million EUR. Model 5 confirms the riskiness of young firms because the risk of failure is very high for newly founded firms (less than 6 years old), as 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 (because of the large sample) but negative and very close to zero. Finally, the coefficients of all variables in Model 8 are directly comparable to those in Models 3–7.

6.2. All Data: Performance of the Z”-Score Models

Table 4 shows the 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 because AUC = 0.743 refers to AR = 0.486, which is approximately average accuracy (0.5). However, the score gives relatively good

Table 4. *Test Data AUCs for Different Countries, Based on All Data Model Versions. Comparisons are with the 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

Table 4 *(Continued)*

	Test data AUC for different models								
Country	*Benchmark*	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*	*Model 8*
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
U.K., liquidation dataset	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 firms, DL	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 rankings; Model 8 = The LR model estimated for all data with all variables.

Table 5. *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, DL	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
U.K., liquidation dataset	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

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.

results (AUC > 0.8) for China (ST firms), Poland, Finland, Estonia, the 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 AUCs for the countries included only in the test data. The performance of the score is very low in the Chinese CN (primarily private) and DL (delisted) samples and for liquidation firms in the United Kingdom.

The Appendix shows the medians of the four ratios (X_1–X_4) by status and country. Table 5 presents the differences of these medians between non-failed and failed firms by country. This table also presents the AUC of the *Z*"-Score and its correlation with the difference of medians, which 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, which implies, with the exceptionally high AUC, that ST firms systematically differ from non-ST firms, although the differences are not extremely large. The differences between the medians are very large in Poland for each ratio, justifying the high AUC, and in Finland and the Czech Republic, where the difference in EBITTA is average. In Germany, Latvia, China (CN and delisted), and the U.K. (liquidation), the differences in all four ratios are below average, which is obviously associated with a low AUC. In the sample of Chinese delisted firms, the differences in RETA and EBITTA are even negative. In Iceland and Ireland, the differences only in EBITTA are exceptionally small.

Model 2 in Table 4 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 that for the original model (0.743), supporting *H1* only very weakly, if at all. The classification accuracy in terms of AR (0.490) is at approximately the average level. The re-estimation of the coefficients has led to improved classification accuracy in a number of countries, especially Bosnia, China (ST), Norway, and Greece. However, it has impaired the classification accuracy in the United Kingdom and China (delisted). The "Benchmark" column reports the results for the benchmark model (*Z*"-Score LR model), showing the effect of the estimation method. For the benchmark model, the AUC in all data is 0.748, which is higher than that for Models 1 and 2. The differences among AUCs are very small, only weakly supporting *H2* (estimation method). The LR model (benchmark) and the MDA model (Model 2) give nearly identical AUCs for each country. This result was expected because the coefficients of the models are directly comparable. The similar results for the models may also indicate that the independent variables conform to multinormality. Nevertheless, these similarities support what most researchers in the field of default classification models have concluded: that the accuracy levels of MDA and logistic regression models are extremely similar.

Model 3 (LR model with year dummies) leads in all test data to a higher AUC (0.752) than the benchmark model, supporting *H3* (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 the 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 per cent, while for the countries with positive effects, it is only a few 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 of only a few failed firms in the latter countries, which leads to the observed effects. *H3* is supported by evidence at the level of the whole sample. However, in some country samples (at the individual country level), *H3* is not supported.

Model 4 (LR model with size variables) performs better than the benchmark model, which lends support to *H4* (size effect) and leads to AUC = 0.760, indicating AR = 0.520, and to significant improvements in the AUC for, e.g., China (delisted) and Austria. For China (delisted), the increase in AUC is extremely strong, and the AUC also improves for, e.g., Estonia, Italy, Slovakia, Spain, and the U.K. Model 4 also led to lower AUCs in a few countries, but this 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 when approaching its maximum value in middle-sized firms. The countries with a positive change in AUC typically have data in which the percent of failed micro firms (in all failed firms) is relatively low (40–60 per cent), while there are many middle-sized failed firms. Because the size effect is strongest for middle-sized firms, the AUC increases. In contrast, the countries with a negative change in AUC have data in which the percentage of failed micro firms is exceptionally high (70–80 per cent).

Model 5 (LR model with age category dummies) gives, for all test data, nearly the same AUC as the benchmark model (AUC = 0.748). However, the difference is positive and statistically significant, giving at least marginal support to *H5* (age effect). For nearly all countries, the effect of age on AUC is small. For Austria, however, this effect is positive and significant. For Austria, the percentage of young firms (less than 6 years) among non-failed firms is only approximately 10 per cent, whereas this percentage among failed firms is more than 20 per cent. Because Model 5 includes a strong positive age risk effect for young firms (D_2), it increases the risk of many failed firms but of only a few non-failed firms, which obviously leads to an improvement in AUC.

Model 6 (LR model with industry dummies) outperforms the benchmark model in AUC, supporting *H6* (industry effect). It gives an AUC = 0.751, indicating an AR = 0.502. However, its AUC is notably higher than the benchmark AUC in only a few countries, such as France, Latvia, Portugal, and Spain. Model 6 has a negative effect on failure risk for, e.g., restaurants, hotels and the information technology industry, but it has a positive effect on the construction and manufacturing industries. For countries with a positive effect on AUC, the percent of non-failed firms in restaurants, hotels and information technology is high, while that of failed firms is low. For risky industries (construction and manufacturing), these distributions are reversed. Thus, Model 6 gives a positive (negative) risk effect for many failed (non-failed) firms and a negative (positive) risk effect for only a few failed (non-failed) firms. Therefore, the AUC increases. This situation is reversed for the countries with a negative effect on AUC (Austria and Slovenia). The samples of these countries include a high percentage of non-failed firms in the manufacturing industry, leading to a decrease in AUC.

Model 7 (LR model with the 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 *H7* (country of origin effect). However, for each country, the resultant AUC is nearly identical to that given by the benchmark model. This result was expected due to the negligible coefficient (−0.003) of the country risk measure (SP country rating rank) in Model 7. This result implies that country risk has no 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 and leads to a sizeable increase in AUC (AUC = 0.771) compared to the benchmark AUC in all test data. However, the effect on the AUCs largely varies and is either negative or positive in different countries. The positive effect is large in several countries, such as Estonia, France, Iceland, Italy, Latvia, and China (delisted). However, it also has a negative effect on AUCs in several countries, such as Bosnia, Hungary, and Norway. These results show that the inclusion of additional variables in the original model will usually increase the AUC, but not in every country.

The "all data" benchmark also performs fairly well for the United States and Colombian samples (with the U.S.A. firms, unlike the majority of other firms in this study, being listed or delisted companies). The poor performance of the predominantly private (CN) and delisted (DL) Chinese firm samples is associated with very small differences between the medians of the non-failed and failed groups, as shown in Table 5. It is clear that the "delisted" status is not comparable with "bankruptcy" status. 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 (Wang and Campbell, 2010; Zhang et al., 2010). 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" make it difficult for a uniform all data model to increase AUCs across all countries. Table 6 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 country data, this benchmark is clearly outperformed by the resulting MDA (Model 1) and LR (Model 2) models in only a few countries (Bulgaria, France, Latvia, Spain, and Sweden). However, these results give only weak support for *H1* at the country level because the effects are not significant. In addition, the benchmark leads to higher AUCs than in Models 1 and 2, at least in Austria, Bosnia, Ireland, Slovenia, and the United States. The dif-

Table 6. *Test Data AUCs for Different Estimated Country Models. Comparisons are With the 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, 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

Table 6 *(Continued)*

Country	*Test data AUC for different models*							
	Benchmark	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*
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.

ferences in AUCs given by Models 1 and 2 are generally small. Model 1 is clearly outperformed by Model 2 in Romania only. Thus, country-level evidence does not support *H2* (estimation method).

Model 3 (LR model with year dummies) leads in most countries to a clearly higher AUC than the benchmark model. This evidence gives support to *H3* (bankruptcy year effect). Model 3 leads to a lower AUC than the benchmark model in Austria, Bosnia, Slovenia, and the United States. Model 4 (LR model with size variables) leads to improved performance in nearly every country, supporting *H4*. This improvement is significant, however, only in the United States, where the AUC increases to 0.816 in Model 4, in contrast to the benchmark model, for which it is only 0.710. This improvement is also notable in Bulgaria, France, Latvia, Spain, and the United Kingdom. Model 5 (LR model with age category dummies) gives higher AUCs than the benchmark model for several countries, but none of the improvements are significant. The positive effect is particularly strong in Bulgaria, France, and Iceland. Although there are also negative effects on AUC, this evidence only weakly supports *H5* because 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. However, the only significant effect is the negative effect found in Austria. In Bulgaria and Slovenia, the AUC notably decreases due to the industry dummies. However, there are notable positive effects on the AUC in, for example, France, Iceland, Latvia, and Sweden. Thus, the effect is not systematic and gives only weak support to *H6*. Model 7 (LR model with all variables) leads to a remarkable increase in AUC compared with the benchmark model in Colombia, Iceland, and the Unites States and to notable improvements in AUC in Belgium, Denmark, Estonia, France, Germany, Hungary, Latvia, Norway, Spain, and the United Kingdom. Negative effects on AUC are found only in Austria and Slovenia. Thus, classification performance of the *Z*"-Score model with "all test" data remarkably increases in most countries when different effects are taken into account by additional variables. This increase is found in most European countries and in Colombia and the United States. In China, the AUC for the ST sample is extremely high for the *Z*"-Score LR model, and it can be improved only slightly by additional variables.

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) using a very large international dataset. We test how the original version of the *Z*"-Score model performs in different countries and how re-estimations using another statistical method and different additional variables affect the classification performance when the data are very heterogeneous. For this kind of testing, seven research hypotheses on classification performance are formulated. These hypotheses are tested for all data and separately for country data (country-level analysis). The estimation data are from 31 countries, and the results are validated for 34 countries. The countries are mainly from Europe, but three non-European countries are included (China, Colombia, and the US). The statuses used in the classification are mainly bankruptcy/active, but receivership firms are also considered to have failed. In the Chinese data, ST (special treatment) and delisted firms are also separately analyzed as failed firms.

The analyses for 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 the classification performance, thus weakly supporting the obsolescence hypothesis (*H1*) or, put differently, showing that the original coefficients are extremely robust across countries and over time (opposite to, e.g., Grice and Ingram, 2001). This same conclusion holds for the re-estimation of the model using LRA because the performance results are very similar to those of MDA (*H2*). The use of additional variables in the model generally improves the classification accuracy of the original model, but the results for countries are dependent on the distribution of failed and non-failed 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, the performance is generally improved, but the improvement is not strong and the effects vary by country. Thus, the evidence gives weak support for the effects of all additional variables. For the effects of bankruptcy year (*H3*) and size

(*H4*), the effects are stronger, but the variations in the effects between countries are also stronger. The effects of age (*H5*), industry (*H6*), and country (*H7*) are marginal. When all additional variables are included in the same model, the performance generally significantly increases, but at the same time, the variations among countries become stronger.

In summary, our evidence 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 for non-European countries using the four original variables accompanied by 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 (a conclusion similar to Xu and Zhang, 2009). 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. 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 then study the phenomenon of interest. In such 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 dataset, work consistently well internationally and are easy to implement and interpret. Thus, this kind of accounting-based model can be used by all interested parties, especially internationally active banks or other financial institutions, not only for failure or distress prediction but also for other managerial purposes such as provisioning and economic capital calculation. Internationally active banks need to develop a universal tool that can be applied in all subsidiaries and branches to control risk across the whole banking group.

Further research should focus on other modifications and extensions than those presented in our paper, such as using alternative modeling techniques (e.g., panel data analysis), introducing new variables (e.g., macroeconomic data), and testing its usefulness with data from other countries (e.g., emerging markets).

Notes

1. See Altman and Saunders (1997) for a review of research over this 20 year period.
2. Dichev (1998) compares the Altman Z-score and Ohlson O-score approaches.
3. Grice and Dugan (2003) present a re-estimation of Ohlson's and Zmijewski's models.
4. We use bankruptcy, failure, default and financial distress as equivalents.
5. Research devoted to the application of the Z-Score model before 2000 was reviewed by Grice and Ingram (2001).
6. Most of the models focused on stock-exchange-listed firms; thus, the Z'-Score and Z''-Score models were not used.
7. See Zhang et al. (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.
8. This is done because the results about predictability were also good for such a small sample.
9. These firms are included only in the test data because the predictability of failure was exceptionally poor.
10. From the weighting procedure, it follows that the score (cut-off-value) that best separates failures from non-failures is 0.50 (or, alternatively, 50 per cent). Although the score (logit) in principle has a probability interpretation, the "probabilities" estimated using this weighting scheme in this study do not, however, represent empirical PD's. It would still require calibration procedures for the models to obtain PD's that correspond to associated empirical PD's in the population. But this is not attempted in the study, as our focus is more general (the classification accuracies of the models across countries). It is also worth noting that the original Z''-Score does not have a PD interpretation either.

References

Acosta-González, E. and F. Fernández-Rodríguez, "Forecasting Financial Failure of Firms via Genetic Algorithms," *Computational Economics* 43 (2014), pp. 133–157.

Agarwal, V. and R. Taffler, "Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models," *Journal of Banking and Finance* 32 (2008), pp. 1541–1551.

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

Altman, E.I., *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing With Bankruptcy* (Hoboken: Wiley Interscience, John Wiley and Sons, 1983).

Altman, E.I., "Revisiting the Recidivism – Chapter 22 Phenomenon in the U.S. Bankruptcy System," *The Brooklyn Journal of Corporate, Financial & Commercial Law* 8 (2014), p. 253.

Altman, E.I. and E. Hotchkiss, *Corporate Financial Distress & Bankruptcy*, 3rd edn (Hoboken, NJ: John Wiley, 2006).

Altman, E.I. and A. Saunders, "Credit Risk Measurement: Developments Over the Last 20 Years," *Journal of Banking and Finance* 21 (1997), pp. 1721–1742.

Balcaen, S. and H. Ooghe, "35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems," *The British Accounting Review* 38 (2006), pp. 63–93.

Bauer, J. and V. Agarwal, "Are Hazard Models Superior to Traditional Bankruptcy Prediction Approaches? A Comprehensive Test," *Journal of Banking and Finance* 40 (2014), pp. 432–442.

Beaver, W.H., "Financial Ratios as Predictors of Failure," *Journal of Accounting Research* 4 (1966), pp. 71–111.

Beaver, W.H., "Alternative Accounting Measures as Predictors of Failure," *The Accounting Review* 43 (1968), pp. 113–122.

Bharath, S.T. and T. Shumway, "Forecasting Default With the Merton Distance to Default Model," *Review of Financial Studies* 21 (2008), pp. 1339–1369.

Campbell, J.Y., Hilscher, J. and J. Szilagyi, "In Search of Distress Risk," *Journal of Finance* 63 (2008), pp. 2899–2939.

Chava, S. and R.A. Jarrow, "Bankruptcy Prediction With Industry Effects," *Review of Finance* 8 (2004), pp. 537–569.

Das, S.R., P. Hanouna and A. Sarin. "Accounting-Based Versus Market-Based Cross-Sectional Models of CDS Spreads," *Journal of Banking and Finance* 33 (2009), pp. 719–730.

Dichev, I.D., "Is the Risk of Bankruptcy a Systematic Risk?," *Journal of Finance* 53 (1998), pp. 1131–1147.

Dimitras, A.I., Zanakis, S.H. and C. Zopoundis. "A Survey of Business Failures With an Emphasis on Prediction Methods and Industrial Applications," *European Journal of Operational Research* 90 (1996), pp. 487–513.

Grice Jr, J.S. and M.T. Dugan, "Re-Estimations of the Zmijewski and Ohlson Bankruptcy Prediction Models," *Advances in Accounting* 20 (2003), pp. 77–93.

Grice, J.S. and R.W. Ingram, "Tests of the Generalizability of Altman's Bankruptcy Prediction Model," *Journal of Business Research* 54 (2001), pp. 53–61.

Hillegeist, S.A., Keating, E.K., D.P. Cram and K.G. Lundstedt. "Assessing the Probability of Bankruptcy," *Review of Accounting Studies* 9 (2004), pp. 5–34.

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

Jackson, R.H.G. and A. Wood, "The Performance of Insolvency Prediction and Credit Risk Models in the UK: A Comparative Study," *British Accounting Review* 45 (2013), pp. 183–202.

Kumar, P. and V. Ravi, "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review," *European Journal of Operational Research* 180 (2007), pp. 1–28.

Kwak, W., Shi, Y., J.J. Cheh and H. Lee. "Multiple Criteria Linear Programming Data Mining Approach: An Application for Bankruptcy Prediction," *Data Mining and Knowledge Management* 3327 (2005), pp. 164–173.

Lyandres, E. and A. Zhdanov, "Investment Opportunities and Bankruptcy Prediction," *Journal of Financial Markets* 16 (2013), pp. 439–476.

Merkevicius, E., Garšva, G. and S. Girdzijauskas "A Hybrid SOM-Altman Model for Bankruptcy Prediction," *International Conference on Computational Science* 3994 (2006), pp. 364–371.

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

Ooghe, H. and S. Balcaen, "Are Failure Prediction Models Widely Usable? An Empirical Study Using a Belgian Dataset," *Multinational Finance Journal* 11 (2007), pp. 33–76.

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

Pindado, J., Rodrigues, L. and C. de la Torre. "Estimating Financial Distress Likelihood," *Journal of Business Research* 61 (2008), pp. 995–1003.

Reisz, A.S. and C. Perlich, "A Market-Based Framework for Bankruptcy Prediction," *Journal of Financial Stability* 3 (2007), pp. 85–131.

Ribeiro, S.P., Menghinello, S. and K.D. Backer. 'The OECD ORBIS Database: Responding to the Need for Firm-Level Micro-Data in the OECD'. No. 2010/1 (Paris: OECD Publishing, 2010).

Shumway, T., "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *Journal of Business* 74 (2001), pp. 101–124.

Smith, M. and D. Liou, "Industrial Sector and Financial Distress," *Managerial Auditing Journal* 22 (2007), pp. 376–391.

Taffler, R.J., "Empirical Models for the Monitoring of UK Corporations," *Journal of Banking and Finance* 8 (1984), pp. 199–227.

Tinoco, M. and N. Wilson, "Financial Distress and Bankruptcy Prediction Among Listed Companies Using Accounting, Market and Macroeconomic Variables," *International Review of Financial Analysis* 30 (2013), pp. 394–419.

Vassalou, M. and Y. Xing, "Default Risk in Equity Returns," *Journal of Finance* 59 (2004), pp. 831–868.

Wang, Y. and M. Campbell, "Business Failure Prediction for Publicly Listed Companies in China," *Journal of Business and Management* 16 (2010), pp. 75–88.

Wu, Y., Gaunt, C. and S. Gray. "A Comparison of Alternative Bankruptcy Prediction Models," *Journal of Contemporary Accounting and Economics* 6 (2010), pp. 34–45.

Xu, M. and C. Zhang, "Bankruptcy Prediction: The Case of Japanese Listed Companies," *Review of Accounting Studies* 14 (2009), pp. 534–558.

Zhang, L., Altman, E.I. and J. Yen. "Corporate Financial Distress Diagnosis Model and Application in Credit Rating for Listing Firms in China," *Frontiers of Computer Science in China* 4 (2010), pp. 220–236.

Zmijewski, M.E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research* 22 (1984), pp. 59–82.

Appendix. 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, DL	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

Appendix *(Continued)*

Country	WCTA Non-failed	WCTA Failed	RETA Non-failed	RETA Failed	EBITTA Non-failed	EBITTA Failed	BVETD Non-failed	BVETD Failed
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
U.K., liquidation dataset	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[a]	0.153	−0.092	0.213	−0.050	0.048	−0.025	0.555	0.075

WCTA, Working Capital/Total Assets; RETA, Retained Earnings/Total Assets; EBITTA, EBIT/Total assets; BVETD, Book Value of Equity/Total Liabilities.
[a]Because the two datasets for non-failed U.K. firms are identical, the non-failed medians are used only once in the calculation of the averages.