MASTER'S PROGRAMME IN FINANCE

THESIS

Verifying the Validity of Altman's Z Score as a Predictor of Company Failure: the Italian Case

Supervisors:

Martin Alexy

Patrizio Tirelli

Student:

Cojocariu Eduard Sebastian

INDEX

ABSTRACT

This thesis investigates the efficacy of Altman’s Z-Score in predicting company failure within the context of Italian Small and Medium-sized Enterprises (SMEs). Recognizing the pivotal role SMEs play in the Italian economy, the study assesses the traditional financial model's accuracy and explores the enhancement brought by incorporating non-financial variables. The data of the thesis consists of financial statements of Italian SME manufacturing companies. All the companies are unlisted. Data was collected from the Bureau van Dijk's Orbis database. The results demonstrate that the inclusion of non-financial variables significantly improves the model's discriminatory power and classification accuracy, reducing the Type I error rate and thereby providing a more reliable tool for stakeholders. This study contributes to the literature by validating the continued relevance of Altman's Z-Score model in contemporary settings and underscores the importance of a comprehensive approach to bankruptcy prediction.

1. Introduction

Increasing globalization is significantly increasing the competition between companies. Demand and supply may not always align, resulting in some companies losing customers. Only a small fraction of businesses remain in the market when the weakest ones go under. Business failures are a natural occurrence in our economic system, with firms entering and exiting based on overall business activity and expectations (Altman & Loris, 1976, p.1). Corporate failure is the eventual outcome due to systematic and non-systematic factors. Financial and accounting literature has repeatedly reinforced the belief in ratio analysis as an effective predictor of corporate failure. A model that can predict future bankruptcy as early as possible is very useful for different stakeholders. Specifically, a model that can produce reliable analysis, for large numbers of companies, and quickly and cheaply is certainly the desired tool for various stakeholders. The first multivariate bankruptcy prediction model was developed by E.I. Altman (1968) from New York University in the late 1960's . After this pioneering work, the multivariate approach to failure prediction spread worldwide among researchers in finance, banking, and credit risk. Failure prediction models are important tools for bankers, investors, asset managers, rating agencies, and even for the distressed firms themselves. The approach used for bankruptcy prediction has been evolving over time. Beaver (1966, 1968) used univariate analysis for selected ratios and detected that some of them had a very good predictive power. Altman (1968) moved significantly forward since he developed a multiple discriminant analysis model (MDA) called the Z-Score Model with 5 ratios. The next two decades brought even more financial distress research (e.g. Ohlson (1980), who used the logit model, Taffler (1982), who developed a Z-score model for the UK) which was summarized by Zmijewski (1984), who used a probit approach in his own model.

In spite of the vast research on failure prediction, the original Z-Score Model introduced by Altman (1968) has been the dominant model applied all over the world. Thus, although the Z-Score Model has been in existence for more than 45 years, it is still used as a main or supporting tool for bankruptcy or financial distress prediction or analysis, both in research and practice.

The reason of its success is in the fact that the Z-score model is easy to understand and can be used by anyone, even in the absence of adequate knowledge and skills in the field of business insolvency risk analysis. As we will see further on, the model uses easily obtainable data, from both balance sheets and statements as well as mean market share values relative to the reference period. It has been ascertained that the Z-score model can predict the distress of industrial companies listed on the Italian Stock Exchange (Borsa di Milano) Celli (2015). The success and accuracy rates shown by the model when applied to the Italian context are indeed fairly high, and the corresponding error rates quite low.

In our case we will focus particularly in Small and Medium Size enterprises (SMEs) since are reasonably considered the backbone of the economy of many countries and most of the operating and bankrupt companies are in this category.

Moreover, for OECD members, the percentage of SMEs out of the total number of firms is greater than 97 per cent.

After the Basel II regulatory framework banks have realized that small - and medium - sized companies are a distinct kind of clients with peculiarities that require specific risk management tools and methodologies (eg, Berger and Udell, 2002). While there have been many successful models developed for corporate distress prediction purposes, and at least two are commonly used by practitioners on a regular basis, none was developed specifically for SMEs .

Recently, Altman and Sabato (2007) apply, with some success, a distress prediction model estimated specifically for the US SME sector based on a set of financial ratios derived from accounting data. They demonstrate that banks should not only apply different procedures (in the application and behavioral process) to manage SMEs compared to large corporate firms, but these organizations should also use scoring and rating systems specifically addressed to the SME portfolio.

Our main objective is to verify if the Z-score model is a true indicator of financial failure for Italian SMEs even after the recent crisis or if more in-depth studies have to be made. This need has been made all the more crucial by the global effects of the COVID-19 pandemic, which, while having a great impact on companies of all sizes, proved to affect SMEs to a larger extent because of their physiologic financial weakness (Ciampi 2015), as well as their prevalence in the industries and countries more exposed to the effects of the pandemic.

Our study is focused on the Z'-score version for private manufacturing firms in order to check its vitality. For the analysis, we use a sample of 89.417–Italian SMEs so distributed: 88.149 viable firms and 1.268 non-viable firms. We collected annual firm financial data from the ORBIS (Bureau Van Dyke) database over the period 2021-2022.

2. Literature Review

There has been extensive research in forecasting bankruptcy, resulting in a wide array of outcomes. Newer, improved, and more precise methods are continually being developed to predict bankruptcy. However, it appears that most methods rely on financial ratios such as inefficiency, high leverage, and poor liquidity.

Although numerous studies have created prediction models for bankruptcy using a variety of statistical techniques, a significant portion of the research has used US data to extend Beaver’s (1966) univariate methodology and Altman’s (1968) multiple discriminant analysis model. Beaver (1966) was the first to introduce a novel approach to bankruptcy prediction by developing a univariate model focused on evaluating the ability of individual key figures to forecast bankruptcies. This seminal study included 158 companies divided into two groups: bankrupt and non-bankrupt. Beaver identified significant differences between the financial ratios of these two groups. Beaver’s research was a pivotal academic advancement because it showed that a warning about an impending crisis could be detected at an early stage. Moreover, the study's methodology can be easily applied in practical business settings. However, a limitation of the forecasting model based on individual ratios is the ambiguity of the results, as different indicators can give varying forecasts for the same company. In his study, Beaver suggested that a multi-ratio model would be more accurate for predicting bankruptcy than just using individual financial ratios. The transition from single-variable analysis to multivariable analysis was a significant impetus for further research. Edward Altman was the pioneering researcher who advanced this model.

He criticizes the ability of one variable to describe a company’s financial situation because it does not consider all the company’s operating conditions. For example, if a company has poor profitability, it is considered a potential bankruptcy company, but in case company has good liquidity ratios the situation might be considered different. Thus, Altman's vision was to build a single predictable model with many financial ratios included.

He advanced upon Beaver's work by incorporating four more variables into the model to give an overall more precise prediction of manufacturing corporate failure. Altman's multi-discriminant analysis (MDA) model differed to Beaver's model in relation to the ratios chosen of highest prediction. Altman classifies the companies into two mutually exclusive groups; bankrupt and non-bankrupt (Altman, 1968, p.591). Altman's discriminate analysis became a dominant model used in corporate failure prediction literature due to its simplicity and accuracy. His multi-discriminant approach allowed him to develop the equation into a combination of five ratios consisting of liquidity, profitability, financial leverage, solvency, and sales activity (sales to total assets). This linear equation distinguished between failing and non-failing companies. The result of the combination of ratios gives rise to a discriminant score otherwise known as the 'Z score'. The data for the study is collected from the balance sheets and income statements. Previous studies have shown that many financial variables are significant in predicting the financial problems of companies, thus Altman formed a list of 22 potentially helpful ratios for evaluation. He divided these ratios in to five categories; liquidity, profitability, leverage, solvency and activity ratios. The criteria for selecting these 22 ratios were the popularity in the literature and the potential relevancy for research. From the original list of 22 variables he finally selected five different variables. Altman did not choose the most significant variables measured independently because the selection was based on the ability of the variables to do best overall job together and form the best bankruptcy prediction model. (Altman 1968).

2.1 Z-Score Models

The purpose of the Altman's model is to find out the ideal combination of different financial variables that best predict bankruptcies and then combine these together into a single weighted index which defines the company's probability of default. Altman named this index as Z-score. The original formula developed by Altman below:

Altman's original Z-Score:

$$Z = 1.2(X1) + 1.4(X2) + 3.3(X3) + 0.6(X4) + 1.0(X5)$$

where

X1 = working capital/total assets,

X2 = retained earnings/total assets,

X3 = earnings before interest and taxes/total assets,

X4 = market value equity/book value of total liabilities,

X5 = sales/total assets

Boundary values:

$Z > 2.99$ Safe Zone: Considered financially healthy

$1.81 < Z < 2.99$ Grey Zone: Could go either way

$Z < 1.81$ Distress Zone: Risk that company will go bankrupt within two years

The Z-Score is calculated by multiplying each of the financial ratios by an appropriate coefficient and then adding the results together. The lower the score, the greater is the risk of financial distress, as a company with a Z-score of -2 is in worse condition than one with a score of 1. The coefficients describe the importance of each ratio, since larger coefficients affect the Z-score more. Each of the ratios is discussed below (Altman 2000):

X1: Working Capital/Total Assets

The working capital to total assets ratio serves as an indicator of a firm's net liquid assets in relation to its total capitalization. Working capital is calculated as the difference between a company's current assets and current liabilities, while total assets encompass both current and fixed assets. This ratio effectively incorporates liquidity and size characteristics, providing a comprehensive measure of the firm's financial health. A high working capital to total assets ratio typically suggests that a company has a substantial amount of liquid assets, which can be used to meet short-term obligations. Conversely, a low ratio may indicate potential liquidity problems, as the company may struggle to cover its short-term liabilities with its current assets. This is particularly important for assessing a company's ability to sustain operations and avoid financial distress. In practical terms, a company experiencing consistent operating losses will often see a decline in its current assets relative to its total assets. This is because ongoing losses can erode the firm's liquid resources, reducing its working capital. As a result, the working capital to total assets ratio will decrease, signaling potential financial instability.

X2: Retained Earnings/Total Assets

Retained earnings report the accumulated reinvested earnings and/or losses of a firm. It is found in the Stockholders Equity section of the Balance Sheet. The ratio measures the cumulative long-term profitability of the company and implicitly considers the age of a firm. Studies have shown that corporate failures are much more common in a firm’s earlier years, as many firms that go bankrupt are relatively young ones that have not yet had the time to build up its cumulative earnings. Hence, it makes sense that young companies are more likely to default on their obligations. In addition, X2 measures the leverage of a firm. Companies with high retained earnings relative to total assets have to a greater extent financed their assets through retention of earnings rather than debt financing, which may reduce the likelihood of bankruptcy.

X3: Earnings Before Interest and Taxes/Total Assets

This ratio demonstrates the efficiency of the company's assets prior to accounting for tax or leverage influences. Companies rely on the effective utilization of their assets' earning potential to ensure their long-term sustainability. The return on total assets seems especially relevant for forecasting bankruptcies, given its significant weighting in each of the Z-Score models. EBIT is located in the company's Income Statement.

X4: Market Value of Equity/Book Value of Total Liabilities

The market value of equity represents the aggregate market value of all shares of common and preferred stock. Total liabilities encompass all current and long-term liabilities listed on the firm’s Balance Sheet. X4 measures how much the company’s assets can depreciate before liabilities surpass assets, leading to insolvency. The equity-to-debt ratio also highlights the firm's leverage. A higher debt level relative to equity signifies a higher risk for the firm. Additionally, this ratio introduces a market perspective to the Z-Score, indicating that declining stock prices might signal impending issues. This ensures that systematic risk is factored into the model, which is crucial during financial crises.

Moreover, the market value of equity partially reflects a company's credit risk and bankruptcy risk. When stock market sentiment is positive and stock prices are high, companies find it easier to borrow money and raise capital through equity issues. Therefore, the market value's effect on X4 partly captures the funding accessibility of companies in the Z-Score model, implying that a low ratio could indicate potential difficulties in securing financing. This aspect is also significant in the context of financial crises.

X5: Sales/Total Assets

The asset-turnover ratio is a common financial metric that assesses the efficiency of a company's assets in generating sales and management's effectiveness in handling competitive pressures. Sales are listed as revenues in the company's Income Statement. It is advisable to use net sales, which account for deductions like returns, allowances, and discounts. Altman determined that X5 is the least critical on its own, yet it holds considerable importance due to its distinct connection with the other ratios in the Z-Score for manufacturing firms.

To summarize, Z-Score combines liquidity, profitability, solvency and efficiency ratios to draw a conclusion about the overall score. The higher the value of the z-score, the less likely it is that the company will face bankruptcy.
Altman established a critical threshold for the results. Companies with a score above 2.99 were deemed functional and in the "safe zone," while those with a score below 1.81 were classified as bankrupt. Thus, if the score exceeds 2.99, the model accurately identifies all companies as healthy, and if the score is 1.81 or lower, it correctly categorizes the company as bankrupt. The range between these

two critical values is known as the "gray area." In this range, the model cannot fully predict outcomes, leading to classification errors. After testing the discriminant analysis model, Altman concluded that the model is a reliable predictor of failure, accurately classifying 95 percent of companies into their actual groups. Additionally, the function proved accurate in several other samples where reliability was tested. However, the predictive power varies with different time horizons. The model accurately forecasts two years prior to bankruptcy, with accuracy diminishing significantly as the time horizon extends.

In 1983, Altman further refined his research and tailored the model for small, privately held companies. He recalculated the original coefficients for the variables, replacing the market value of equity with the book value. Since the original Z-model relied on market value, it was unsuitable for non-public companies. Altman therefore substituted the market value with the book value and re-estimated the coefficients of the ratios. This revised model is known as Z'. The Z'-model is utilized to assess the bankruptcy potential of manufacturing firms. The model and its variables are presented below:

$$\mathbf{Z' = 0.717X1 + 0.847X2 + 3.107X3 + 0.420X4 + 0.998X5}$$

Where:

X1: Working Capital/Total Assets

X2: Retained Earnings/Total Assets

X3: EBIT/Total Assets

X4: Book Value Equity/Total liabilities

X5: Sales/Total Assets

Source: Altman (1983:122)

This model was further developed to create the Z” Score model (Altman, 1995). This was adapted to predict corporate failures for developing countries firms (Mexican companies), emerging market companies and for non-manufacturers. This model kept the first four variables, the asset turnover ratio (Sales / Total Assets), X_5 variable, has been completely removed from the model since it is industry specific variable. Altman did this in order to minimize the sensitivity of the industry effect, which makes the model useful for a wider range of non-manufacturing companies (Altman 2000).

Z-score estimated for non-manufacturers below:

The modified Z"-Score model:

$$Z'' = 3.25 + 6.56 (X1) + 3.26 (X2) + 6.72 (X3) + 1.05 (X4)$$

Boundary values:

$Z > 5.85$ Safe Zone

$4.35 < Z < 5.85$ Grey Zone

$Z < 4.35$ Distress Zone

2.2 SME Studies

There has been a lot of research in predicting bankruptcy and there is a considerable amount of different results. New, better and more accurate methods are constantly being developed to predict bankruptcy. However, in spite of the vast research on failure prediction, still few research focused on the small and medium - size enterprises.

There is no common definition of the segment of small and medium sized enterprises across different countries. The European Union has had a common definition since 1996 that was updated in 2003 and it's still valid today. The number of employees and the annual turnover of a firm are the criteria considered (less than €50 million in sales or less than 250 employees).

Typically, the distinctive characteristics of small and medium-sized enterprises (SMEs) require the development of default prediction models customized to address SME-specific challenges. These models rely not only on conventional financial ratios but also on qualitative data, as highlighted in works by Ciampi (2015, 2018) and Norden and Weber (2010). Over the past twelve years, research in SME default prediction has grown into a significant field, exploring various emerging issues and methodologies, as noted by Ciampi (2015).

After the Basel Accord for bank capital adequacy (Basel II) some analysts started focusing on the SME segment (see e.g., Saurina and Trucharte, 2004; Altman and Sabato, 2007; Berger, 2006). As a result, a significant number of studies aim to analyze and predict the bankruptcy risk of SMEs in different geographical contexts.

Another aspect of the literature suggests that despite being the driving force behind economic progress, small and medium-sized enterprises (SMEs) face greater financial limitations compared to larger corporations. Access to finance emerges as a critical impediment to their growth trajectory.

Focusing on the US, Altman and Sabato (2007) use a logit regression technique to develop a 1-year default prediction model for a panel of 2000 firms over the period 1994 to 2002. They demonstrate that a convenient way for banks to set their credit risk system is to separate SMEs from generic firms.

These studies have dealt with the problem of the possible effects of Basel II on bank capital requirements, but the problem of modelling credit risk specifically for SMEs has either not been addressed or only briefly considered (Altman and Sabato, 2008). Later on, there has been a growing focus in finance and management literature on enhancing forecasts for small and medium-sized enterprise (SME) defaults. Scholars have noted the ongoing absence of a widely recognized model for predicting SME defaults, as highlighted by Ciampi et al. (2021).

Dietsch and Petey (2004) develop a one-factor credit risk model to assess estimates of stationary default probabilities and asset correlation for class of small medium firm, in order to assess the effects of new regulation of Basel II. They show that, on average, SMEs are riskier than large businesses; asset correlations in the SME population are very weak (1-3%) and decrease with size. Moreover, they do not find evidence supporting a negative relationship between asset correlations and prediction default across rating grades as assumed by Basel II.

Tabouratzi, Lemonakis and Garefalakis (2017) run a panel regression model with correction for fixed effects on a sample of 3600 Greek manufacturing firms from 2003-2011. They find that firms presenting higher performance in terms of ROA and sales and higher leverage levels that enhance their liquidity as well are healthier in terms of Z-score than their less profitable counterparts and acquire lower rates of probability of default.

A more recent study has been made by (Altman, Balzano, Giannozzi and Srhoj, 2023) in which they use LASSO techniques and other machine-learning techniques on a sample of 2040 SMEs from 2015-2019 in the context of Croatia. They find argue that introducing management and employee-related variables into SME prediction models can improve their predictive power.

Focusing on Italian firms, Pederzoli and Torricelli (2010) adopt a logit model to predict the default probability for a specific region, that is, Emilia Romagna, based on financial ratios. They find that the equity ratio, the EBIT over asset ratio, the long-term liabilities over asset ratio and the sales over asset ratio are sufficient to fit the default event in their sample.

In a similar context, for a sample of 232 Italian SMEs, Dainelli et al. (2013) develop a logit model for a 1- year estimation of the probability of default. In addition to standard financial indicators (profitability, solvency and liquidity), they include credit relationship quality indicators. They find that both profitability and credit relationship quality are important determinants of the probability of default.

Ciampi and Gordini (2013) provide a methodological contribution and focus on a large sample of 7000 small Italian enterprises (those with a turnover of less than 1.8 million euros). They compare results obtained with different techniques and find that neural network analysis makes a better contribution to the small enterprise credit risk evaluation than traditional techniques, that is, MDA and LR.

Luppi, Marzo, Scorcu (2007) use a multiple-factor credit risk model to provide new estimates of default probabilities in a sample of 3900 SMEs and show that show that, on average, SMEs are riskier than large businesses within the retail segment.

Gordini (2014) compares the potential of genetic algorithms (GAs) with those of logistic regression (LR) and support vector machine (SVM) to a sample of 3.100 Italian manufacturing SMEs, three, two and one year prior to bankruptcy. He finds that GAs are a very effective and promising instrument in assessing the likelihood of SMEs bankruptcy compared with LR and SVM and shows that GAs prediction accuracy rate increases when the model is applied according to size and geographical area.

Ciampi (2015) apply a logistic regression to a sample of 934 Italian small enterprises (SEs) from 2010 and found that CEO duality, owner concentration, and a reduced number of outside directors on the board (no more than 50%) are significantly and negatively correlated with small company default and that corporate governance variables significantly improve the SE default prediction accuracy rates.

Modina and Pietrovito (2014) use a logistic regression model on a database of 9208 Italian limited liabilities SMEs in a time frame of 3 years, over the period 2006 to 2010. They find that the capital structure (both in terms of internal and external funds and in terms of the source of external financing) and interest expenses are more relevant than economic variables as determinants of SMEs' default.

Pozzoli and Paolone (2016) apply the Z'-Score model on a sample of 335 Italian manufacturing companies (S.p.A. and S.r.l.) which went bankrupt within the first quarter of 2016. Their results confirm a good predictive effectiveness in relation to bankrupted companies with significant discrepancies between the different, analyzed juridical entities.

Calabrese, Marra and Osmetti (2016) introduced a binary regression accounting-based model for bankruptcy prediction of SMEs. They employed the BGEVA model, logistic additive regression as well as log-log additive regression on a sample of 49 738 Italian SMEs from the period 2006-2011.

Altman, Esentato and Sabato (2020) used a logistic regression technique in order to build a multivariate model for predicting the probability of default on a sample of 14,420 Italian SMEs over the period 2004–2013. Their results confirmed that the Z'-Score model successfully classified and predicted default or non-default on large samples of Italian SMEs.

The Italian SME sector, including the micro-enterprises, is the largest in the EU, with 3.730 million firms (ISTAT, 2021). Moreover, focusing on SME's, their number is around 211 thousand, which remarkably account for 41% of the total revenue generated in Italy, employ 33% of the private sector workforce, and contribute 38% to the country's value added, all on their own. For this reason, our analysis will focus on them.

2.3 Limitations

While this thesis aims to cover many aspects of firm bankruptcy analysis, there are some limitations to our work. The Z-Score models are widely recognized as reliable tools for predicting corporate insolvencies in various markets. However, these models are not universally applicable and have faced various criticisms. Altman tested the original Z-Score model on companies of all sizes, showing its ability to evaluate even large corporations. A common criticism is that financial ratios can reduce statistical significance by size, thus diminishing the impact of size-related factors (Altman, 2000). Additionally, these models use unadjusted accounting data, making them vulnerable to significant fluctuations due to one-time write-offs between quarters. They can also be affected by false accounting practices, as Altman noted, where retained earnings can be manipulated through corporate reorganizations and stock dividend declarations, introducing bias (Altman, 2000). Since the Z-Score replaces the market value of equity with the book value, it may miss bankruptcies caused by factors not reflected on the balance sheet, such as unexpected business disruptions. This makes the Z-Score particularly prone to potential manipulation of accounting data. Many researchers have pointed out that the methodology used in these models often violates two key assumptions of the MDA technique: the multivariate normal distribution of independent variables and the equivalence of variance-covariance matrices (Barnes, 1982; McLeay and Omar, 2000). Recognizing these issues, researchers often use Logistic Regression Analysis (LRA) as an alternative approach. LRA uses a parameter-nonlinear model to estimate the likelihood of bankruptcy. Unlike probit models, which assume a cumulative normal distribution, logit models assume a logistic distribution. A benefit of this technique is that it does not rely on the strict assumptions that MDA does.

For these reasons, we will use the logit model in our case. Despite these concerns, the Z-Score models are still among the best-known and widely used measures of financial distress. These credit risk models have proven to be important tools for analyzing corporate health and the possibility of bankruptcy. To strengthen and verify the results, the models can be supplemented with other analytical tools.

Moreover, using only two non-financial variables may have limited the impact on the model's accuracy. Academic literature, such as studies by Altman, Sabato, and Wilson (2008) and Altman et al. (2017), has pointed out that incorporating different non-financial attributes of firms significantly enhances the predictive capability of risk models. Non-financial variables such as management quality, corporate governance, and market position could provide additional insights into the risk profiles of SMEs.

Another limitation of our study is the limited time period analyzed, focusing only on the years after the COVID pandemic. This could result in a lack of generalizability of the findings, as the economic conditions during the pandemic and the immediate aftermath were atypical and might not reflect the broader economic trends. Partially to not broaden the study more and due to time limitations, some potential quality improvements are left out. Future studies could benefit from analyzing a more extended period, including both pre-pandemic and post-pandemic years, to provide a more comprehensive understanding of the factors influencing SME bankruptcies.

3. Objectives & Hypothesis

3.1 Goals

This thesis primarily aims to assess the predictive capability of the revised Altman Z-score model (1983) for privately owned firms in Italy. Exploring this predictive capability is intriguing as it can demonstrate the applicability of the Altman Z-score model within the Italian context. In essence, the research seeks to determine whether the model can serve as an effective forecasting or early warning tool for Italian companies in their decision-making processes. The 1983 version of Altman's Z - score model is an updated iteration of the original 1968 model, which employs multiple discriminant analysis to generate a linear combination of financial ratio variables that distinguish between bankrupt and non-bankrupt firms. In our study we will use the 1983 version which is called Z' - score model.

3.2 Hypothesis

The aim of our analysis is to assess the classification performance and the transferability of the re-estimated Altman's Z' - score model. To predict bankruptcies, financial ratios need to vary between an operational and functioning company and a bankrupt one. Hence, the initial hypothesis of this research is as follows: *H1: Financial ratios vary between solvent and insolvent companies.*

Even though Altman's Z-Score models has been used in Italy, also analyzing the SME's sector, we are interested in finding if Altman's model still has good degree of performance by classifying bankrupt and non-bankrupt companies in the Italian context after COVID-19. *Therefore, our second hypothesis (H2) is that there is a positive relationship between the Altman Z'-score model variables and predicting bankruptcy for Italian SMEs.*

Many studies such as Grunert, Norden, and Weber (2005), Altman, Sabato, and Wilson (2010), Altman et al. (2017), which we are going to look over in the following section, concluded that non-financial company-specific variables contribute significant information to failure prediction. *This is why our third hypothesis (H3) is that non-financial variables improve discriminatory performance and classification accuracy of the model.*

4. Data & Methodology

3.1 Dataset

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 modified Z'- Score model in classifying bankrupt and non-bankrupt firms in the Italian context mainly after the COVID-19 crisis.

Building an appropriate database for our empirical analysis, we deal with a number of issues. The first one is related to the definition of SMEs. To this end, we follow the Basel II rules and the definition provided by the European Union (Commission Recommendation 967280/EC), which takes into account both the number of employees and the amount of sales.

Table 1. Enterprise Size Classification Criteria

Enterprise category	Headcount: annual work unit (AWU)	Annual turnover		Annual balance sheet total
Medium-sized	<250	≤ EUR 50 million	or	≤ EUR 43 million
Small	<50	≤ EUR 10 million		≤ EUR 10 million
Micro	<10	≤ EUR 2 million		≤ EUR 2 million

As we can see in Table 1, firms with fewer than 250 employees and sales lower than 50 million euros are considered SMEs. These requisites have also been confirmed recently by Basel III. Secondly, default is intended as the end of the SME' s activity, that is the status in which the SME needs to liquidate its assets for the benefit of its creditors. In practice, we consider a default to have occurred when a specific SME enters a bankruptcy procedure as defined by the Italian law.

3.2 Sample

Our identification process starts from Bureau Van Dijk's database. The data are from ORBIS Europe that is a commercial database which at the moment of sampling contained administrative information on over 50 million European firms. Moreover, in our sample the number of employees is restricted from a minimum of 1 to a maximum of 250.

We require that the company to be selected must operate in manufacturing industries and had filed for bankruptcy in 2023, with a regular balance sheet from 2021 to 2022. The economic–financial variables were based on the 2022 and 2021 balance sheet and income statement data. This means that the data set has financial ratios one (T-1) and two (T-2) years prior to bankruptcy, occurred in 2023. The financial ratios of the bankrupt and non-bankrupt cases are always from the same calendar years (i.e. at T-1 from the 2022 balance sheet, at T-2 from the 2021 balance sheet). This will be our final validation sample on which we are going to analyze Altman’s model. Regarding the training sample on which we will do a comparison in order to validate our analysis, it consists of companies that filed for bankruptcy in 2022 with a regular balance sheet from 2021. We do not include the year 2020 for the training sample for reasons explained below.

We analyze the bankruptcy prediction only one year and two years prior to failure and do not include year 2020 in the analysis for two different reasons. Firstly, with the COVID-19 pandemic outbreak in 2020, many companies were forced to stop their normal operations and their balance sheet may mislead the analysis. Secondly, following the work of Altman (1968), he found that after the second year the model accuracy falls off. The most logical reason for this occurrence is that after the second year, the discriminant model becomes unreliable in its predictive ability.

ORBIS (2023) divides active firms into 6 subclasses (*active, active rescue plan, active default of* payment, active insolvency proceedings, active reorganization, and active dormant), and inactive companies in 8 subclasses (in liquidation, bankruptcy, dissolved merger or take-over, dissolved demerger, dissolved liquidation, dissolved bankruptcy, dissolved, inactive no precision). In selecting the active firms, we select only the 'active' ones following the literature as did by Altman (2017), while regarding the failed firms, we try to avoid ambiguity as much as possible by considering a firm failed if its status in ORBIS is stated as bankruptcy or liquidation. We remove companies for which we are not able to compute the Z'-score from the sample. After eliminating these data, our data set for the validation sample consists of 89.417 Italian SMEs. Among them, 1.268 were defaulted cases and 88.149 were not. Regarding the training sample, it contains 98.630 Italian SMEs of which 1.293 were defaulted and 97.337 were not.

Table 2. Breakdown of Sample by Year and Default

	Financial Year	Non-defaulted	Defaulted	Total
Training Sample	2021	97.337	1.293	98.630
Validation Sample	2022	88.149	1.268	89.417

This table shows the structure of the Italian SME's development sample. In the second and third row the number of non-defaulted and defaulted firms are shown.

4.1 Methods

Regarding methodology, investors frequently utilize matched sampling when applying the MDA technique (Altman, 2000; Beaver, 1966; Begley et al., 1996; Mselmi, Lahiani, & Hamza, 2017). This method involves aligning assets, employee counts, and sales figures of bankrupt firms with those of non-bankrupt counterparts. However, it's important to note that matched sampling may not always be the best choice due to the potential for selection bias. This bias occurs when the selection of non-bankrupt firms is not randomized, leading to a lack of representativeness within the sample compared to the broader population of non-bankrupt firms. To address this bias, researchers often resort to random sampling, ensuring that each firm in the population has an equal chance of being included in the sample (Marshall, 1996). However, random sampling introduces its own issue in the form of sampling error, where sample components deviate from the population parameters. This problem can be mitigated by significantly increasing the sample size (Marshall, 1996). Therefore, I choose to utilize all available non-bankrupt firms with comprehensive datasets, ensuring that all relevant variables are taken into account when computing ratios to represent the non-bankrupt group. As defined by the early work of Ohlson (1980) and Zavgren (1983), studies using MDA approach encounter several challenges. Firstly, they are subject to stringent statistical prerequisites concerning the distributional characteristics of the predictors. Notably, the variance-covariance matrices of the predictors must exhibit uniformity across both groups (i.e., failed and non-failed firms). Additionally, the reliance on normally distributed predictors poses a hindrance to the incorporation of dummy independent variables. Secondly, the outcome yielded by the MDA model, in the form of a score, lacks intuitive interpretation, primarily functioning as an ordinal ranking mechanism rather than offering substantive insight. Thirdly, concerns arise regarding the efficacy of the "matching" procedures commonly employed in MDA studies. The process of matching failed and non-failed firms based on criteria such as size and industry is inherently arbitrary, introducing potential biases into the analysis. For these reasons, the authors and the succeeding body of academic work as Altman and Sabato (2007), Altman, Sabato, and Wilson (2010), and Altman et al. (2017) employed a logistic regression model, known as logit.

This approach involves utilizing a set of diverse financial indicators to estimate probabilities of binary outcomes. Unlike methods such as MDA (Multivariate Discriminant Analysis) or LRA (Linear Regression Analysis), logit does not impose stringent assumptions on the data, rendering it a widely favored technique in the realm of binary outcome modeling. This approach doesn't hinge on multivariate normally distributed variables, nor does it depend on the assumption of equal variance-covariance matrices within the two classification groups. Consequently, we opt for logit regression as our preferred method due to its flexibility and widespread applicability. Considering the challenges associated with MDA over the past three decades, a significant portion of academic research has turned to logit analysis for default prediction (e.g., Altman & Sabato, 2007; Ciampi & Gordini, 2008, 2009, 2013b; Daily & Dalton, 1994a, 1994b; Keasey & Watson, 1987; Lee & Yeh, 2004). This body of work has demonstrated that, statistically speaking, logistic regression is well-suited to address the complexities of default prediction, particularly when the dependent variable is binary (default/non-default) and the groups are discrete, non-overlapping, and distinguishable. The logit model generates a score ranging from zero to one, providing a convenient probability of default for the client. Moreover, logistic regression analysis has become the predominant method for modeling company default prediction, widely adopted by banks and credit rating agencies. Hence, in alignment with the predominant literature on default prediction modeling this study also employs logistic regression analysis. In general, such a technique derives the probability of an event by calculating coefficients for the financial ratios. These coefficients are interpreted as the effect of a unit change in a financial ratio on the probability of failure. A function transforms the predicted values to a probability function Zavgren (1983). If the logistic transformation function is used, we have the logistic regression model:

$$P(x) = (\beta_0 + \bar{\beta}\bar{X}_i)\,\frac{e^{(\beta_0+\bar{\beta}_i\bar{X}_i)}}{1+e^{(\beta_0+\beta_i X_i)}} = \frac{1}{1+e^{-(\beta_0+\bar{\beta}_i\bar{X}_i)}}$$

Zavgren (1983).

Where:

P (x) = Probability of failure

β_0 = Regression intercept

$\vec{\beta}$ = Vector of slope parameters for each of the i = 1 to 5 independent variable

$\vec{X}$ = Independent variable vector of each of the five financial variables.

We use as dependent variable a dummy variable equal to 1 for defaulting firms and to 0 for non-defaulting firms. As independent variables we use the set of financial ratios of Altman.

Based on data from one financial statement prior to default, we assess the Type I (or false negative) and Type II (or false positive) error rates as we can see in Table 3.

Table 3. Overview of Accuracy Matrix

	Actual outcome for firm		
	Failed	Non-failed	Total
Predicted outcome:			
Failed	*TP*	*FP*	*TP+FP*
Non-failed	*FN*	*TN*	*FN+TN*
Total	*TP+FN*	*TN+FP*	*TP+FP+TN+FN*

TP (True Positive) is the correct classification of a failed firm; and *TN* (True Negative) is the correct classification of a non-failed firm. A false negative outcome (*FN*) is predicting a firm to survive when it actually fails (type I error); and a false positive outcome (*FP*) is predicting a firm to fail when it actually survives (type II error).

The Type I error measures the percentage of defaulted firms that are classified as non-default and the Type II error measures those firms classified as default but which did not default. The Type I error in binary classification is committed when failed company is marked as non-failed, i.e. it is false negative. The Type II error is false positive, i.e. non-failed company is marked as failed. In credit risk modeling the Type I error is more costly, since there are higher costs associated with lending a company that will bankrupt eventually than it is to let go of potentially profitable client just because the model marked him as bankrupted.

4.2 Default Prediction Methods

There exist various methodologies for assessing the discriminatory performance of binary classification models on out-of-sample data. Previous research predominantly relied on metrics such as the proportion of accurately and inaccurately predicted outcomes, or the overall correctness of predictions. However, evaluating classification accuracy necessitates the establishment of a particular threshold, making accuracy contingent upon this threshold. In our analysis we will use as cut-off point the proportion of defaulted companies in training sample.

Prior to that, we will evaluate the discriminatory performance of the models on out-of-sample data using the area under the ROC (receiver operating characteristic) curve, which circumvents the necessity for a predetermined threshold. This is done by following the academic work of Altman, Sabato and Wilson (2008) and Kacer, Ochotnicky and Alexy (2019).

The ROC curve plots the true positive against the false positive rate as the threshold to discriminate between failed and non-failed firms' changes. It illustrates the model's ability to discriminate between positive and negative outcomes. A fundamental aspect of the ROC curve is its portrayal of the model's performance spectrum, ranging from random guessing (represented by the diagonal line) to optimal classification (approaching the top-left corner), thereby offering a comprehensive visualization of the model's discriminatory power, as shown by Engelmann, Hayden and Tasche (2003).

Figure 1.

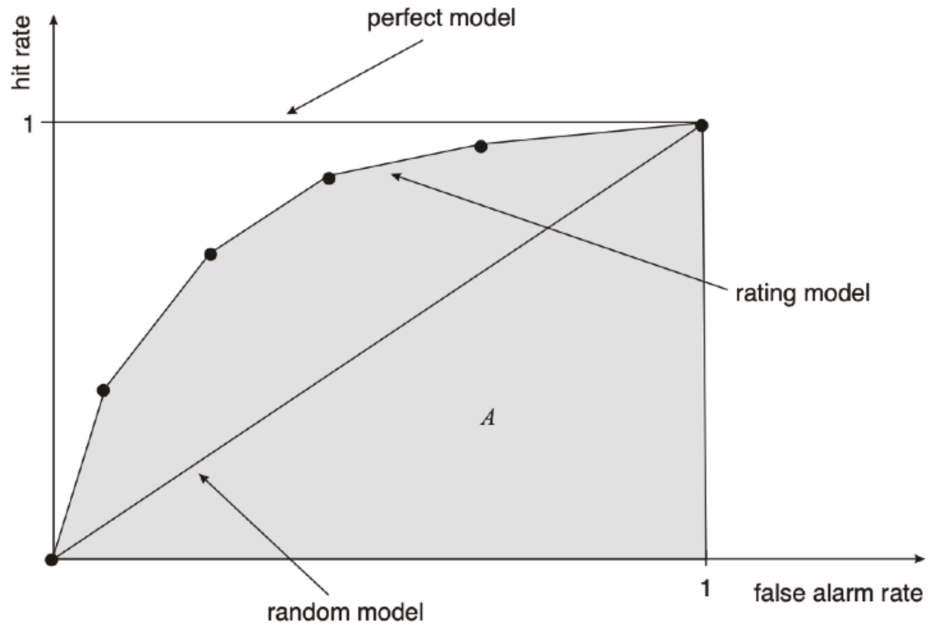

Figure 1 is taken from Engelmann, Hayden and Tasche (2003) for a better understanding of the ROC curve.

Complementing the ROC curve, the AUC metric encapsulates the area beneath the ROC curve, providing a singular quantitative assessment of model performance. This scalar value serves as a condensed representation of the model's ability to distinguish between different classes of financial risk. AUC values range from 0 to 1, where 0.5 indicates a model with no discrimination ability (random guessing) and 1 represents a perfect model that perfectly separates the two classes. Generally, an AUC value above 0.7 is considered acceptable for many financial applications. The accuracy ratio, AR, of each model, following Englemann, Hayden, and Tasche (2003), is a linear transformation of the area under the ROC curve: AR = 2(θ - 0.5).
The reason why we use the AUC is that it is commonly used as a performance metric because it is threshold-independent, meaning it considers the overall performance of the model across all possible threshold values.

4.3 Variables

In this research the functionality of the Altman model is examined with new, up to date data, and thus no attempt is made to find new financial indicators that might increase the accuracies of Altman's Z-models. The financial ratios of the bankrupt and non-bankrupt cases are always from the same calendar years. Explanatory variables are taken from the respective works of Altman (1968) and Altman (1983). Mitigating the effect of extreme values we remove the firms with values of any of the five X variables in the highest and lowest 0.5% of the observations for that variable in the full data set. Similar as Altman et al. (2016), who winsorized the independent variables at the 1 and 99 percent level to minimize extreme outliers, we winsorize it at the 5 percent and 95 percent level.

Table 4. Description of Explanatory Variables

Variable	Description
X1	Working Capital to Total Assets
X2	Retained Earnings to Total Assets
X3	Earnings before Interest and Taxes to Total Assets
X4	Net Worth to Total Liabilities
X5	Sales to Total Assets
X1W5	Working Capital to Total Assets, winsorized at the 5^{th} and the 95^{th} percentile
X2W5	Retained Earnings to Total Assets, winsorized at the 5^{th} and the 95^{th} percentile
X3W5	Earnings before Interest and Taxes to Total Assets, winsorized at the 5^{th} and the 95^{th} percentile
X4W5	Net Worth to Total Liabilities, winsorized at the 5^{th} and the 95^{th} percentile
X5W5	Sales to Total Assets, winsorized at the 5^{th} and the 95^{th} percentile

We will utilize the winsorized variables at 5% level as shown in Table 4.

4.3.1 Non-financial Variables

The financial variables seen till now, encompassing metrics pertinent to liquidity, profitability, leverage, and activity among others, serve as a comprehensive reflection of firms' operational efficacy. Considering that default primarily manifests as a financial debacle, denoting a company's failure to meet its financial obligations punctually and in full, it comes as no surprise that financial variables stand as principal prognosticators of default, hence, constituting integral components of failure prognosis frameworks. Nevertheless, there exist compelling rationales to supplement these financial metrics with non-financial counterparts. Unlisted enterprises lack access to market-derived insights, relying solely on accounting data for assessment and in this situation non-financial indicators assume critical significance in increasing default analysis. In fact, earlier studies and recent literature concluded that quantitative variables are not sufficient to predict SME default and that including qualitative variables (such as the number of employees, the legal form of the business, the region where the main business is carried out, the industry type, etc.) improves the models' prediction power. Grunert, Norden, and Weber (2005) propose that integrating both financial and non-financial elements leads to more precise default forecasts. They conducted an analysis on 409 instances spanning various company-years and employed the bootstrap method to validate their findings. Altman, Sabato, and Wilson (2010) observed that incorporating non-financial attributes of firms significantly enhances the predictive capability of risk models. Their investigation centered on UK SMEs, encompassing over 5.8 million observations across company-years. Altman et al. (2014), Káčer, Ochotnicky and Alexy (2019), along with Altman et al. (2017), all affirm that introducing supplementary non-financial variables enhances the accuracy of model classifications.

In our analysis we will use two different non-financial variables: *age of the firm and the size of the firm.* Regarding age, when examining its impact on business failure, two conflicting trends emerge. Firstly, as a company matures, it accrues experience and profits over time, leading to a decrease in its likelihood of failure. This is attributed to the accumulation of expertise and financial stability. Conversely, there's the notion that failure requires time, and young companies often possess startup capital to sustain themselves, despite inadequate revenue generation. Combining these perspectives, it's expected that failure rates are relatively low in the initial years when startup capital supports the company. However, as this capital diminishes, and the company continues to evolve, the likelihood of failure tends to peak during this transitional period marked by ongoing experiential learning. Even if Altman (1983) noted that the age of a firm is implicitly considered in the Retained Earnings/Total Assets ratio (X2) we will still explicitly include it in the model. Since we use a logit model, the variable will be transformed in log of age (log_AGE). Concerning the size of the company, we know from previous literature that the boundary between bankrupt and non-bankrupt firms is different for small and larger firms. In our case we are working with medium and small enterprises and not large ones but we will try to analyze if there is an improvement of the model by including the size category. The size variable will be a dummy variable where 0 equals small size company and 1 equals medium size company.

5. Empirical Results

5.1 Descriptive Statistics

An overview of the descriptive statistics of the full sample is presented in Tables 5A and 5B. The first one refers at the variables of the training sample while the second one at the validation sample. The descriptive statistics (mean, standard deviation and quartiles) are calculated both for failed and non-failed companies. The last column indicate p-values of the t-test of the difference of means for explanatory variables in the defaulted and non-defaulted groups.

Table 5A. Descriptive statistics for training sample

			Non-defaulted companies					Defaulted Companies			
			N = 97 337					N = 1 297			T-value
Variable	Mean	SD	p25	p50	p75	Mean	SD	p25	p50	p75	
X1W5	0.25	0.24	0.08	0.26	0.44	0.13	0.32	-0.22	0.08	0.40	18.25***
X2W5	0.04	0.07	0.00	0.02	0.07	-0.03	0.09	-0.10	-0.07	0.01	35.18***
X3W5	0.06	0.08	0.01	0.04	0.10	-0.01	0.11	-0.10	-0.07	0.03	32.44***
X4W5	0.30	0.22	0.11	0.26	0.45	0.16	0.26	-0.05	0.03	0.32	20.93***
X5W5	1.05	0.53	0.65	0.97	1.37	1.08	0.71	0.43	0.91	1.64	-1.90
log_AGE	2.66	0.86	2.07	2.77	3.36	2.43	0.84	1.79	2.40	3.09	9.41***
SIZE	0.56	0.49	0	1	1	0.21	0.41	0	0	0	25.02***

Table 5B. Descriptive statistics for validation sample

			Non-defaulted companies					Defaulted Companies			
			N = 88 149					N = 1 268			T-value
Variable	Mean	SD	p25	p50	p75	Mean	SD	p25	p50	p75	
X1W5	0.26	0.24	0.09	0.26	0.44	0.14	0.33	-0.20	0.07	0.43	17.58***
X2W5	0.04	0.06	0.00	0.02	0.07	-0.04	0.09	-0.10	-0.11	0.00	41.62***
X3W5	0.06	0.08	0.01	0.04	0.10	-0.02	0.10	-0.10	-0.10	0.02	37.12***
X4W5	0.30	0.21	0.12	0.27	0.46	0.17	0.26	-0.04	0.02	0.35	21.47***
X5W5	1.10	0.52	0.71	1.02	1.42	1.15	0.70	0.49	1.01	1.76	-3.6
log_AGE	2.75	0.79	2.19	2.83	3.4	2.55	0.81	1.94	2.56	3.25	9.06***
SIZE	0.58	0.49	0	1	1	0.25	0.44	0	0	1	23.57***

The tables shows descriptive statistics of the explanatory variables. SD stands for standard deviation, p25, p50 and p75 stand for lower quartile, median and upper quartile, respectively. The statistical significance is indicated with asterisks (*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$).

As we can see on average all the ratios are larger for non-bankrupt firms, which is in line with expectations. Altman (2000) publishes similar results as all ratios of the non-bankrupt firms are on average higher than the bankrupt firms. Moreover, the five financial factors appear effective in distinguishing between failed and non-failed companies, as their averages are notably lower for defaulted firms compared to those that have not defaulted. The difference between the means is statistically significant as indicated by t-test.

As in Altman (1968) the X5 variable has the highest mean and the T-value is not statistically significant. Also, as seen in Ohlson (1980), the ratios deteriorate as one moves from non-bankrupt firms to two years prior to bankruptcy to one year prior to bankruptcy.

This confirms our first hypothesis, meaning that financial ratios vary between solvent and insolvent companies.

5.2 Estimation Results

5.2.1 Training Sample

We are going to analyze the model in the training sample in order to be able to validate the results done with the validation sample. The primary purpose of dividing the dataset into a validation set is to avoid overfitting. This means ensuring that the model doesn't just excel at categorizing the examples in the training set but also can generalize well and make precise predictions on new, unseen data.

First, we estimate the model based on the five financial ratios used in model developed by Altman (1983). As stated before the model is estimated using logistic regression with 1 = failed; 0 = non-failed so that we expect that a negative coefficient indicates a reduced risk of insolvency and a positive coefficient an increased risk of insolvency.

Table 6 presents estimation results. We estimated six models. The dependent variable is the indicator of default and the explanatory variables are described in Table 4.

As said the first model contains the financial variables from the revised Altman's Z' - score model. All variables are statistically significant beside one, variables X1 (working capital to total assets), X2 (retained earnings to total assets), X4 (net worth to total liabilities) and X5 (sales to total assets) are significant at 1% level. X3 (EBIT to total assets) is not statistically significant.

The model achieves relatively low McFadden pseudo-R^2, even though its in-sample discriminatory ability measured by the area under ROC curve (AUC) is relatively high.

Table 6. Estimation Results

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
X1W5	-0.377***	-0.375***	-0.390***	-0.444***	-0.446***	-0.450***
	(0.132)	(0.132)	(0.133)	(0.130)	(0.130)	(0.130)
X2W5	-16.27***	-14.80***		-11.75***	-12.94***	
	(1.851)	(0.490)		(1.930)	(0.480)	
X3W5	1.302		-12.40***	-1.050		-10.91***
	(1.585)		(0.435)	(1.651)		(0.421)
X4W5	-0.580***	-0.590***	-0.789***	-0.392**	-0.386**	-0.520***
	(0.175)	(0.175)	(0.175)	(0.173)	(0.173)	(0.173)
X5W5	0.260***	0.268***	0.321***	0.323***	0.316***	0.371***
	(0.0496)	(0.0486)	(0.0490)	(0.0495)	(0.0485)	(0.0488)
log_AGE				0.0809**	0.0807**	0.0825**
				(0.0358)	(0.0358)	(0.0359)
SIZE				-1.283***	-1.280***	-1.322***
				(0.0723)	(0.0721)	(0.0720)
Constant	-4.333***	-4.317***	-4.120***	-4.103***	-4.116***	-3.948***
	(0.0764)	(0.0738)	(0.0716)	(0.117)	(0.115)	(0.113)
Observations	98,623	98,623	98,623	98,622	98,622	98,622
Log likehood	-6165.15	-6165.48	-6297.77	-5975.66	-5975.87	-5993.06
McFadden's R^2	0.1050	0.1050	0.1003	0.1320	0.1320	0.1295
AUC (training)	0.746	0.747	0.749	0.783	0.783	0.784
AUC (validation)	0.775	0.775	0.770	0.800	0.800	0.793

The table shows estimation results for default model. The parameters are estimated using logistic regression. Standard errors are denoted in parentheses and the statistical significance is indicated by asterisks (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

The X3 and X2 variables attract opposite signs and in fact the first one is not statistically significant and second one is. Since it may be due to multicollinearity, we are going to check the pairwise correlation between the variables. The correlations matrix is a comparison of how closely related two variables are.

Table 7 shows that there are not very high correlations between variables beside the pairwise correlation of X2 and X3, which are nearly perfectly correlated (0.973).

Table 7. Correlation Matrix

	X1W5	X2W5	X3W5	X4W5	X5W5
X1W5	1.000				
X2W5	0.344	1.000			
X3W5	0.326	0.973	1.000		
X4W5	0.568	0.410	0.369	1.000	
X5W5	0.026	0.247	0.296	-0.173	1.000

Table 7 presents the correlation matrix of the variables.

Since the two variables indicate both profitability of a company (retained earnings and EBIT) and they are nearly perfectly correlated, we try to exclude one variable at time and re-estimate the model without them.

As we can see from table 6, the second model is estimated without the variable X3, yet the log-likelihood, McFadden R^2 and AUC are basically unchanged.

All variables are now significant at 1% level. We can compare the results with model 3 which is estimated without the variable X2. As we can see, the log-likelihood, McFadden R^2 and AUC are unchanged again. Also in this case all variables are significant at 1% level. Looking at the area under ROC curve (AUC), it slightly increased in the second and third model, compared to the first one, but it can not be considered noteworthy. This confirms the fact that the variables X2 and X3 together do not contribute new information to the model due to their high correlation.

We want to analyze now the effect of the non-financial variables on the model to see if the accuracy improves. We can see the results in models 4,5 and 6.

Model 4 contains all the original five financial variables and the two non-financial one. All financial variables maintain their original sign and are statistically significant at 1% level beside X3, which becomes negative but still not statistically significant. The size of the company is statistically significant at 1% while the age variable is statistically significant at 5% level. What we understand from the sign of the size variable is that small-sized companies are more risky than the medium-sized ones.

Looking at the models with non-financial variables we can see that the discriminatory performance improved when compared to models without, so clearly these variables provide additional information for the models. Even if the area under ROC curve (AUC) increased just by roughly 3%, we can still state that the non-financial variables improve the performance of the model. The results are similar to those of Altman and Sabato (2008). They found that when adding non-financial and compliance information to the basic accounting model the core variables retain their signs, as in our analysis. Regarding the significance of the non - financial variables, they add value to the model (AUC of 0.80) with an improvement of over 8% compared with the AUC of the model using only financial information (0.74).

In our analysis we have the same performance of the original model with only financial variables (AUC of 0.74) and we get nearly the same results adding our two non-financial variables (AUC of 0.78). We could have even a better improvement of the model by adding more qualitative info, in fact this is one of our limitation.

5.2.2 Validation Sample

Now we are going to do the analysis on our validation sample in order to compare them to the results obtained previously.

In Table 6, for each estimated model, the AUC for the training sample (in-sample performance) is shown in the second-to-last row, while the AUC for the validation sample (out-of-sample performance) is displayed in the last row. Comparing these two values allows us to determine whether our models captured genuine relationships among variables or merely random noise unique to the estimation sample. As we can see from figure 2 and 3, the area under ROC curve (AUC) of models for the validation sample reaches 0.77 using only financial variables and 0.80 introducing the non financial ones.

Figure 2. ROC curve of Model 2.

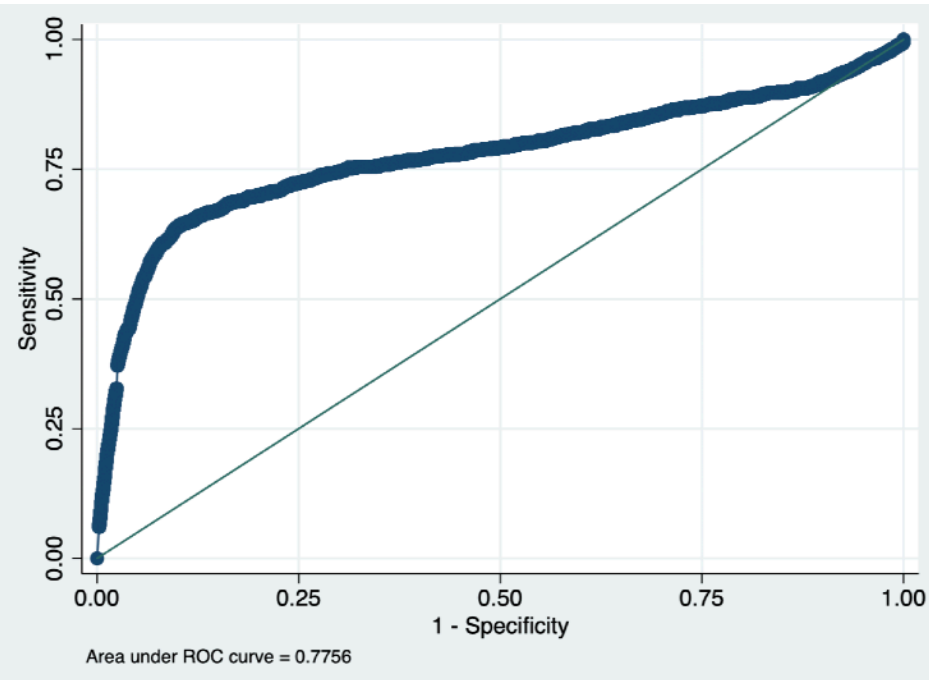

Figure 2 shows the area under ROC curve (AUC) of model 1.

Figure 3. ROC curve of Model 5.

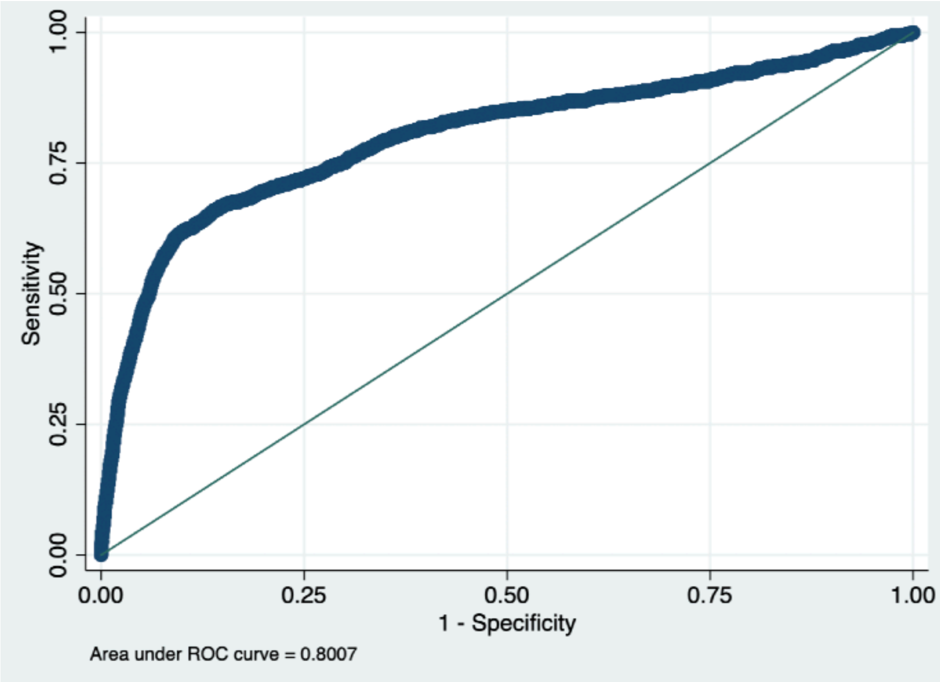

Figure 3 shows the area under ROC curve (AUC) of model 5.

From the standpoint of the obtained AUC values, there are no universally accepted benchmarks or ranges, as these values are context-dependent. Typically, individual values exceeding 0.8 are regarded as very good, while those over 0.9 are deemed excellent. Nevertheless, in terms of our research hypotheses, AUC facilitates straightforward comparison of the models and allows for tests of statistical significance.

As in the training sample, the area under ROC curve (AUC) increased just by roughly 3% but we can still say that the non-financial variables improve the performance of the model. The most important outcome of this analysis is that the performance of models in the validation sample is similar to that in the training sample, i.e. the models are not over-fitted and the relations between variables in the training sample captured by the models continue to hold in the validation sample as well.

5.3 Two-years prior bankruptcy

We want to see how the model and its accuracy perform by doing the analysis with the financial statements of the companies two years before bankruptcy. Altman (1968) evaluated the model up to five years prior to bankruptcy. We have to state that the analysis is made just to observe how the model would behave. But it cannot be considered valid since we do not have a training sample to verify it. Results are shown in Table 8.

Table 8. Estimation Results

VARIABLES	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
X1W5	0.281**	0.284**	0.265*	0.183	0.184	0.171
	(0.136)	(0.136)	(0.137)	(0.133)	(0.133)	(0.133)
X2W5	-15.64***	-10.90***		-10.97***	-9.388***	
	(1.877)	(0.577)		(1.923)	(0.544)	
X3W5	4.082***		-8.407***	1.363		-7.440***
	(1.540)		(0.487)	(1.587)		(0.457)
X4W5	-1.262***	-1.300***	-1.501***	-0.940***	-0.949***	-1.080***
	(0.181)	(0.181)	(0.180)	(0.179)	(0.179)	(0.179)
X5W5	0.120**	0.150***	0.183***	0.164***	0.175***	0.215***
	(0.0562)	(0.0549)	(0.0556)	(0.0561)	(0.0546)	(0.0552)
log_AGE				-0.0281	-0.0279	-0.0215
				(0.0405)	(0.0406)	(0.0406)
SIZE				-1.220***	-1.224***	-1.256***
				(0.0679)	(0.0677)	(0.0676)
Constant	-3.976***	-3.930***	-3.791***	-3.468***	-3.453***	-3.354***
	(0.0804)	(0.0786)	(0.0770)	(0.131)	(0.129)	(0.129)
Observations	89,417	89,417	89,417	89,416	89,416	89,416
Log likehood	-6312.79	-6316.03	-6342.40	-6112.55	-6112.91	-6127.67
McFaddens' R^2	0.051	0.051	0.047	0.081	0.081	0.079
AUC	0.697	0.699	0.693	0.746	0.746	0.743

The table shows estimation results for default model. The parameters are estimated using logistic regression. Standard errors are denoted in parentheses and the statistical significance is indicated by asterisks (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Similarly to the previous analysis, all variables are statistically significant and X2 - X3 still attract opposite sign due to high correlation (0.97). We can see that the McFadden R^2 and area under ROC curve (AUC) are lower in all models compared to the analysis on one year before bankruptcy. Still, the AUC is relatively high and also in this case the addiction of non-financial variables increases the model accuracy by roughly 5%.

This is perfectly in line with the literature stating that the model loses its accuracy by increasing the time horizon of the financial statements.

5.4 Classification Accuracy

We are going to analyze the classification accuracy for both samples.

As mentioned before, our cut-off value is fixed at the level of the sample default rate, which for the training sample is (0.0131496) and for the validation sample (0.0141807). Table 9 shows the classification accuracy figures for each model with the respective cut-off point using training sample.

Table 9. Classification Accuracy of the Models Using Training Sample

				Failed Companies				Non-failed Companies			
		Average Accuracy	Overall Accuracy	Failure Accuray (True positives)		Type I error (False negatives)		Non-failure Accuracy (True negatives)		Type II error (False positives)	
		%	%	Obs.	%	Obs.	%	Obs.	%	Obs.	%
Model 1	cut-off	70.84%	73.10%	886	68.52%	407	31.48%	71.216	73.16%	26.121	26.84%
Model 2	cut-off	70.84%	73.02%	887	68.60%	406	31.40%	71.139	73.09%	26.198	26.91%
Model 3	cut-off	70.36%	70.95%	902	69.76%	391	30.24%	69.079	70.97%	28.258	29.03%
Model 4	cut-off	72.37%	71.83%	943	72.93%	350	27.07%	69.911	71.82%	27.426	28.18%
Model 5	cut-off	72.43%	71.87%	944	73.01%	349	26.99%	69.942	71.86%	27.395	28.14%
Model 6	cut-off	72.25%	71.37%	946	73.16%	347	26.84%	69.451	71.35%	27.886	28.65%

The table shows classification accuracy figures for the six models of the training sample with the cut-off point of (0.0131496).

The overall accuracy is a weighted average of true positives and true negatives percentages; the weights are number of observations in each group. The average accuracy is a simple average of true positives and true negatives percentages to account for unbalanced sample. Since we have a sample with much more non-defaulted firms, we are going to consider more valid the average accuracy. The first three models with only financial variables achieve a similar average accuracy, about 70%, with Type I error around 30%. Regarding the Type II error i.e, those firms classified as default but which did not default, it is around 29% for the first three models. Looking at the last three models with all explanatory variables included, we can see that the accuracy slightly increases to 72%, confirming the fact that with non-financial variables the models achieve better results. Most importantly, the Type I error decreases to around 27% while the Type II error remains stable at 28%. While minimizing both types of errors is ideal, financial analysts and firms often focus more on reducing Type I errors because the immediate and tangible costs tend to be higher.

Similar results can be seen in the validation sample showed in Table 10.

Table 10. Classification Accuracy of the Models Using Validation Sample

				Failed Companies				Non-failed Companies			
		Average Accuracy	Overall Accuracy	Failure Accuray (True positives)		Type I error (False negatives)		Non-failure Accuracy (True negatives)		Type II error (False positives)	
		%	%	Obs.	%	Obs.	%	Obs.	%	Obs.	%
Model 1	cut-off	75.26%	80.03%	898	70.82%	370	29.18%	67.971	77.11%	20.178	22.89%
Model 2	cut-off	75.47%	79.37%	900	70.98%	368	29.02%	68.089	77.24%	20.060	22.76%
Model 3	cut-off	73.04%	74.13%	907	71.53%	361	28.47%	67.147	76.17%	21.002	23.83%
Model 4	cut-off	74.33%	75.74%	926	73.03%	342	26.97%	64.435	73.10%	23.714	26.90%
Model 5	cut-off	74.11%	75.16%	926	73.03%	342	26.97%	64.413	73.07%	23.736	26.93%
Model 6	cut-off	73.77%	74.11%	923	72.79%	345	27.21%	64.745	73.45%	23.404	26.55%

The table shows classification accuracy figures for the six models of the validation sample with the cut-off point of (0.0141807).

All six models have an higher average accuracy and lower Type I and Type II errors. Models with only financial variables achieve an average accuracy of about 73% - 75%, with Type I error around 29% and Type II error raging 22-23%. Also in this case the introduction of non - financial variables decreases the Type I error and increases the average accuracy. Interestingly, the reduced model 2 achieves a slightly better score confirming our conjecture that it is better to remove non-significant variables or those with negative signs.

We are going to examine now the precision of the model two years prior bankruptcy. Results are shown in Table 11.

Table 11. Classification Accuracy two years prior bankruptcy

				Failed Companies				Non-failed Companies			
		Average Accuracy	Overall Accuracy	Failure Accuray (True positives)		Type I error (False negatives)		Non-failure Accuracy (True negatives)		Type II error (False positives)	
		%	%	Obs.	%	Obs.	%	Obs.	%	Obs.	%
Model 1	cut-off	64.25%	60.23%	867	68.38%	401	31.62%	52.994	60.12%	35.155	39.88%
Model 2	cut-off	64.51%	59.69%	881	69.48%	387	30.52%	52.497	59.55%	35.652	40.45%
Model 3	cut-off	64.04%	58.52%	884	69.72%	384	30.28%	51.444	58.36%	36.705	41.64%
Model 4	cut-off	68.41%	68.29%	869	68.53%	399	31.47%	60.196	68.29%	27.953	31.71%
Model 5	cut-off	68.44%	68.27%	870	68.61%	398	31.39%	60.179	68.27%	27.970	31.73%
Model 6	cut-off	68.50%	68.00%	875	69.01%	393	30.99%	59.930	67.99%	28.219	32.01%

The table shows classification accuracy figures for the six models of the validation sample two years prior bankruptcy with the cut-off point of (0.0141807).

The cut-off point is the same as before since we are working on the validation sample. In the first three models, compared to one year before bankruptcy, we can see that the average accuracy decreased to 64% while the Type I and Type II errors increased to 30% and 40% respectively. Also in this case the models with non - financial variables have a better average accuracy (68%) and a lower Type II error (32%).

The results confirm the third hypothesis (H3) of the thesis, reinforcing the argument that non-financial variables provide valuable additional information that enhances the discriminatory performance and classification accuracy of the models. This is in line with the findings of Altman and Sabato (2007) and Altman, Sabato and Wilson (2008). The empirical findings consistently demonstrate that the hybrid models achieve higher overall accuracy, improving the precision of predictions. Notably, these models exhibit a significant reduction in Type I errors, which are critical in the context of financial distress prediction. Type I errors, which occur when distressed firms are misclassified as healthy, can have severe financial repercussions. Therefore, the reduction of these errors is of paramount importance to financial analysts, investors, and firms themselves, who rely on these predictions to make informed decisions and mitigate potential losses.

Comparing to the literature on Italian SME's prediction, Modina and Pietrovito (2014) use a logistic regression model on a database of 9208 Italian limited liabilities SMEs in a time frame of 3 years, over the period 2006 to 2010 and found a percentage of correctly predicted defaulted companies of 80%.

Ciampi (2015) apply a logistic regression to a sample of 934 Italian small enterprises (SEs) from 2010 and got a percentage of correctly predicted defaulted companies of 84%.

Calabrese, Marra and Osmetti (2016) employed the BGEVA model, logistic additive regression as well as log-log additive regression on a sample of 49 738 Italian SMEs from the period 2006-2011 and got an AUC of 0.81 with the logistic model. We can say that our findings accuracy are slightly below the level of the existing literature but still perfectly in line so we can also confirm our second hypothesis (H2) that there is a positive relationship between the Altman Z'- score model variables and predicting bankruptcy for Italian SMEs.

6. Conclusion

In this thesis, we have explored the validity of Altman's Z-Score as a predictor of company failure, specifically focusing on the Italian context. The research aimed to evaluate the effectiveness of this model when applied to Small and Medium-sized Enterprises (SMEs), which constitute a significant portion of the Italian economy. Our findings indicate that Altman's Z-Score remains a robust tool for bankruptcy prediction, even decades after its development. The model demonstrated a satisfactory accuracy rate in predicting financial distress among Italian SMEs. However, the inclusion of non-financial variables improved the model's predictive power. This enhancement was evident in the reduction of Type I errors and the overall increase in classification accuracy. Through the analysis, we confirmed that while financial ratios are crucial indicators of a firm's health, non-financial variables provide additional insights that can enhance prediction accuracy. This is particularly important for SMEs, where non-financial factors may play a more significant role in their stability and performance. The study also highlighted the model's diminishing accuracy with increasing prediction horizons, aligning with existing literature that stresses the need for timely and frequent evaluations of financial health. The findings confirmed our three initial hypothesis. In conclusion, Altman's Z-Score, augmented with non-financial variables, presents a valuable tool for stakeholders in the Italian market to assess the risk of corporate failure.

While this study attempted to replicate the Altman model, future research could explore beyond this framework. Further research could perhaps focus more on other, unexamined financial ratios or even build Italian-specific ratios, if such could be created and deemed necessary. Additionally, leveraging more advanced statistical techniques or machine learning algorithms could potentially improve the predictive accuracy of bankruptcy models. These approaches could incorporate a wider variety of data sources, including macroeconomic indicators, industry-specific trends, and real-time financial data, to better capture the dynamics affecting SME stability.

7. References

Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, The Journal of Finance, Vol. 23 No. 4, pp. 589-609.

Altman, E.I. (2000). Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models.

Altman, E.I. (1983). Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy. New York: John Wiley & Sons. ISBN 978-0-471-08707-6.

Altman, E.I., Balzano, M., Giannozzi, A., Srhoj, S. (2023). The Omega Score: An improved tool for SME default predictions. Journal of the International Council for Small Business, Taylor & Francis Journals, vol. 4(4), pp. 362-373.

Altman, E.I., Esentato, M., Sabato, G. (2020). Assessing the credit worthiness of Italian SMEs and mini-bond issuers. Global Finance Journal, Elsevier, vol. 43(C).

Altman, E.I., Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U.S. Market. Abacus, *43*, No. 3, pp. 332 – 357.

Altman, E.I., Sabato, G., Wilson, N. (2008). The Value of Quantitative Information in SME Risk Management. Working Paper, Stern School of Business: 1-40.

Altman, E.I., Sabato, G., Wilson, N. (2010). The Value of Non-financial Information in Small and Medium-sized Enterprise Risk Management. Journal of Credit Risk, *6*, No. 2, pp. 95 – 127.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. Journal of International Financial Management & Accounting, *28*, No. 2, pp. 131 – 171.

Altman, E.I., Loris, B. (1976). A Financial Early Warning System for Over-the-Counter Broker-Dealers. The Journal of Finance, 31, 1201-1217.

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

Beaver, W. H. (1968). Alternative accounting measures as predictors of failure, The Accounting Review, Vol. 43 No. 1, pp. 113-122.

Berger, A.N. and Udell, G.F. (2002). Small Business Credit Availability and Relationship Lending: The Importance of Bank Organizational Structure. The Economic Journal, 112, 32-53.

Calabrese, R., Marra, G., Osmetti, S. A. (2016). Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. Journal of the Operational Research Society, Palgrave Macmillan; The OR Society, vol. 67(4), pp. 604-615.

Celli, M. (2015). Can Z-Score Model Predict Listed Companies' Failures in Italy? An Empirical Test. International Journal of Business and Management, 10 (3), 57-66.

Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. Journal of Business Research, Elsevier, vol. 68(5), pp. 1012-1025.

Ciampi, F., Giannozzi, A., Marzi, G., Altman, E.I. (2021). Rethinking SME default prediction: a systematic literature review and future perspectives, Scientometrics, Springer; Akadémiai Kiadó, vol. 126(3), pp. 2141-2188.

Ciampi, F., Gordini, N. (2013). Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. Journal of Small Business Management, Taylor & Francis Journals, vol. 51(1), pp. 23-45.

Daily. C. M., Dalton, D. R. (1994). Corporate governance and the bankrupt firm: An empirical assessment. Strategic Management Journal, Wiley Blackwell, vol. 15(8), pp. 643-654.

Dainelli, F., Bini, L., Giunta, F. (2013). Signaling strategies in annual reports: Evidence from the disclosure of performance indicators. Advances in accounting, Elsevier, vol. 29(2), pp. 267-277.

Dietsch, M., Petey, J. (2004). Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. Journal of Banking & Finance, Elsevier, vol. 28(4), pp. 773-788.

Engelmann, B., Hayden, E., Tasche, D. (2003). Measuring the Discriminative Power of Rating Systems. Deutsche Bundesbank Discussion Paper, Series 2 (no 01).

Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. EXPERT SYSTEMS WITH APPLICATIONS, 41(14), 6433-6445.

Grunert, J., Norden, L., Weber, M. (2005). The role of non-financial factors in internal credit ratings. Journal of Banking & Finance, Elsevier, vol. 29(2), pp. 509-531.

Káčer, M., Ochotnický, P., Alexy, M. (2019). The Altman's Revised Z'-Score Model, Non-financial Information and Macroeconomic Variables: Case of Slovak SMEs. In Ekonomický časopis/Journal of Economics, vol. 67, no.4, pp. 335-366.

Keasey, K., Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses. Journal of Business Finance and Accounting, 14(3), 335–354.

Lee, T. S., Yeh, Y. H. (2004). Corporate Governance and Financial Distress: evidence from Taiwan. Corporate Governance: An International Review, Wiley Blackwell, vol. 12(3), pp. 378-388.

Luppi, B., Marzo, M., Scorcu, E. (2007). A credit risk model for Italian SMEs. Working Papers 600, Dipartimento Scienze Economiche, Universita' di Bologna.

Marshall, M.N. (1996). Sampling for Qualitative Research. Family Practice, 13, 522-525.

Mcleay, S., Omar, A. (2000). The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios. The British Accounting Review, 32(2), 213-230.

Modina, M., Pietrovito, F. (2014). A default prediction model for Italian SMEs: the relevance of the capital structure. Applied Financial Economics, 24(23), 1537–1554.

Mselmi, N., Lahiani, A., Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. International Review of Financial Analysis, Elsevier, vol. 50(C), pp. 67-80.

Norden, L., Weber, M. (2010). Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers. The Review of Financial Studies, 23, 3665-3699.

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, *18*, No. 1, pp. 109 – 131.

Pederzoli, C., Torricelli, C. (2010). A parsimonious default prediction model for Italian SMEs. Centro Studi di Banca e Finanza (CEFIN) (Center for Studies in Banking and Finance) 0022, Università di Modena e Reggio Emilia, Dipartimento di Economia "Marco Biagi".

Pozzoli, M., Paolone, F. (2016). An Overlook at Bankruptcy Prediction in Italy in 2016: An Application of the Altman‟s Model on Failed Italian Manufacturing Companies In The 2016 - First Quarter. International Journal of Accounting and Financial Reporting 6(2):293-309.

Saurina, J., Trucharte, C. (2004). The Impact of Basel II on Lending to Small - and Medium-Sized Firms: A Regulatory Policy Assessment Based on Spanish Credit Register Data. Journal of Financial Services Research, Springer; Western Finance Association, vol. 26(2), pp. 121-144.

Tabouratzi, E., Lemonakis, C., & Garefalakis, A. (2017). Determinants of failure in Greek manufacturing SMEs. Corporate Ownership & Control, 14(3), 45-50.

Taffler, R. J. (1982). TForecasting company failure in the U.K: Using discriminant analysis and financial ratio data. Journal of Royal Statistical Society, 145(3), 342–358.

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. Journal of Accounting Research, *22*, No. 2, pp. 59 – 82.