Universidade do Minho
Escola de Economia e Gestão

Roxana Sera

Financial Distress Prediction for Portuguese SMEs

October 2020

Roxana Sera Financial Distress Prediction for Portuguese SMEs

Uminho | 2020

Universidade do Minho
Escola de Economia e Gestão

Roxana Sera

Financial Distress Prediction for Portuguese SMEs

Master Degree Project
Master in Finance

Supervisor:
Florinda Conceição Cerejeira Campos da Silva, PhD

October 2020

Acknowledgments

I would like to thank all of those who have made possible the completion of this study.

I would like to thank my advisor, Professor Florinda Silva, for always being open and eager to share her knowledge with a lot of patience and good disposition.

I would also like to thank Professor Sónia Silva for her invaluable comments; allowing me to share her passion for econometrics helped increase mine.

My gratitude also goes to nBanks company, for making this project possible and always being there for support.

Heartfelt thanks also go to Mr. José Fernandes with additional tribute to his late wife Mrs. Alice Cracel Fernandes, who will always live in our hearts, for their enormous help and constant encouragements.

In addition, I must express my very profound gratitude to my parents, Vasile and Valeria, to my spouse Mizuaki, and to my children Alexia and Ivan for providing me with unfailing support and love. You are always there for me.

Finally, I thank all my family and friends for their continuous support and sympathetic ear.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Em Portugal, as Pequenas e Médias Empresas (PMEs) representam 99.9% do número total de empresas e são um fator chave para a geração de emprego, com uma contribuição elevada para a economia geral do país. Considerando o papel estratégico desempenhado e o fato de que a maior fonte de recursos para as PMEs são as instituições financeiras, é fundamental que essas tenham tanto facilidade de acesso à instrumentos financeiros diversificados, quanto a possibilidade de apresentar a sua atividade e resultados obtidos de uma maneira adequada que lhes garante acesso a esses instrumentos.

Nesse contexto, a aplicação de um modelo de previsão de insolvência baseado na análise de rácios financeiros é uma maneira de interpretar a informação disponível sobre uma empresa de uma forma clara, concisa e eficiente. A análise facilitada por tal instrumento beneficia tanto as instituições financeiras, que podem interpretar os resultados obtidos para melhor entender a situação geral da empresa, quanto os gestores da empresa, para quais facilita a detecção e prevenção de eventuais problemas financeiros.

O objetivo deste estudo é identificar os principais rácios financeiros relevantes para distinguir entre empresas em dificuldades financeiras e empresas saudáveis, estimar com base neles um modelo de previsão de insolvência e utilizar os parámetros estimados para previsão de dificuldades financeiras nas PMEs portuguesas.

Para obter uma amostra mais equilibrada de empresas foi aplicado o método *Propensity Score Matching,* com pareamentos de um-para-um e um-para-muitos. O modelo foi estimado com base nos dados financeiros de empresas insolventes de um ano antes da insolvência. Testes de validação foram feitos em amostras de um, dois e três anos antes da insolvência, amostra de um a três anos antes da falência, bem como no inteiro conjunto de empresas com dados disponíveis, até seis anos antes da insolvência.

As cinco variáveis que mostraram melhor capacidade de previsão da insolvência são: Ativo Corrente/ Total do Ativo, Fluxo de Caixa Operacional/ Total do Ativo, Fluxo de Caixa Operacional/ Total do Ativo, Resultados Transitados/ Total do Ativo e Património Líquido/ Total do Passivo. A capacidade total preditiva do modelo é acima de 85%, o que leva à conclusão de que o modelo pode ser aplicado ao mercado Português, no contexto das PMEs.

Palavras-chave: Insolvência, Rácios Financeiros, Modelo de Previsão de Dificuldade Financeira, Modelos Logísticos, *Propensity Score Matching,* Pequenas e Médias Empresas

Abstract

In Portugal, small and medium-sized enterprises (SMEs) represent 99.9% of the total number of companies and are key generators of employment and contributors to the country`s economy. Given their key role and the fact that their main source of funding comes from financial institutions, it is vital that they have easy access to diversified financing instruments as well as the capacity of presenting their activity and results in an efficient way in order to gain access to them.

In this context, a way of interpreting the information available about a company in a clear, concise and efficient manner is through the application of an accounting - based financial distress model. The analysis provided by such an instrument is beneficial to both financial institutions, that can use the results in order to understand the general situation of the company, and to the company`s management, who can foresee and prevent eventual financial problems.

The objective of this study is to identify the main financial ratios that are relevant in order to discriminate between financially distressed and healthy companies and estimate financial distress prediction models based on them then use the estimated parameters to predict the probability of financial distress in Portuguese SMEs.

In order to obtain a more balanced data set of companies the propensity score method, with matching of one-to-one as well as one-to-many, was applied. The model estimation was made with insolvent companies` data from one year prior to insolvency. Validation tests were performed on data samples for one, two and three years before insolvency, as well as for years one to three in a joint data set and also for the entire set of insolvent companies available, up to six years prior to insolvency.

The five variables found to be the best predictors of insolvency are Current Assets to Total Assets, Operating Cash Flow to Total Assets, Operating Cash Flow to Debt, Retained Earnings to Total Assets and Equity to Debt. The overall forecasting accuracy of the final model was of over 85%, by which we conclude that the model could be successfully applied to the Portuguese market, in the context of the SMEs.

Keywords: Insolvency, Financial Ratios, Financial Distress Prediction Model, Propensity Score Matching, Logistic Models, Small and Medium-sized Enterprises

Table of Contents

Resumo iv

Abstract vi

List of Abbreviations viii

List of Tables ix

List of Figures x

1 Introduction 1

2 Presentation of NBanks Company 4

3 Theoretical Framework 6

3.1 SME Definition and Insolvency 6

3.1.1 SME Definition 6

3.1.2 Insolvency 6

3.2 Literature Review 9

3.2.1 Univariate Models 9

3.2.2 Multivariate Models 10

3.2.3 Linear Probability Models 14

3.2.4 Other Models 16

4 Methodology and Data 18

4.1 Objectives 18

4.2 Methodology and Variables 18

4.3 Data Set 20

4.4 Descriptive Statistics 27

4.5 Ratio Selection 33

5 Analysis and Discussion of Results 35

5.1 Considerations on Results for Project Hosting Company nBanks 45

6 Conclusions 47

References 49

Appendices 53

List of Abbreviations

AMADEUS – Analyse Major Databases from European Sources

CIRE – Insolvency and Business Recovery Code

GOF – Goodness-Of-Fit

IAPMEI - Portuguese Institute of Support to Small and Medium-sized Enterprises and Innovation

LDA – Linear Discriminant Analysis

LRA – Logistic Regression Analysis

NACE – Statistical Classification of Economic Activities in the European Community

NUTS – Nomenclature of Territorial Units for Statistics

PCA – Principal Component Analysis

PS – Propensity Score

PSM – Propensity Score Matching

SME – Small and Medium-Sized Enterprises

List of Tables

Table 1: Causes of Insolvency 8

Table 2: Altman`s Z-Score Model - Variable Means and Test of Significance 11

Table 3: Altman`s Z-Score Model Predictive Accuracy 12

Table 4: Predictive Accuracy of Ohlson`s Models 15

Table 5: Initial Ratios 20

Table 6: Distribution of Insolvent Companies by Year of Insolvency 22

Table 7: Model Estimation Data Sets - Difference in Means After PSM 1 to 1 and 1 to 10 24

Table 8: Data Sets Composition Before and After PSM 1 to 1 and PSM 1 to 10 24

Table 9: Model Estimation Data Sets - Number of Observations by Year of Financial Statement 25

Table 10: Model Estimation Data Sets - Distribution by Region 26

Table 11: Distribution of Insolvent Companies by Division and Section of Activity 27

Table 12: Distribution of Portuguese SMEs by Size 27

Table 13: Descriptive Statistics – After PSM 1 to 1 28

Table 14: Test of Equality of Means Between Active and Insolvent Companies – After PSM 1 to 1 29

Table 15: Test of Equality of Means Between Active and Insolvent Companies - After PSM 1 to 10 30

Table 16: Pearson`s Correlation Coefficients for the Financial Ratios After PSM 1 to 1 31

Table 17: Pearson`s Correlation Coefficients for the Financial Ratios After PSM 1 to 10 32

Table 18: Coefficients Estimates for Model 1 and Model 2 37

Table 19: GOF - Pearson`s Chi-square Test 40

Table 20: GOF - Hosmer-Lemeshow Test 40

Table 21: Predictive Accuracy 42

Table 22: Number of Observations/ Companies Used for Testing Forecast Accuracy 43

Table 23: Forecast Accuracy for Models 1 and 2 (%) 44

List of Figures

Figure 1: Distribution of Insolvent Companies by Region 26

Figure 2: Area Under Curve – Model 1 (PSM 1 to 1) 39

Figure 3: Area Under Curve – Model 1 (PSM 1 to 10) 39

Figure 4: Model 1 (PSM 1 to 1) Estimation – Sensitivity and Specificity vs Probability Cutoff 41

Figure 5: Model 2 (PSM 1 to 10) Estimation – Sensitivity and Specificity vs Probability Cutoff 42

1 Introduction

The present study is part of a research project collaboration between the School of Economics and Management of the University of Minho and nBanks company, a FinTech enterprise that aims to help companies better manage their financial operations and processes and intermediate between them and financial institutions.

Financial distress and business failure prediction are considered essential issues in economics and finance. Their importance was made even more evident by the 2007-2009 global financial crisis, one of the most devastating of its kind, that brought many businesses to the brink of collapse and caused the loss of a considerable amount of jobs and income. At present, the world is affected by the coronavirus pandemic, which is showing considerable impact over the economic activity.

Business failure prediction is important for all involved: owners, stakeholders, managers, employees, financial institutions, government and society in general. Early prediction can help take action to prevent failure as well as provide a measure for financial institutions with regard to which companies are eligible for credit granting. In the actual globalised environment, with world economy more and more interconnected, "failure prediction is a field of world-wide interest" (Dimitras, Zanakis & Zopoundis, 1996, p. 491).

In Portugal, small and medium-sized enterprises (SMEs) represent 99.9% out of the total number of companies, are providers of the main source of employment and generate over half of the total value added. Given the importance of SMEs, it could prove extremely beneficial to be able to forecast bankruptcy and prevent financial distress in due time to allow for measures that might restore equilibrium to the financial situation of these companies that can be considered pillars of Portuguese economic growth.

One way to forecast financial distress is by developing a business failure prediction model which is a classification model that aims to distinguish between firms in distress and firms in normal active business operating conditions.

The host of this project, nBanks, is a company that presents an innovative business model that proposes a platform for optimal interaction between clients and financial institutions. With the motto "Freedom of Choice", nBanks aims to set the basis of a world-wide new banking relationship model, reinforce global banking literacy and help to improve banking related decisions made by their clients (NBanks, 2019).

In order to attain this goal, nBanks company has the purpose of acting as a bridge facilitating the encounter between the need for financial products and the appropriate offer that satisfies that need. For both needs to be met, a deep understanding of the client`s characteristics, circumstances and history is required to set a foundation for the search for a concrete financial product or approach that will prove the most suitable for each specific financial circumstance.

On the offer side, financial institutions are endowed with teams of specialists that create products for different profile customers and that perform the complex analyses required in the process. On the customer side, the companies, for the great majority SMEs, do not possess the same ease of understanding in relation to the financial products and many times are not fully prepared to present and explain the exact situation of need or the characteristics of the company in a way that meets the criteria of the financial institution.

Therefore the role of nBanks is to bridge that gap in a mutually comprehensive way, and one of the tools that may be of use would be a financial distress prediction model, with the aid of which it could build a classification system that indicates on a scale, the degree of financial health of each of their customers and help the SMEs to better understand the degree of their default risk. This system would serve companies and financial institutions alike, creating a common reference for measuring financial health, predicting and eventually helping to prevent situations of moderate to extreme financial distress and default.

There are many financial distress prediction models that have been developed around the world, such as Altman`s (1968) Z Score, also Altman`s (1983) Z` and Z`` Score, Ohlson`s (1980) O Score, Zmijewski`s (1984) probit model, among others. This study`s contribution is to develop a model based on information that is more specific to the reality of the Portuguese SMEs and from a more recent period – the last 10 years available.

The methodology is the logistic regression analysis model, which employs maximum likelihood estimation. The Propensity Score Matching method will also be applied. This method allows to pair, in the same proportion or to different proportions, treated (declared insolvent) companies and control (healthy, active) companies, based on common, comparable characteristics such as industry and size, per same year.

The data set used is from AMADEUS database, published by Bureau Van Dijk/ Moody`s Analytics. It comprises data on SMEs of Portugal for the years 2010 – 2018, from all business sectors except the financial sector.

Two models were estimated and then validation tests were performed on data sets of one, two, three years prior insolvency, of all three years combined prior to insolvency and on all data available, which comprise one up to six years prior insolvency.

The results show a predictive accuracy of above 80%. The use of the propensity score matching made the two groups of companies more homogenous and comparable in terms of size, industry and year, and reduced the unbalance in numbers of insolvent and active/ healthy companies. These models could be additionally improved by the addition of macroeconomic variables, or by performing tests and estimations on companies by sector of activity which would render the financial information even more comparable and contribute to the predictive power.

This study is divided into six sections. Section 2 is dedicated to the presentation of the company that supported this project. Section 3 presents the theoretical framework based on which this study was developed, with emphasis on the main existing models. Section 4 describes the methodology used in terms of statistical model estimation and validation and also introduces the database from where information was sourced, the data set used, its characteristics and selection standards. Section 5 is dedicated to the presentation of results and the respective analysis of the estimation and validation process. Section 6 briefly presents the main conclusions reached by the analysis of the results obtained.

2 Presentation of NBanks Company

One important part of financial transactions is the evaluation of business performance. Financial institutions have demands that are not always easily met by businesses. In this context, *fintech* company nBanks aims to bridge up the gap between financial institutions and their customers, in this case SMEs from Portugal and other South European countries.

The company started in officially in September of 2018 and offers products and services that penetrate and integrate areas of the financial system with the aim of changing the business landscape in the financial area. Some of the them are:

- Consolidated information that comprises bank account details, transactions, business associates, administration functions, documents, etc., for easier and faster processing. This integration of all necessary information allows for faster processing and more precise analysis of company`s business performance and tax compliance and administration. This standardised processing of information is applied to all customer companies, thus creating patterns that enable a more efficient processing of this information and even getting on the brink of predicting possible future outcomes for each company.
- Intelligent product search, which is a consolidated search engine that offers access to the descriptions of various financial products available on the market for the companies (such as short/ long-term loans, investments, leasing etc.), enabling the company to select the best product, contact the respective financial institution and start negotiations.
- Platform integration with IRB – *Índice de Risco Bancário* (Bank Risk Index), where customer companiescan make a simulation of the way financial institutions evaluate their business performance based on financial statement information.
- Communication hub available for interaction between the company and its accounting services, which makes possible real-time access to business partners and transactions information, eliminating the time lag needed for e-mail and other such communication that many times delays the accounting processing.

Communication system with the banks, through which the financial institutions can know more about a customer company that is willing to acquire a certain financial product. For example, the bank can have access to the IRB – Bank Risk Index - and thus better understand at a glance the profile of the company, which makes the whole process faster. This helps both parties to save time and offers to the bank a more independent evaluation of the customer company, once this evaluation is not done by the company itself but by nBanks.

In this context, the estimation a financial distress prediction model based on recent data and on the reality of Portuguese SMEs would provide an useful tool for nBanks to apply in practice in order to attend the necessities of their customer companies.

3 Theoretical Framework

3.1 SME Definition and Insolvency

3.1.1 SME Definition

A SME, as defined by the Decree-Law 81/2017 issued by the Portuguese Government, in accordance with the European Union Commission Recommendation 2003/361 of 6 May 2003[1], definition also adopted by the Portuguese Institute of Support to Small and Medium-sized Enterprises and Innovation (IAPMEI), is an enterprise that employs fewer than 250 persons, has an annual turnover not exceeding EUR 50 million and/or an annual balance sheet total not exceeding EUR 43 million.

3.1.2 Insolvency

In research and in practice alike it is difficult to define insolvency and what exactly separates it from bankruptcy and many definitions of default or financial failure also exist. As mentioned by Ohlson (1980, p. 111), "there is no consensus on what constitutes `failure`".

Armour (2001, p. 3), starting from the commonly accepted sense of the word "*insolvency*" which is an inability to pay creditors, tries to establish a distinction between six different meanings of this term which are: the accounting concept of *balance sheet insolvency*, *cash flow insolvency* (or "financial distress"), *economic failure* (or "economic distress"), and the judicial concepts of *liquidation*, *reorganisation* and *insolvency proceedings* (or "bankruptcy").

The distinctions are specified by Armour (2001) as follows. Balance sheet insolvency means that the book value of its assets is exceeded by that of its liabilities. Cash flow insolvency means a firm is unable to pay its obligations as scheduled. The expression "financial distress" is commonly used to refer to a company which has difficulty in paying its creditors, while "economic distress" alludes to a lack of economic viability. The last is related to financial distress by the fact that "all firms which are economically distressed will also become financially distressed" (Armour, 2001, p. 4).

The term *liquidation* refers to one of the possible outcomes of financial distress and means "the conversion into cash, through sale, of a firm`s assets" (Armour, 2001, p.4), and while it can also happen under administrative receivership, it is a necessary part of the closing proceedings. *Insolvency* is a condition, and *liquidation* is an event (Armour, 2001).

Altman (1983) sums the generic terms which refer to unsuccessful business enterprises to three: *failure*, *insolvency* and *bankruptcy*.

Failure, "by economic criteria, means that the realized rate of return on invested capital (...) is significantly and continually lower than prevailing rates on similar investments." (Altman, 1983, p. 6), but this does not imply the discontinuance of the entity. When the company can no longer meet the legally enforceable demands of its creditors it enters *legal failure* (although this may happen without formal legal action involved). *Business failures* (as also used by Dun & Bradstreet) include businesses that cease operation following bankruptcy or after loss to creditors after execution, foreclosure or attachment, that voluntarily compromise with creditors, or voluntarily withdraw leaving unpaid obligations.

Insolvency is used technically to mean a lack of liquidity resulting in the firm not being able to meet its current obligations and indicates "a chronic rather than a temporary condition" (Altman, 1983, p. 6), and the real net worth of the firm is negative.

Bankruptcy is described by Altman (1983) as being of two types: one in which the net worth of the firm is negative and another where there is a formal declaration of bankruptcy in court, together with a petition to either liquidate its assets or try recovery.

Portugal`s *Insolvency and Business Recovery Code* (CIRE) regulates proceedings regarding insolvency and business recuperation processes. It states, in Article 3, that enterprises are considered in a state of insolvency when the book value of its liabilities surpasses the book value of its assets. It also states, in Article 7, that insolvency is not the same thing as bankruptcy since the impossibility of paying as scheduled does not automatically imply that the company is no longer economically viable or that it cannot recover from a financial point of view.

This project, due to data availability, will abide by the definitions provided by the database from which the data were sourced, AMADEUS. Company status definitions are as follows.

Active = the company is active. The control group companies used in this study belong to this category.

Active (insolvency proceedings) – the company is declared insolvent and although remaining active it is in administration or receivership or under a scheme of arrangement, placed under the protection of the law and continues operating and repaying creditors and tries to reorganise and return to normal operating. At the end, the company will either return to normal operating or will be reorganized or will be liquidated. The insolvent companies used in this study only include companies from this group that did not return to normal operating and were finally characterised as "in liquidation" at the end of the process.

In liquidation = the company is in the process of liquidation and its assets are being sold. The next step will be that the company is dissolved and will no longer exist. In some cases the need for liquidation proceedings stems from the need of self-addressing creditor problems, since when an insolvent`s assets are insufficient to meet the claims of all creditors it will be in the creditor`s best interest to try and recover its claim before other creditors can do the same.

The insolvency of a company has various causes. Table 1 presents some of the elements that may result in a state of insolvency, which can be divided into internal and external causes. Internal causes are related to the management of the company, such as liquidity problems, poor management, lack of quality of the product, fraud, among others. Liquidity problems due to lack of finance are closely related to the subject of this study since many SMEs face this type of problem, and this project is part of the nBanks company attempts to help with this issue by making easier for their SMEs customers to adequately present their situation to financial institutions in order to obtain the necessary funds. External causes are macroeconomic situations brought on by the environment outside the company, among which are harsher competition, economic situation difficulties, bad debt, natural disaster and so on (Kucher, Mayr, Mitter, Duller & Feldbauer-Durstmuller, 2018).

Table 1: Causes of Insolvency

Internal causes	Liquidity problems due to lack of finance
	Poor business-economic competences
	Unqualified management
	High cost pressure
	Poor quality of goods or services
	Conflicts between managers or owners
	Fraud
External causes	Competition increase, price fights
	Economic slowdown
	Bad debt
	Natural disasters

Source: adapted from Kucher et al. (2018)

3.2 Literature Review

There is a vast literature on default prediction. Over time there have been developed several financial distress prediction models, employing different techniques. These models aim to predict the likelihood of business failure of firms, based on a selection of most relevant financial ratios that reflect the companies` financial health and probability of default.

3.2.1 Univariate Models

First statistical models used univariate analysis for selected ratios, with notable contributions from Beaver, who introduced a technique that permitted classification of firms into healthy and failing, by using "financial ratios as predictors of important events – one of which is the failure of the firm" (Beaver, 1966, p. 72). Univariate models are based on the analysis of the financial ratios in isolation and comparing their values between financially distressed companies and healthy ones, in order to differentiate them.

The sample used by Beaver (1966) comprised of 79 failed companies and 79 non-failed ones, with financial statements of the failed companies obtained for five years prior to failure. The data set extended between the years 1954 to 1964, 10 years. For analysis were tested 30 ratios, divided into 6 categories. From each of these categories the ratio with the highest discriminating power was selected, with the following results:

- Cash flow to total debt;
- Net income to total assets;
- Total debt to total assets;
- Working capital to total assets;
- Current ratio;
- No-credit interval.

The ratios for the companies were classified in ascending order and an optimal cutoff point was set for each given ratio, in order to minimise incorrect predictions, then the percentage of misclassification was calculated. Beaver`s conclusion was that the strongest ability to predict failure was in the Cash Flow to Total Debt ratio, with failures of only 13% in the first year and 22% in the fifth.

As further development, Beaver suggested a multi-ratio analysis that "would predict even better than the single ratios" (Beaver, 1966, p.100).

3.2.2 Multivariate Models

Default risk forecasting models thus evolved to multivariate studies, the most notable being the study by Altman (1968), in which an Multiple Discriminant Analysis (MDA) model, called the Z-Score model, was developed.

MDA is "a statistical technique used to classify an observation into one of several a priori groups, dependent on the observation`s individual characteristics (...), data are collected for the objects in the groups; MDA then attempts to derive a linear combination of these characteristics which `best` discriminates between groups" (Altman, 1968, pp. 591-592).

The original Z-Score model is a model aiming to forecast bankruptcy of manufacturing firms, which was developed on a sample of 66 United States companies divided into two groups of 33 failed and 33 non-failed firms, using the estimation of a linear combination of five variables, with the final discriminant function being as follows:

$$Z = 1.2 \cdot X_1 + 1.4 \cdot X_2 + 3.3 \cdot X_3 + 0.6 \cdot X_4 + 1.0 \cdot X_5$$

where $X_1 =$ Working Capital/ Total Assets

$X_2 =$ Retained Earnings/ Total Assets

$X_3 =$ EBIT/ Total Assets

$X_4 =$ Market Value of Equity/ Book Value of Total Liabilities

$X_5 =$ Sales/ Total Assets

X1 – Working Capital/Total Assets measures a company`s net liquid assets relative to total capitalisation. For company having consistent losses current assets will be diminishing in relation to its total assets, and this leads to a decreasing working capital. Altman concluded that this ratio was the most valuable of the liquidity ratios evaluated.

X2 – Retained Earnings/Total Assets is a measure of cumulative profitability over time and implicitly reflects the age of the firm, since a relatively young firm would not have had the time to make this kind of reserve.

X3 – EBIT/Total Assets is a measure of the real productivity of the assets of a company, eliminating any tax or leverage effects. This ratio is important because insolvency happens when a company`s total liabilities exceed its total assets.

X4 – Market Value of Equity/Book Value of Total Debt shows how much the company`s assets can decline in value before liabilities exceed assets and the company becomes insolvent. Altman found this ratio to be a more effective predictor than the more commonly used Net Worth/Total Debt ratio.

X5 – Sales/Total Assets, the capital turnover ratio, illustrates the sales generating ability of the company`s assets and measures the capability of management to deal with competition. Even though this ratio presented a very low F value, in the final model it ranked second in discriminating ability due to its relationship to the other variables in the model.

Table 2 presents the results of the F test. The higher the F ratio the better the predictive ability of the respective financial ratio.

Table 2: Altman`s Z-Score Model - Variable Means and Test of Significance

Variable	**Bankrupt Group Mean** n = 33	**Non-Bankrupt Group Mean** n = 33	**F Ratio**
X1	-6.1%	41.4%	32.6*
X2	-62.6%	35.5%	58.86*
X3	-31.8%	15.3%	26.56*
X4	40.1%	247.7%	33.26*
X5	150.0%	190.0%	2.84
Source: Adapted from Altman (1968)			

*significant at the .001 level

The interpretation of the Z-score results is as follows:

a) Z > 2.99 – safe zone (non-failed company);
b) 1.80 < Z < 2.99 – grey zone (uncertainty);
c) Z < 1.80 – danger zone (failed company).

Table 3 shows the predictive accuracy of Altman`s initial model.

Table 3: Altman`s Z-Score Model Predictive Accuracy

Years prior to insolvency	Number of Observations	Hits	Misses	Predictive accuracy
1	33	31	2	95%
2	32	23	9	72%
3	29	14	15	48%
4	28	8	20	29%
5	25	9	16	36%
Source: Adapted from Altman (1968)				

Altman, Haldeman and Narayan (1977) developed a new model based on the Z-Score, called the ZETA model, in collaboration with Zeta Services Inc., due to which the final formula is not publicly available. This new model was adapted to the new reality that included large companies. The data set used was comprised of 53 insolvent and 58 healthy U.S. companies, from years between 1969 and 1975, and started with 27 variables. The final ratios included in the model are:

- X1 – EBIT/Total Assets;
- X2 – Standard error of estimate around a ten-year trend in X1;
- X3 – log (EBIT/Total interest payments);
- X4 – Retained Earnings/Total Assets;
- X5 – Current Assets/Current Liabilities;
- X6 – Common Equity/Total Capital;
- X7 – log (Total Assets).

The predictive capacity of this model surpassed that of the original Z-Score, with 90% hit rate for one year prior to insolvency and about 70% up to five years ahead of insolvency. The discrimination of companies into the categories of insolvent or healthy is subject to two types of errors:

- Type I error is classifying an insolvent company as healthy. This is considered the costliest error, since this means that the model does not predict insolvency. Any investment based on this misclassification will be lost;
- Type II error is classifying a healthy company as insolvent. This misclassification would cause a missed investment opportunity and the loss implied would be only of the possible gains not received.

Accordingly, minimising Type I error is the most important, since this kind of misclassification is the most financially prejudicial.

Altman also developed extensions of the original Z-Score model, which was originally developed for U. S. publicly traded firms, based on market data.

The extensions developed are the Z`-Score model adapted for private companies, which also has five variables, and the $Z^{``}$ Score model, with four variables, adapted for non-manufacturers and emerging markets (Altman, 2002; Altman, Iwanicz-Drozdowska, Laitinen and Suvas, 2017).

$$Z^{`} = 0.717 \cdot X_1 + 0.847 \cdot X_2 + 3.107 \cdot X_3 + 0.420 \cdot X_4 + 0.998 \cdot X_5$$

where $X_4 =$ Book Value of Equity/ Book Value of Total Liabilities and other variables the same as those in the original.

The new model estimation had a small change in the cutoff value so that the interpretation of this new score is as follows:

a) Z > 2.90 – safe zone (non-failed company);
b) 1.23 < Z < 2.90 – grey zone (uncertainty);
c) Z < 1.23 – danger zone (failed company).

$$Z^{``} = 3.25 + 6.56 \cdot X_1 + 3.26 \cdot X_2 + 6.72 \cdot X_3 + 1.05 \cdot X_4$$

The interpretation of the $Z^{``}$Score model is as follows:

a) Z ≥ 1.10 – safe zone (non-failed company);
b) Z < 1.10 – distressed condition.

Altman further enhanced and improved this model, with re-estimations also considering Basel II[1] environment (Altman, 2002).

In an international context, Altman et al. (2017) analysed the performance of the $Z^{``}$Score model using a sample of firms from 31 European and three non-European countries, for private and public,

[1] Basel II is a set of international regulations by the Basel Committee on Bank Supervision which introduced capital requirements for financial institutions, such as the minimum capital to be maintained in a percentage based on risk-weighted assets (see Basel, 2001).

manufacturing and non-manufacturing firms, adapting the original model by making the necessary modifications. The study chooses to focus on the accounting - based versions of the Z-Score model even though it is occasionally outperformed by other models, since these versions do not rely on market data which are not usually available for privately held firms, which are the most common firms operating in business. The study used extensive international data and trying to assess the effects of different factors – year of bankruptcy, size of firms, age of firms, industry and country - on the predictive performance of the model. The results show that the $Z^{``}$ Score model performs very satisfactorily in an international context. Nevertheless, the study reached the conclusion that it is "possible to extract a more efficient country model for most (...) countries using the four original variables accompanied by a set of additional background variables" (Altman et al, 2017, p. 167).

The MDA technique developed by Altman inspired many authors to produce works in this line of research, authors such as Deakin (1972). One common idea behind their research is that a company has more chances to fail if it is unprofitable, highly leveraged and has difficulties with cash flow.

One critique to the studies based on MDA is that, as according to Altman and Sabato (2007), most of them violate two basic assumptions of this technique: that the independent variables are normally distributed and that the variance=covariance matrices of the failing and of the non-failing group are equal.

Altman`s Z Score is a seminal model, which has been often applied and adjusted to other countries or to other type of samples. For example, Taffler (1984) proposed a Z-Score model for the United Kingdom, Xu and Zhang (2009) adapted the model for Japan, Tinoco and Wilson (2013) adapted again for the U.K., Singh and Singla (2019) did a re-estimation for India.

3.2.3 Linear Probability Models

For many years, MDA was the mainly used statistical technique for failure prediction models. However, as mentioned above, it is subject to the critique that two of its basic restrictive assumptions, namely that the independent variables are normally distributed, and that the variance-covariance matrices are equal across the insolvent and healthy groups, end up being often violated in practice.

3.2.3.1 Logistic Regression Analysis (LRA)

LRA or logit analysis uses the non-linear maximum log-likelihood technique to estimate the probability of firm failure under the assumption of a logistic distribution.

Logit analysis was proposed for the prediction of business failure by Ohlson (1980), with the goal of predicting company bankruptcy up to three years before actual failure. He was a pioneer of using logistic regression for business failure prediction. The model developed by Ohlson (1980), the O-Score, was based on a data set composed of 105 bankrupt companies and 2,058 non-bankrupt companies over the period between 1970 and 1976. The model considers nine variables, seven financial ratios and two binary variables, selected based on the fact they were the most mentioned in literature.

In Ohlson`s (1980) logit model:

Logistic function = Probability of firm failure = $\frac{1}{1+e^{-Z}}$

where $Z = \alpha + \beta_1 SIZE + \beta_2 \frac{TL}{TA} + \beta_3 \frac{WC}{TA} + \beta_4 \frac{CL}{CA} + \beta_5 OENEG$

$$+\beta_6 \frac{NI}{TA} + \beta_7 FUTL + \beta_8 INTWO + \beta_9 CHIN$$

and where SIZE is the natural logarithm of GDP-deflated total assets; TL/TA is total liabilities divided by total assets; WC/TA is working capital divided by total assets; CL/CA is current liabilities divided by current assets; OENEG is a dummy variable equal to one if total liabilities exceed total assets, and zero otherwise; NI/TA is net income divided by total assets; FUTL is funds from operations (pre-tax income plus depreciation and amortization) divided by total liabilities; INTWO is a dummy variable equal to one if net income was negative over previous two years, and zero otherwise; and CHIN is the scaled change in net income calculated as (NIt-NIt-1)/(|NIt|+|NIt-1|), where NIt is the net income for the most recent period.

In his study Ohlson (1980) estimates three different models to capture the probability of insolvency over different periods: one year, two years and three years before failure. The predictive capacity for all three models was over 90% (see Table 4).

Table 4: Predictive Accuracy of Ohlson`s Models

Models	Predictive accuracy
1	96.12%
2	95.55%
3	92.84%

Source: Adapted from Ohlson (1980)

Other authors followed to apply the same methodology. Zavgren (1985) criticised Ohlson`s model for having a rather weak theoretical basis as well as for not having a balanced sample of companies. The author used the logit method to develop and test a new bankruptcy prevision model able to identify signals and estimate the probability of insolvency five years before the fact. The sample used by Zavgren (1985) was composed of 45 insolvent and 45 healthy companies, with data from years 1972 to 1978 and estimated statistically significant models for each of the five years before insolvency. Among the findings of the study are the fact that efficiency ratios are more relevant in the long term, liquidity ratios showed that insolvent companies were more concerned over liquidity than over investment opportunities one year prior failure, debt ratios were found relevant but profitability ratios were not found significant in order to discriminate insolvent companies from healthy ones.

3.2.3.2 Probit Analysis

Probit models are similar to logit models. The main difference is that for the calculation of the probability for a binary dependent variable probit regression uses the cumulative normal distribution function. Maximum likelihood estimation is employed in the same way as in logit analysis.

Studies that use probit analysis are much fewer than those using logit, possibly because more computational power is needed because it involves non-linear estimation (Gloubos & Grammatikos, 1988). One of the most representative study is Zmijewski (1984), who estimated a probit model analysing data of 40 bankrupt and 800 non-bankrupt companies for a period from 1972 to 1978, using only three independent variables: ROA (Net Income to Total Assets), FINL (Financial Leverage = Total Debt to Total Assets) and LIQ (Liquidity = Current Assets to Current Liabilities).

3.2.4 Other Models

Other models developed are the Artificial Intelligence models, which are based on the primary assumption that data can be incomplete and can change over time and take these changes into account in a dynamic manner, being often described as learning systems. Some approaches are genetic algorithms (GAs), which are applications of biologically inspired algorithms, and artificial neural networks (ANNs) (Bisogno, Restaino & Di Carlo, 2018; Balcaen & Ooghe, 2006).

There is also a rapidly growing type of models which are the contingent claim models, based on the option pricing theory of Black and Scholes (1974) and Merton (1973), also hazard models, support vector machines, rough sets, decision tree models and others (Alaka et al., 2018).

There have been made many reviews of the existing models. For example, Agarwal and Taffler (2008) have found that the predictive accuracy of accounting-based models and market-based models is similar, but also arrive to the conclusion that, although market-based models are conceptually attractive, rather lack empirical superior performance, while the accounting-based models, although criticised for lack of theoretical basis, are able to correctly capture the years of poor corporate performance that precedes failure. Also, according to Agarwal and Taffler (2008, p. 1550), “the double entry system of accounting ensures that window dressing the accounts or change in accounting policies will have minimal effect on a measure that combines different facets of accounting information simultaneously”.

4 Methodology and Data

4.1 Objectives

The objective of this project is to estimate a logit model of financial distress prediction appliable to Portuguese SMEs, starting from several financial ratios that are considered relevant in the literature and trying to identify the most relevant among them in the context of the data set used and calculating the coefficients to be used in order to obtain a classifying score. This financial health scoring model might help SMEs to obtain better knowledge of their own default risk, which can help in early detection and prevention of problems and also result in a better and easier relationship with financial institutions.

In this project we use the logistic regression (LR) approach to identify the relevant control variables for a logit model adapted to the Portuguese context of the SMEs. These variables will be used as predictors for default.

According to Lacerda and Moro (2008), the logit model is widely used and practical because its score is calibrated as probability of default. Also, logit methodology does not require the restrictive assumptions of MDA: it allows working with a disproportional sample and also does not require multivariate normal distribution of the data (Altman & Sabato, 2007).

The research consist of three parts: sample selection and data collection, selection of methods and specific variables (ratios) to obtain a prediction model, and model validation through statistical significance and results accuracy.

4.2 Methodology and Variables

As previously mentioned, the scoring model will be based on a logistic regression model. The dependent variable is binary and represents the status of the company, coded as $Y = 1$ for distressed companies and $Y = 0$ for healthy, active companies. The logit model evaluates the probability of financial distress in function of one or more independent variables, using a maximum likelihood estimator. The model produces a score between zero and one which represents the probability of financial distress for the respective company.

The balance sheet data used is from one year previous to the year of default (e.g. Butera & Faff, 2006, Altman & Sabato, 2007).

In the logit model the relationship between the probability of financial distress (p) and the independent variables is a S curve ranging from 0 to 1. The independent variables are quantitative and are financial ratios. The reason lies with the assumption that the values of the financial ratios deteriorate as a company starts to approach insolvency status.

How to predict financial distress based on ratios? According to Altman (1968, p.591), "the question becomes, which ratios are more important in detecting bankruptcy potential, what weights should be attached to those selected ratios, and how should the weights be objectively established." According to Dimitras et al. (1996), the financial ratios that can predict failure are different depending on country, sector and period of time. The aim of this study is to identify what are the most important ratios in the context of Portuguese SMEs and objectively establish the weights that should be attached to them in the model in order to have an accurate level of financial distress prediction. To do so, we shall consider financial variables such as liquidity, profitability, activity, solvency, leverage ratios that may prove to be relevant to the SMEs in the specific context of Portugal.

Studies such as Lehmann (2003) indicate that using not only quantitative variables but including qualitative variables (such as location, existence of export activity, etc.) is better for predicting default in SMEs. Nevertheless, this study will not include qualitative variables because of the following reasons. First, it is aimed at constructing initially a prediction model for the general reality of Portuguese SMEs (an extension considering separate analysis by industry, for example, being left as a possibility for further research). Secondly, although this project initially contemplating the analysis including a qualitative variable that indicated the existence of export activity within companies, the AMADEUS database does not contain this information.

In congruence with previous studies, this study uses ratios from five categories: profitability, liquidity, solvency, leverage, and activity. Profitability is expected to be of key importance in discriminating probability of financial distress since "a firm with poor profitability (...) may be regarded as a potential bankrupt." (Altman, 1968, p.591) and profitability is negatively related to credit risk (Doumpos, Kosmidou, Baourakis & Zopounidis, 2002). Liquidity is also an important determinant. Companies with good liquidity positions are more capable of meeting the obligations to their creditors (Doumpos et al., 2002). Liquidity is also negatively related to credit risk. Solvency ratios measure the capacity of a company to generate internal funds (Canovas & Solano, 2006). Leverage ratios are widely analysed as classic indicators of financial risk, high values increasing the probability of default (Lacerda & Moro, 2008). Activity ratios measure the effectiveness with which an asset contributes to the profitability of investment in that asset category (Butera & Faff, 2006).

For each ratio category we have selected a number of financial ratios among those found relevant in most studies, as presented in Table 5, and we tested the various ratios in order to select those most potentially able to integrate the estimated model. In total, 19 ratios were selected.

Table 5: Initial Ratios

Liquidity	X1	Current Ratio	Current Assets / Current Liabilities
	X2	Working Capital to Total Assets	Working Capital / Total Assets
	X3	Quick Ratio	(Cash + Accounts Receivable) / Current Liabilities
	X4	Cash Ratio	Cash / Current Liabilities
Profitability	X5	Current Assets to Total Assets	Current Assets / Total Assets
	X6	EBIT to Total Assets	EBIT / Total Assets
	X7	Operating Cash Flow to Total Assets	Cash Flow / Total Assets
	X8	Operating Profit Margin	EBIT / Operating Revenue
	X9	ROA	Net Income / Total Assets
Leverage	X10	Debt to Equity	Total Liabilities / Shareholders` Funds
	X11	Debt to EBITDA	Total Liabilities / EBITDA
	X12	Operating Cash Flow to Debt	Cash Flow / Total Liabilities
	X13	Retained Earnings to Total Assets	(Other Shareholders`Funds + Net Income) / Total Assets
	X14	Debt to Asset	Total Liabilities / Total Assets
Solvency	X15	Interest Coverage	EBIT / Interest Paid
	X16	EBITDA to Interest Coverage	EBITDA / Interest Paid
	X17	Equity to Debt	Shareholders`Funds / Total Liabilities
Activity	X18	Total Assets Turnover	Operating Revenue / Total Assets
	X19	Working Capital Turnover	Operating Revenue / Working Capital

Source: Author

4.3 Data Set

Historical accounting and financial data is collected from AMADEUS, a database published by Bureau van Dijk /Moody`s Analytics, which contains financial and business information on over 21 million European companies, providing standardised annual accounts, financial ratios, sectoral activities and ownership data, with up to ten years archive. This study uses a data set for Portuguese SMEs for the last ten years available, between 2010 and 2018.

As previously stated, the financially distressed group consists of companies with the following statuses, standardized by AMADEUS:

"*In liquidation*" - which means the end of the firm`s activity. This category in AMADEUS includes voluntary liquidation and dissolution but there is no indication in the database to distinguish between voluntary and compulsory.

"*Active (insolvency proceedings)*" – from this category only companies that did not return to normal operating status and were subsequently characterized as "*in liquidation*" were selected.

The control group comprises companies registered in AMADEUS as "*Active*".

From this first sample we selected only the SMEs from Portugal that comply with the following criteria:

- Companies from all activity sectors except activity codes NACE 64, 65, 66, 68, corresponding to financial and real estate activities;
- Unlisted companies.

This initial sample comprised 281,925 Portuguese enterprises, out of which 5,479 insolvent and 276,446 active companies.

Subsequently the following filters were applied, in order to select only:

- Companies with year of incorporation up to and including 2016, thus ensuring a minimum of three years of activity, since during the first years of their lives young and healthy companies often show a financial structure similar to failing companies (du Jardin, 2010);
- Companies with all the accounting information needed to calculate all 19 ratios considered in the first selection of independent variables for the model;
- Companies attending criteria for SMEs: less than 250 employees, and less than 50 million EUR turnover/or less than 43 million EUR total assets;
- In order to eliminate the very small firms, since those tend to present gaps and potential distorted values, we also eliminated companies with less than 100,000 EUR total assets (Altman, 2017, Balcaen and Ooghe, 2006);
- Finally, the data were winsorized at the 1% and 99% levels to eliminate outliers.

After this second selection, the dataset comprises 65,997 companies, out of which 1,504 insolvent and 64,493 active companies.

Table 6 shows the distribution of the insolvent companies by year of insolvency.

Table 6: Distribution of Insolvent Companies by Year of Insolvency

Insolvency Year	Companies
2014	68
2015	171
2016	221
2017	326
2018	336
2019	382
Total	1,504
Source: Author	

In face of this final sample being quite disproportional considering the number of insolvent and active companies a method was needed in order to obtain a more balanced/ homogenous sample. In order to do so this study applied the propensity score matching method (PSM), where "matching is a method of sampling from a large reservoir of potential controls to produce a control group of modest size in which the distribution of covariates is similar to the distribution in the treated group." (Rosenbaum & Rubin, 1983, p. 48). This matching technique was developed by Rosenbaum and Rubin (1983), and its purpose is to find, for every individual in the treatment group (in our case, the insolvent companies), a statistical twin that possesses similar characteristics in the non-treated/ control group (in our case, the active companies), so that the sample can be considered randomly selected and direct comparisons be more meaningful. This is important because if the individuals of the treatment and the control group are not randomly selected the sample runs the risk of suffering from selection bias.

The common support condition makes sure that the propensity scores of both groups overlap and all participants have a counterpart in the control group, which means that only firms that are sufficiently alike each other are matched.

The covariates for the matching are assumed as not affected by the treatment, either pre or post treatment. The covariates used in this study are industry (NACE level 2 - division), the year of the financial statement (which in case of the insolvent companies is one year prior to insolvency) and size, for which the logarithm of total assets was used as proxy, to control for the size effect and allow comparisons of ratios (du Jardin, 2010).

We performed PSM selecting the nearest neighbour with replacement, which allows for a control firm to be used more than once as a match. This helps to decrease bias since control firms similar to several treated firms can be used multiple times as needed. In order to ensure a better matching quality, we have set the maximum permitted difference between matched individuals (caliper) to 0.25 of the propensity score standard deviation, following Cochran and Rubin (2004). This also reduces the number of matches that can be performed, but it does not negatively affect this study due to the large size of the data set available.

The first step in PSM is calculating the propensity score, in order to assign to each insolvent company a similar active one. For this the sample is split into five sets of intervals and tested separately to asses if the balancing properties are satisfied, which means that there are no significant statistical differences between the two groups regarding the distribution of covariates (Dehejia & Wahba, 1999). The propensity score is then calculated by a probit model:

$$prob(D_i = 1) = \alpha_i + \varphi Z_{i;t-1} + \varepsilon_{it}$$

where D_i is a dummy variable with the value of 1 if the company is insolvent and 0 otherwise and Z the set of control variables.

The one-to-one propensity score matching selects for each distressed company an active company with the nearest distance to the distressed one as indicated by the propensity score. The financial variables used for the estimation of the model, for the insolvent companies, are those from the year prior to insolvency (N-1). Appendix 1 shows the kernel density plot of the propensity score.

In order to mitigate for the limitation of the propensity score matching procedure, besides one-to-one matching we also used another criterion which is one-to-many matching and performed the below tests on these matched data sets as well. The one-to-many PSM selects a specified number of active companies within the nearest distance to the insolvent one, the maximum distance (caliper) being fixed at 25% of the standard deviation of the propensity score, computed following Cochran and Rubin (2004). In this study we performed two selections: one-to-five (1 to 5) which matches 1 insolvent company to 5 active ones, and one-to-ten (1 to 10), which matches 1 insolvent company to 10 active ones. The performance of the model obtained from the 1 to 5 matching was very similar to that obtained from the 1 to 10 matching.

Since the results from the 1 to 10 matching data set were slightly better, this study will focus on reporting these.

After the propensity score matching, the difference between the means of the proxy used for size (logarithm of Total Assets) between active and insolvent companies is reduced and no longer statistically significant, which proves that the matching was successful, as shown in Table 7. Appendix 2 shows the results for PSM 1 to 5 matched data set.

Table 7: Model Estimation Data Sets - Difference in Means After PSM 1 to 1 and 1 to 10

Variable	Sample	Mean		%bias	t-test	
		Treated	Control		t	p>|t|
Log (Total Assets)	Unmatched	13.757	13.37	29.2	5.29	0.000
	Matched 1 to 1	13.757	13.697	4.5	0.54	0.589
	Matched 1 to 10	13.757	13.728	2.2	0.26	0.795
Source: Author						

Since the PSM method used was selecting the nearest neighbour with replacement, some of the active companies were used more than once, thus matching 1 to 1 returned 288 active companies with 289 observations, and matching 1 to 10, 2,108 active companies with 2,171 observations. Table 8 shows the composition of the data sets, in terms of number of companies and number of observations, before and after PSM 1 to 1 and PSM 1 to 10. Appendix 3 shows the composition of the data set after PSM 1 to 5.

Table 8: Data Sets Composition Before and After PSM 1 to 1 and PSM 1 to 10

Before PSM	**Insolvent**	**Active**
Companies	1,504	64,493
Observations	3,282	235,711
After PSM 1 to 1	**Insolvent**	**Active**
Companies	289	288
Observations	289	289
After PSM 1 to 10	**Insolvent**	**Active**
Companies	289	2,108
Observations	289	2,171
Source: Author		

Table 9 shows the distribution of the matched data sets PSM 1 to 1 and PSM 1 to 10 by year. Appendix 4 shows the distribution of the PSM 1 to 5 data set by year.

Table 9: Model Estimation Data Sets - Number of Observations by Year of Financial Statement

Year	Insolvent	PSM 1 to 1	PSM 1 to 10
2013	34	62	449
2014	27	27	189
2015	48	21	220
2016	60	46	354
2017	69	67	463
2018	51	66	496
Total	289	289	2,171
Source: Author			

This table presents the number of observations by year of financial statement. Column Year represents the year of the financial statement. Columnt Insolvent represents the number of observations for insolvent companies at one year prior to insolvency. PSM 1 to 1 represents the number of observations for active companies matched to the insolvent by PSM 1 to 1. PSM 1 to 10 represents the number of observations for active companies matched to the insolvent by PSM 1 to 10.

Table 10 shows the distribution of the matched data set by region of Portugal – NUTS II[2]. Insolvent companies are concentrated mainly in the North region and in Lisbon Metropolitan Area, with these two regions together with Central region accounting for over 85% of the insolvent companies (see Figure 1), which is representative of the distribution of all companies (active or not) over these regions, which is of 84%. Appendix 5 shows the distribution of the PSM 1 to 5 matched data set by region of Portugal.

[2] Nomenclature of territorial units for statistics (NUTS) is a common statistical classification of territorial units for harmonised regional statistics in the European Union (EU). NUTS classification divides each Member State of EU into NUTS level I territorial units, each of which is subdivided into level II units, these again subdivided into level III units (Regulation EC 1059, 2003).

Table 10: Model Estimation Data Sets - Distribution by Region

Region-NUTS II	Insolvent	PSM 1 to 1	PSM 1 to 10
PT11 - North	117	100	933
PT17 - Area Metropolitana de Lisboa	69	66	437
PT16 - Centro	61	72	536
PT18 - Alentejo	16	31	110
PT15 - Algarve	13	11	86
PT20 - Regiao Autonoma dos Acores	11	3	40
PT30 - Regiao Autonoma da Madeira	2	5	29
TOTAL	289	288	2,171
Source: Author			

This table presents the distribution of the model estimations data sets by region. Insolvent represents the number of insolvent companies per region, PSM 1 to 1 and PSM 1 to 10 represent the number of matching active companies selected by these respective methods, by region.

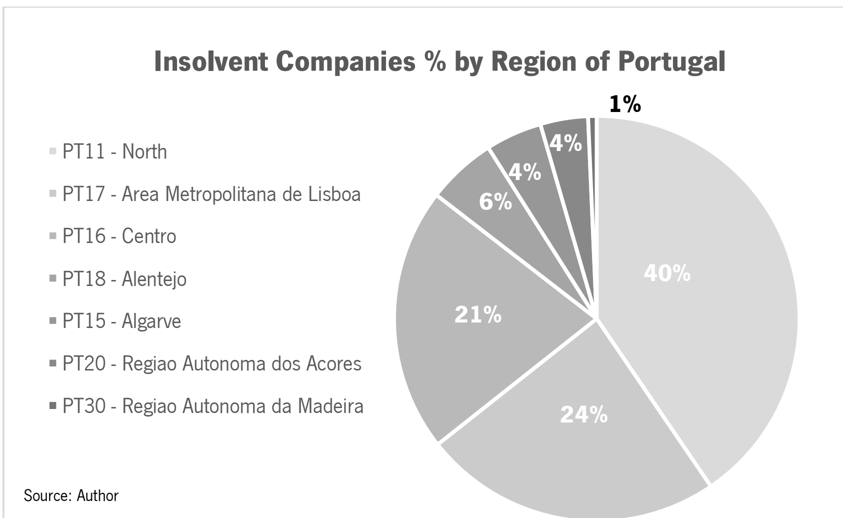


Figure 1: Distribution of Insolvent Companies by Region

According to EUROSTAT – NACE Rev. 2[3] statistical classification of economic activities in the European Community, the companies in the data set analysed belong to mainly the following activity divisions: G – Wholesale and retail trade; repair of motor vehicles and motorcycles, C – Manufacturing, F – Construction.

[3] NACE is the statistical classification of economic activities in the European Community. It consists of a hierarchical structure that contains a first level with sections identified by an alphabetical code (sections), a second level identified by a two-digit numerical code (divisions) and two more levels identifying groups and classes (Regulation (EC) 1893, 2006). In this study we use only the first two levels, sections and divisions, for identification.

These divisions represent over 80% of the insolvent companies. The distribution of insolvent companies per division/sections of activity is presented in Table 11.

Table 11: Distribution of Insolvent Companies by Division and Section of Activity

NACE Rev.2 - Division	Insolvent Companies	%
G- Wholesale and retail trade; repair of motor vehicles and motorcycles (45-47)	114	39%
C- Manufacturing (10-33)	69	24%
F-Construction (41-43)	54	19%
H-Transportation and storage (49)	15	5%
I- Accommodation and food service activities (55-56)	9	3%
J-Information and communication (58-63)	7	2%
M-Professional, scientific and technical activities (69-75)	6	2%
Others	15	5%
TOTAL	**289**	**100%**
Source: Author		

Concerning size, most of the SMEs of Portugal belongs to the micro category, with fewer than 10 employees. Micro and small enterprise categories account for around 90% of the insolvent as well as of the active companies. The distribution is shown in Table 12.

Table 12: Distribution of Portuguese SMEs by Size

Company Size	Insolvent	Insolvent %	Active	Active %
Micro (<10 employees)	137	47%	156	54%
Small (<50 employees)	121	42%	106	37%
Medium-sized (<250 employees)	31	11%	26	9%
Total	**289**	**100%**	**288**	**100%**
Source: Author				

4.4 Descriptive Statistics

Insolvent companies present lower liquidity and profitability ratios, as expected. For these companies, the ratios that have EBIT/ Operating Cash Flow/ Net Income as numerator are all negative. In terms of leverage, the Operating Cash Flow to Debt ratio, which measures creditworthiness, is high for active companies and has negative/ close to zero values for the insolvent group, while the Retained Earnings to Total Assets ratio is also negative for the insolvent companies. Total Liabilities to Total Assets ratio is higher for the insolvent companies, also as expected. As for solvency, interest coverage ratios are much higher for the active companies showing better capability to meet its interest obligations from operating earnings.

Table 13 shows the descriptive statistics of the initial ratios considered, for the data set after PSM 1 to 1. Appendix 6 presents the descriptive statistics for the data set without PSM. Appendices 7 and 8 show the descriptive statistics for the data set after PSM 1 to 5 and PSM 1 to 10, respectively.

Table 13: Descriptive Statistics – After PSM 1 to 1

Ratios	Active Companies				Insolvent Companies			
	Obs	**Mean**	**St. Dev.**	**Median**	**Obs**	**Mean**	**St. Dev.**	**Median**
X1	289	3.815	6.751	1.873	289	1.686	3.003	1.013
X2	289	0.245	0.283	0.187	289	0.119	0.392	0.127
X3	289	1.868	3.343	0.943	289	0.630	0.755	0.402
X4	289	0.772	1.922	0.198	289	0.109	0.230	0.030
X5	289	0.661	0.274	0.719	289	0.682	0.266	0.754
X6	289	0.035	0.108	0.030	289	-0.172	0.305	-0.077
X7	289	0.050	0.112	0.042	289	-0.163	0.301	-0.069
X8	289	0.025	0.197	0.034	289	-0.373	0.896	-0.130
X9	289	0.015	0.100	0.014	289	-0.198	0.320	-0.096
X10	289	2.031	9.754	1.685	289	0.666	42.876	-1.450
X11	289	9.785	19.924	6.930	289	0.070	74.338	-5.347
X12	289	0.135	0.288	0.067	289	-0.117	0.192	-0.061
X13	289	0.247	0.350	0.231	289	-0.693	1.344	-0.259
X14	289	0.656	0.281	0.686	289	1.272	0.851	1.022
X15	289	108.109	648.462	4.107	289	-58.890	263.256	-5.126
X16	289	137.105	743.307	8.421	289	-28.404	112.377	-3.334
X17	289	1.047	1.595	0.459	289	-0.004	0.446	-0.021
X18	289	1.130	1.153	0.880	289	0.904	0.859	0.715
X19	289	0.425	55.267	2.945	289	-3.097	36.934	0.832
Source: Author								

This table presents the number of observations, mean, standard deviation and median for the variables based on the sample composed of insolvent companies with data from one year prior to insolvency, and active companies matched to the insolvent by PSM 1 to 1. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

In order to see if here is a significant difference between the two groups, a t-test was applied to verify the hypothesis that the means of the independent variables (financial ratios) are not equal. Table 14 shows the results of the t-test, associated p-values and the difference between means (active minus insolvent) for the PSM 1 to 1 matched groups. Appendix 9 shows the results of the same tests applied to the data set before PSM and Appendix 10, the same for the data set PSM 1 to 5 matched groups.

The results shown in Table 14 show that the differences between the means of the two groups are statistically significant at the 1% level for all ratios except for the Total Liabilities to EBITDA ratio, which is significant at the 5% level, and the Total Liabilities to Shareholders` Funds and Net Income to Working Capital ratios, for which the difference in the means of the two groups is not statistically significant.

Table 14: Test of Equality of Means Between Active and Insolvent Companies – After PSM 1 to 1

Ratios	Difference between means	t value	Pr(|T| > |t|)
X1	2.128	4.897	0,000***
X2	0.126	4.437	0,000***
X3	1.238	6.139	0,000***
X4	0.662	5.818	0,000***
X5	-0.021	-0.922	0,357
X6	0.207	10.886	0,000***
X7	0.213	11.287	0,000***
X8	0.398	7.370	0,000***
X9	0.213	10.805	0,000***
X10	1.365	0.528	0.598
X11	9.715	2.146	0.032**
X12	0.252	12.354	0,000***
X13	0.940	11.501	0,000***
X14	-0.616	-11.694	0,000***
X15	166.999	4.057	0,000***
X16	165.508	3.743	0,000***
X17	1.050	10.780	0,000***
X18	0.226	2.672	0.008***
X19	3.522	0.901	0.368
Source: Author			

This table presents the test of equality of means (active minus insolvent) between the insolvent companies with data from one year prior to insolvency, and the active companies matched to the insolvent by PSM 1 to 1. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 15 presents the results for the PSM 1 to 10 matched groups. In this case the differences between the means of the two groups of companies are statistically significant at the 1% level for all ratios except for the Debt to Equity ratio, which is significant at the 10% level.

Table 15: Test of Equality of Means Between Active and Insolvent Companies - After PSM 1 to 10

Ratios	Difference between means	t value	Pr(|T| > |t|)
X1	1.196	4.574	0,000***
X2	0.124	7.430	0,000***
X3	0.682	7.438	0,000***
X4	0.341	6.796	0,000***
X5	0.176	25.014	0,000***
X6	0.183	24.838	0,000***
X7	0.187	26.129	0,000***
X8	0.307	19.739	0,000***
X9	0.179	26.316	0,000***
X10	1.645	1.954	0.051*
X11	11.266	4.505	0,000***
X12	19.214	7.068	0,000***
X13	0.746	26.626	0,000***
X14	-0.497	-23.403	0,000***
X15	98.734	4.817	0,000***
X16	93.157	4.137	0,000***
X17	0.784	13.331	0,000***
X18	0.267	5.192	0,000***
X19	6.144	2.712	0.007***
Source: Author			

This table presents the test of equality of means (active minus insolvent) between the insolvent companies with data from one year prior to insolvency, and the active companies matched to the insolvent by PSM 1 to 10. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.
*p < .05 **p < .01 ***p < .001

Table 16 presents Pearson`s correlation coefficients between the ratios analysed for the data set after PSM 1 to 1, considering the full model with all the 19 ratios considered at the beginning of the analysis. Table 17 presents Pearson`s correlation coefficients for the data set after PSM 1 to 10. It is expected that ratios that have similar financial information present strong correlation, such as liquidity ratios Cash and Receivables to Current Liabilities and Operating Cash Flow to Current Liabilities, which have in common Operating Cash Flow and Current Liabilities data, or profitability ratios EBIT to Total Assets, Operating Cash Flow to Total Assets, Net Income to Total Assets, all of which have in common revenue information and Total Assets data. Appendix11 presents Pearson`s correlation coefficients for the data set before PSM. Appendix 12 presents Pearson`s correlation coefficients for the data set after PSM 1 to 5.

Table 16: Pearson`s Correlation Coefficients for the Financial Ratios After PSM 1 to 1

PSM 1 to 1	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
X1	1.0000																		
X2	0.2925*	1.0000																	
X3	0.6875*	0.1978*	1.0000																
X4	0.5684*	-0.0064	0.7847*	1.0000															
X5	0.1527*	0.3937*	0.0855*	-0.0295	1.0000														
X6	0.1125*	0.2513*	0.1300*	0.0980*	-0.0484	1.0000													
X7	0.1035*	0.2322*	0.1254*	0.0977*	-0.1014*	0.9715*	1.0000												
X8	0.1199*	0.1232*	0.0886*	0.0799	0.0992*	0.4315*	0.3851*	1.0000											
X9	0.1175*	0.2580*	0.1313*	0.0998*	-0.0584	0.9902*	0.9808*	0.4189*	1.0000										
X10	-0.0127	-0.0054	-0.0094	-0.0073	-0.0122	0.0330	0.0293	0.0258	0.0328	1.0000									
X11	0.0952*	0.0808	0.0156	0.0007	0.0194	0.0869*	0.0899*	0.1199*	0.0892*	0.0669	1.0000								
X12	0.1314*	0.0861*	0.2167*	0.2146*	-0.0964*	0.6940*	0.6993*	0.3462*	0.6671*	0.0103	0.0564	1.0000							
X13	0.1637*	0.3811*	0.1906*	0.1579*	-0.0256	0.6972*	0.7042*	0.3783*	0.7330*	0.0308	0.0450	0.4383*	1.0000						
X14	-0.1777*	-0.4015*	-0.2034*	-0.1706*	0.0259	-0.5300*	-0.5555*	-0.2900*	-0.5617*	-0.0414	-0.0347	-0.3517*	-0.9276*	1.0000					
X15	0.0762	0.0766	0.1811*	0.1556*	0.0869*	0.1959*	0.1721*	0.1343*	0.1713*	0.0104	0.0359	0.4151*	0.1365*	-0.1195*	1.0000				
X16	0.0745	0.0684	0.1801*	0.1595*	0.0744	0.1923*	0.1749*	0.1016*	0.1697*	0.0064	0.0149	0.4304*	0.1269*	-0.1117*	0.9630*	1.0000			
X17	0.2633*	0.1625*	0.3413*	0.3595*	-0.0529	0.2713*	0.2760*	0.2018*	0.2756*	0.0072	0.0036	0.5246*	0.4241*	-0.5176*	0.2160*	0.2212*	1.0000		
X18	-0.0855*	-0.0996*	-0.0102	0.0014	0.1887*	-0.0074	-0.0118	0.1592*	-0.0284	0.0115	-0.0043	0.0701	-0.1044*	0.0998*	0.1261*	0.1249*	-0.0549	1.0000	
X19	0.0157	0.0775	0.0105	-0.0105	0.0234	0.0596	0.0585	-0.0285	0.0591	0.0227	-0.0021	0.0234	0.0305	-0.0338	0.0116	0.0153	0.0122	-0.0305	1.0000

Source: Author

This table presents Pearson`s correlation coefficients for the data set composed of the insolvent companies with data from one year prior to insolvency, and the active companies matched to these insolvent companies by PSM 1 to 1. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

* denotes statistical significance at 5% or inferior.

Table 17: Pearson`s Correlation Coefficients for the Financial Ratios After PSM 1 to 10

PSM 1 to 10	**wX1**	**wX2**	**wX3**	**wX4**	**wX5**	**wX6**	**wX7**	**wX8**	**wX9**	**wX10**	**wX11**	**wX12**	**wX13**	**wX14**	**X15**	**X16**	**X17**	**X18**	**X19**
X1	1.0000																		
X2	0.3269*	1.0000																	
X3	0.5738*	0.2199*	1.0000																
X4	0.4788*	-0.0452*	0.7536*	1.0000															
X5	0.0694*	0.1385*	0.1833*	0.1694*	1.0000														
X6	0.0269	0.0466*	0.1731*	0.1710*	0.9470*	1.0000													
X7	0.0352	0.0494*	0.1718*	0.1684*	0.9403*	0.9913*	1.0000												
X8	0.0489*	0.0694*	0.1514*	0.1343*	0.6537*	0.6214*	0.6188*	1.0000											
X9	0.0787*	0.1433*	0.1816*	0.1660*	0.9904*	0.9340*	0.9454*	0.6492*	1.0000										
X10	-0.0383	0.0060	-0.0646*	-0.0624*	0.0142	0.0120	0.0161	0.0355	0.0186	1.0000									
X11	0.0361	0.0407*	-0.0372	-0.0230	0.0573*	0.0467*	0.0567*	0.1372*	0.0679*	0.1010*	1.0000								
X12	0.0102	0.0976*	-0.0228	-0.0268	0.0928*	0.0540*	0.0663*	0.1454*	0.1067*	0.0835*	0.2925*	1.0000							
X13	0.1867*	0.2387*	0.3226*	0.2913*	0.6758*	0.6289*	0.6502*	0.5034*	0.6976*	0.0032	0.0289	0.1157*	1.0000						
X14	-0.2389*	-0.2551*	-0.3508*	-0.3110*	-0.5024*	-0.4630*	-0.4870*	-0.3947*	-0.5278*	0.0304	-0.0070	-0.0914*	-0.9083*	1.0000					
X15	0.0978*	0.0257	0.1681*	0.2016*	0.2334*	0.2073*	0.2010*	0.1782*	0.2268*	-0.0246	0.0329	0.0207	0.1827*	-0.1431*	1.0000				
X16	0.1009*	-0.0084	0.1533*	0.1993*	0.1700*	0.1589*	0.1565*	0.1304*	0.1675*	-0.0245	0.0224	0.0107	0.1355*	-0.1111*	0.9158*	1.0000			
X17	0.3781*	0.1496*	0.4723*	0.4710*	0.2916*	0.2642*	0.2793*	0.2465*	0.3072*	-0.1091*	-0.0396*	-0.0044	0.5960*	-0.7111*	0.1778*	0.1611*	1.0000		
X18	-0.1783*	-0.1211*	-0.0232	-0.0008	0.1597*	0.1964*	0.1617*	0.1471*	0.1218*	-0.0162	-0.0412*	-0.0447*	0.0369	0.0397*	0.0102	0.0154	-0.0709*	1.0000	
X19	-0.0205	0.0161	0.0326	0.0338	0.0770*	0.0791*	0.0770*	0.0305	0.0743*	-0.0225	-0.0181	0.0003	0.0693*	-0.0449*	-0.0017	-0.0026	0.0331	0.0790*	1.0000

Source: Author

This table presents Pearson`s correlation coefficients for the data set composed of the insolvent companies with data from one year prior to insolvency, and the active companies matched to these insolvent companies by PSM 1 to 10. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

* denotes statistical significance at 5% or inferior.

The ratios presenting higher correlation coefficients are X3, Quick Ratio (cash plus accounts receivable to current liabilities) and the X4, Operating Cash Flow Ratio (cash flow to current liabilities), which are very similar in their composition. The same thing can be said about X6, Return to Total Assets (EBIT to total assets), X7, Cash Flow to Total Assets and X9, ROA (net income to total assets) which are also correlated and with similar structure, as well as about X15, Interest Coverage (EBIT to interest paid) and X16, EBITDA to Interest Paid.

The data set after PSM 1 to 10 additionally shows correlations between X5, Current Assets to Total Assets and X6, X7, X9 which are also correlated in the PSM 1 to 1 data set.

4.5 Ratio Selection

To identify the ratios with the best discriminating power for the financial distress prediction model we have performed several statistical tests on the initial ratios.The analysis is performed following the steps described below on several combinations of data sets, always keeping the insolvent companies' data for one year prior to insolvency and using different control groups obtained by PSM method as well as a random selection of 80% of all active companies.

The analysis is done taking into consideration the following:

(1) T-test: a large t score indicates that the groups are different, thus statistically significant higher t values for a ratio indicate better discriminating power between the insolvent and the active group;
(2) F-test: a large F ratio indicates that the variability among group means is large relatively to the within group variability, thus the higher the F ratio the higher its discriminating power;
(3) Linear Discriminant Analysis (LDA): provides standardised coefficients for each ratio which indicate the relative contribution of each variable to the model. The higher the standardized coefficient loading of a ratio, the more it contributes to the discrimination between the groups (Mardia et al., 1979). Appendix 13 presents the ranking of the standardised coefficients for the data sets PSM 1 to 1 = composed of insolvent companies`data one year prior to insolvency and: matched active companies by PSM 1 to 1; PSM 1 to 5 = composed of insolvent companies`data one year prior to insolvency and: matched active companies by PSM 1 to 5; PSM 1 to 10 = composed of insolvent companies`data one year prior to insolvency and: matched active companies by PSM 1 to 10; and a random selection of 80% of the active companies`data from

years corresponding to the insolvent companies. According to some authors (e.g., Huberty, 1994), structure coefficients, which measure the correlation between each ratio and the discriminant function, can also be used for interpretation (the greater the structure coefficient the more important the ratio is for discrimination). Variables with a coefficient above 30% are considered as having medium to high discriminating power. Appendix 14 presents the ranking of the structure coefficients for the same data sets as above: PSM 1 to 1; PSM 1 to 5; PSM 1 to 10; and a random selection of 80% of the active companies.

(4) Principal component analysis (PCA): is a technique that reduces the dimensionality of large data sets in order to increase interpretability, while minimising information loss. It does so by creating new uncorrelated variables that successively maximise variance, variables which are called principal components (Jolliffe and Cadima, 2016). Appendices 15, 16 and 18 presents the graphs with the first and second component loadings (which account for the most variance and second most variance) for the data sets PSM 1 to 1, PSM 1 to 10, and the data set with the random selection of 80% of the active companies, respectively

(5) Logit regression: X-standardised coefficients. With full standardisation both the X and the Y variables are standardised to have a mean of 0 and a standard variation of 1. By standardising the X variables only, the relative importance of the X variables can be observed. Thus, the higher its coefficient (absolute value), the higher the discriminating power between the groups of a variable. Appendices 19, 20 and 21 present the output of the logit regression with X-standardised coefficients for the data sets PSM 1 to 1, PSM 1 to 5 and PSM 1 to 10, respectively.

(6) Probit regression: marginal effects provide an estimation of the effect of changing an independent variable by one unit, to the probability of the outcome (dependent variable). The higher its coefficient the more discriminating power between the two groups a variable has. Appendices 22 and 23 present the marginal effects for the data sets PSM 1 to 1 and PSM 1 to 10.

The selection procedure started by progressively eliminating the variables that according to all the above tests do not show to have discriminant capacity, then performing logistic regression to assess the interaction of the variables by adding or replacing variables and testing for performance improvement, by eliminating correlated variables or introducing a variable that showed intermediate discriminating power according to previously performed statistical tests.

5 Analysis and Discussion of Results

As previously stated, the purpose of this study is to empirically test a logit model of financial distress prediction appliable to Portuguese SMEs.

From the several tests performed the ratios that appear as the most relevant ones to predict financial distress in the case of Portuguese SMEs are the following.

X5 = Current Assets / Total Assets. This ratio indicates the extent of total funds invested for the purpose of working capital and helps to measure the liquidity of the company. If this ratio is high the company presents a low risk regarding its ability to cover liabilities in the short term. While a moderate to low value may mean that the company is willing to take some risks but try to compensate for them by increasing profitability via increased investment in fixed assets, if this ratio is low the company presents a higher risk of financial distress.

X7 = Cash Flow / Total Assets. This ratio measures the amount of operating cash flow the company generates from one unit of currency of assets owned. A higher ratio implies a more efficient use of the company`s assets and thus a lower probability of financial distress. In the data sets used in this study, the mean for this ratio is negative in the insolvent firms and positive in the healthy ones.

X12 = Cash Flow / Total Liabilities. This ratio measures the creditworthiness of a company, its ability to generate operating cash flow in order to settle debt. The higher this ratio the more comfortable a company is to pay its obligations and thus the lower the probability of financial distress. When the company`s available cash exceeds its liabilities, it is less likely it will face financial distress (Ong et al., 2011), while insufficient cash flow for the repayment of total liabilities may signal potential distress, which could be associated with high leverage. In the data sets used in this study, the mean for this ratio is negative in the insolvent firms and positive in the healthy ones.

X13 = Retained Earnings / Total Assets. This ratio shows the financial leverage of the company and also shows how much it is relying on debt for the funding of its total assets. It is also a measure of the accumulated profitability of a company over time, and it also indicates if the company has some reserve that might help during eventual poor performance periods. The higher this ratio, the lower the probability of financial distress. For this ratio as well the mean for the insolvent companies is negative while positive for the active ones.

X17 = Equity to Debt. This ratio measures a company`s ability to meet its debt obligations by its shareholders` funds, and the higher the coverage the lower the probability of insolvency. Since shareholders` funds and debt are two sources of financing for a company, appealing more to debt than to shareholders` funds makes the company burdened with high interest expenses and other short-term liabilities. According to Altman (1968, p. 595), this ratio "shows how much the firm`s assets can decline in value (...) before the liabilities exceed the assets and the firm becomes insolvent".

Bellovary et al. (2007) did a comprehensive review of bankruptcy prediction studies published after 1930, and found the ratio Current Assets to Total Assets to have been used in 26 studies, Cash Flow to Total Assets in 15, Cash Flow to Total Liabilities, in 14, and Retained Earnings to Total Assets appears registered with greatest frequency being employed in 42 studies. For instance, Retained Earnings to Total Assets and Equity to Debt are two of the ratios that compose Altman`s Z Score (Altman, 1968, 1983, 2017), Cash Flow to Total Assets is one of the five ratios that compose the logistic model specific for SMEs developed by Altman and Sabato (2007).

Table 18 presents the log odds coefficient estimates for logit Models 1 (estimated from the data set obtained by PSM 1 to 1) and 2 (estimated from the data set obtained by PSM 1 to 10) and the p-value associated with the z-statistic reported by the logit models. No industry dummies are included. For Model 1, all coefficients are statistically significant except X5 – Current Assets to Total Assets, which, on the other hand, is statistically significant in Model 2. In Model 2 all coefficients are statistically significant except for X13 – Retained Earnings to Total Assets. Nevertheless, substituting these ratios for correlated ones or eliminating them altogether resulted consistently in lower performance models. This could be explained by the fact that these ratios are important and add in discriminating power when combined with the other ratios present in the models.

Table 18: Coefficients Estimates for Model 1 and Model 2

VARIABLES	Model 1	Model 2
X5	0.379	4.511**
	(0.433)	(1.972)
X7	6.841***	-10.388***
	(1.889)	(1.985)
X12	-12.285***	-0.005***
	(2.188)	(0.001)
X13	-1.836***	-0.336
	(0.538)	(0.224)
X17	-1.187***	-1.528***
	(0.348)	(0.274)
Constant	0.035	-1.467***
	(0.334)	(0.105)
Observations	578	2,474
Pseudo R-squared	0.417	0.312
Prob > chi2	0.000	0.000
Source: Author		

This table contains the estimation results for the logit models .The dependent variable equals zero if the firm is not financially distressed and one otherwise. The column Model 1 contains the results of the estimation using the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1. The column Model 2 contains the results of the estimation using the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10. X5, Current Assets to Total Assets; X7, Cash Flow to Total Assets; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X17, Equity to Debt. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

According to the LR-statistic, both models are significant at 1% which means they are appropriate for the study of financial distress prediction. We have also estimated two additional models, Model 3, estimated from a data sample composed of insolvent companies`data from one year prior to insolvency and active companies matched by PSM 1 to 5, and Model 4, estimated from a data sample composed of insolvent companies`data from one year prior to insolvency and a random selection of 80% of active companies` data from the same years as the insolvent companies (see appendix 23). Although Model 4, which includes the sample of insolvent companies and a control sample formed of a random selection of 80% of the active companies, presented a very high rate of accuracy, the results might be biased because of the extreme unbalance of the data set. Due to this, when using this model for validating the fit on the out-of-sample test data sets, only a very small proportion of the insolvent companies can be correctly identified.

Because we are trying to model failure, the expected signs of the variable coefficients are counterintuitive and thus we anticipate negative signs for the ratios whose high value means less probability of failure and vice-versa. Although negative values were expected for the coefficients of the ratios Current Assets to Total

Assets and Operating Cash Flow to Total Assets ratios, Model 1 shows positive values for these coefficients, and Model 2 shows a positive coefficient for the ratio Current Assets to Total Assets as well. This could be explained by the fact that, for Current Assets, the difference in the means of the two groups is very small and statistically nonsignificant. It could also be due to the fact that insolvent companies may have high levels of inventory or customers`due payments incorrectly recorded accounts.

The positive coefficient of the ratio Operating Cash Flow to Total Assets in Model 1 could be explained by the fact that the sample contains companies from all business sectors of Portugal which can have very different Operating Cash Flow profiles, due to which this ratio may also reflect their business characteristics. Model 2 as well as Model 3 estimated from PSM 1 to 10 and PSM 1 to 5, respectively, show the expected sign for the coefficient of this ratio which may be due to a larger sample.

In both Models 1 and 2, the ratios Operating Cash Flow to Total Assets and Shareholders` Equity to Total Liabilities are statistically significant, which indicates that these ratios are the better predictors. In Model 1, the ratio Retained Earnings to Total Assets is also statistically significant. The ratio Current Assets to Total Assets is also statistically significant in Model 2.

Overall, the coefficients show that the ratio that has the best predictive power in both models is X7 - Operating Cash Flow to Total Assets ratio. This was found to be an important ratio also by Beaver (1966), who reported its significant relationship with the probability of insolvency, as well as by Vieira (2013).

Following Agarwal and Taffler (2008) and Altman et al. (2017) among others, we assessed the classification performance of the models by the Area Under Curve (AUC) extracted from the ROC (Receiver Operating Characteristic) curve. If a model is incapable of discriminating between insolvent and active companies, the ROC curve will be a 45 degree line; the greater the predictive power of the model the more bowed the ROC curve will be (Charalambakis, E. C., Garrett, I., 2018). AUC is closely connected to the Accuracy Ratio (AR), since AR = 2 x AUC -1. The larger the AUC the better the model is at predicting financial distress. Figures 2 and 3 show the AUC for Models 1 and 2. For Model 1, AR = 79.50% and for model 2, AR = 76.22%, the accuracy of Model 1 being slightly better.

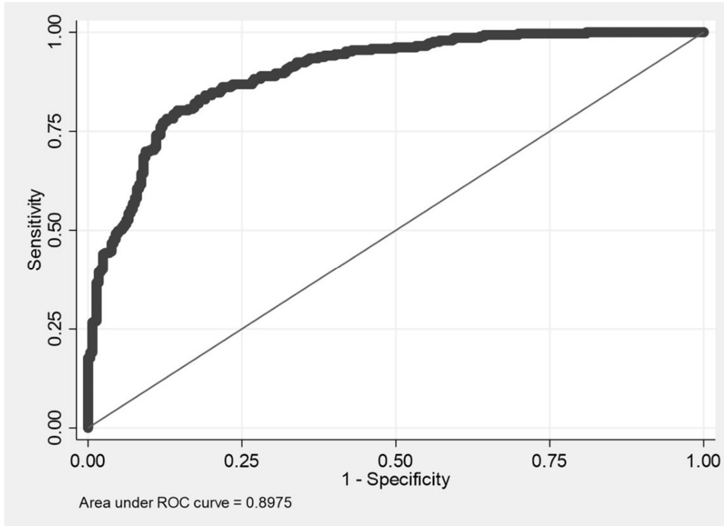


Figure 2: Area Under Curve – Model 1 (PSM 1 to 1)

This figure shows the ROC curve and the Area Under Curve for Model 1 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1.

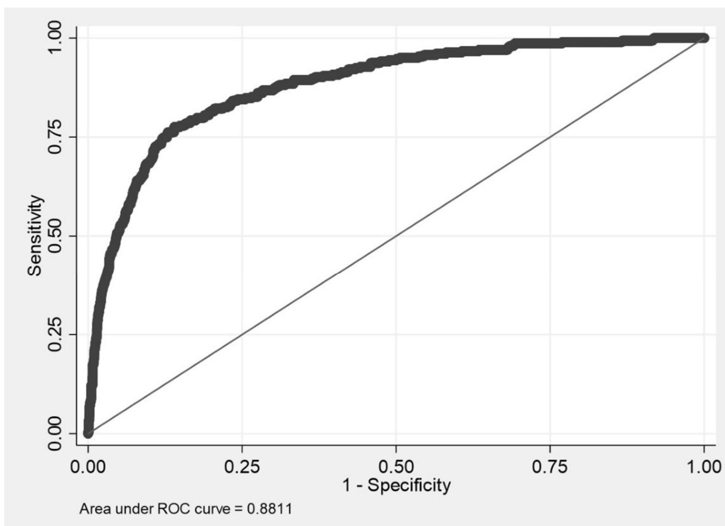


Figure 3: Area Under Curve – Model 1 (PSM 1 to 10)

This figure shows the ROC curve and the Area Under Curve for Model 2 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10.

We have applied goodness-of-fit (GOF) tests based on covariate patterns - Pearson`s Chi-square test, and based on estimated probabilities - Hosmer-Lemeshow test (Hosmer and Lemeshow, 1989). The results, presented in Tables 19 and 20, respectively, show a better fit of Model 1 with both tests showing no statistical significance for this model.

Table 19: GOF - Pearson`s Chi-square Test

	Model 1	Model 2	
Observations	578	2,474	
Covariate patterns	578	2,474	
Pearson chi2 (572)	560.9	Pearson chi2 (2468)	2,751.33
Prob > chi2	**0.622**	0.000	
Source: Author			

This table presents the results of Pearson`s Chi-square tests for for Model 1 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1 and for Model 2, estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10.

Table 20: GOF - Hosmer-Lemeshow Test

	Model 1	Model 2
Observations	578	2474
Groups	10	10
Hosmer-Lemeshow chi2 (8)	14.6	26.96
Prob > chi2	**0.068**	0.001
Source: Author		

This table presents the results of Pearson`s Chi-square tests for for Model 1 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1 and for Model 2, estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10.

Financial distress prediction models assess the ability to predict by counting the total errors, and there are two types of errors that occur when classifying companies, which are: Type I, which is to classify a financially distressed company as healthy, and Type II, which classifies a healthy company as financially distressed (Altman, 1968). A company is classified as financially distressed if its probability of default score is above the cutoff point, and as healthy if the score is below the cutoff point.

According to Weiss (1996), Type I errors have the consequence of loss from lending to firms that end up as insolvent while type II errors incur the opportunity cost of not lending to firms that do not end up as insolvent but continue healthy. According to du Jardin (2010), most models correctly predict healthy firms at a rate higher than that at which they predict failing firms, and this is a common result in the financial literature regardless of the modelling technique. For those who may use the model as a decision tool, du Jardin (2010, p. 2051) finds that "the cost of having a failing company classified as healthy (Type I error) is far greater than the cost of a healthy company classified as failing (Type II error). A Type I error involves the loss of an investment or debt that will not be reimbursed as result of bankruptcy, while a Type II error involves the loss of a potential bargain. Thus, models should avoid above all Type I errors". For example, Altman et al. (1977) find the cost for Type I errors is 70% of the amount lent while for Type

II errors is 2% of the amount that could have been lent, but this estimation does not take into account the size of the company or the loan amount.

The results of predictive ability of Models 1 and 2 are tabulated at first on the basis of a cutoff point of .5. The output of the estimation of a logit model gives results in terms of sensitivity, which (in the terms of this study) is the probability that the test result will be positive for an insolvent company, a true positive rate, and specificity, which again in the terms of this study is the probability that the test result will be negative for a healthy company, a true negative rate. When a higher value of the cutoff is selected, false positives decrease, with increased specificity, but at the same time true positives and sensitivity decreases. When a lower cutoff value is chosen, true positives and sensitivity increase but at the detriment of true negatives and specificity which will decrease (Jackson & Wood, 2013).

In order to account for the unbalance in insolvent and active companies in the data sets, another cutoff was estimated for each model, this time in order to minimize the sum of the errors. The sensitivity/specificity versus probability cutoff graphs, illustrating the optimal cutoff points at the crossing of sensitivity and specificity lines, are shown in Figure 4 for Model 1 and Figure 5 for Model 2.

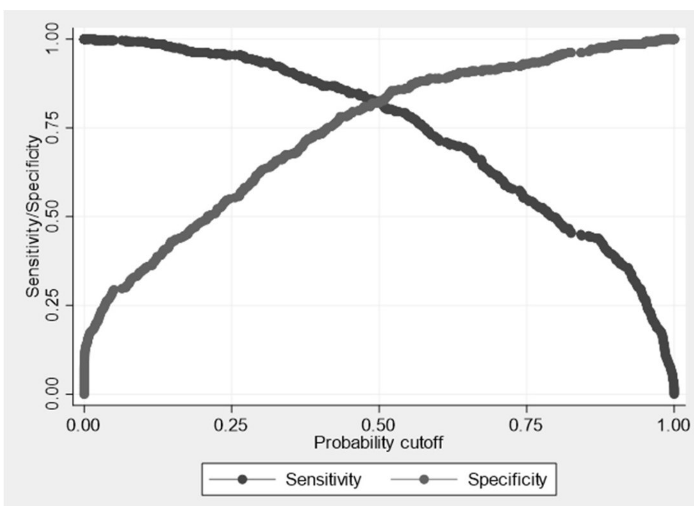


Figure 4: Model 1 (PSM 1 to 1) Estimation – Sensitivity and Specificity vs Probability Cutoff

This figure presents the sensitivity line (positive result for an insolvent company, true positive) and the specificity line (negative result for a healthy company, true negative), and the encounter of these lines at the cutoff point, for Model 1 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1.

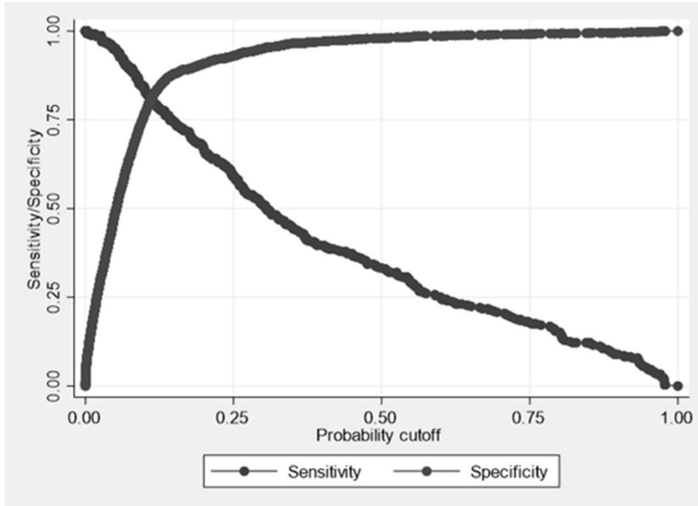


Figure 5: Model 2 (PSM 1 to 10) Estimation – Sensitivity and Specificity vs Probability Cutoff

This figure presents the sensitivity line (positive result for an insolvent company, true positive) and the specificity line (negative result for a healthy company, true negative), and the encounter of these lines at the cutoff point, for Model 1 estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10.

The predictive accuracy of the two models is presented in Table 21.

Table 21: Predictive Accuracy

	True Status	**Distressed**	**Active**
MODEL 1	Distressed	82.01	17.99
	Active	17.65	82.35
Cutoff = .5	Accuracy	82.18	
	Distressed	79.93	20.07
	Active	14.53	85.47
Cutoff = .521	Accuracy	82.70	
MODEL 2	Distressed	33.00	67.00
	Active	2.03	97.97
Cutoff = .5	Accuracy	90.02	
	Distressed	77.56	22.44
	Active	14.00	86.00
Cutoff = .134	Accuracy	84.96	
Source: Author			

This table presents the predictive accuracy for Model 1, estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 1, and Model 2, estimated from the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 10.

For model 1, the cutoff setting to minimise the sum of errors slightly increases the overall predictive ability of the model but increases the Type I error, therefore if using this model as a decision tool, keeping the cutoff of .5 might help to avoid the costs associated with this type of error.

For Model 2, the optimal cutoff is the one that minimises the sum of errors and also greatly reduces the Type I error. This is due to the fact that the sample is unbalanced in the sense that the number of active companies is about 7 times larger (the PSM method was not able to find 10 counterparts for each of all the insolvent companies).

After estimation, the models were tested on the following validation samples: for insolvent companies, data from year N-1 (one year before insolvency), year N-2 (two years before insolvency), year N-3 (three years before insolvency) and for N-1&2&3 = years one, two and three before insolvency, all together; for the corresponding active companies was used the total available sample, respecting the same years of financial data as the insolvent companies in each group. Another testing was done with all data available for insolvent companies, up to 6 years prior to insolvency, and active companies data from the same years.

Table 22 presents the number of observations used for testing estimated models`forecast accuracy.

Table 22: Number of Observations/ Companies Used for Testing Forecast Accuracy

Period	**N-1**		**N-2**		**N-3**		**N-1&2&3**		**N-1to6**	
	Distressed	Active	Distressed	Active	Distressed	Active	Distressed	Active	Distressed	Active
Observations	304	235,711	713	200,456	793	159,376	1,810	235,711	3,282	235,711
Companies	304	64,493	713	64,415	793	64,142	1,032	64,493	1,504	64,493
Years	2013-2018		2013-2017		2013-2016		2013-2018		2013-2018	
Source: Author										

This table presents the composition of the data sets used for testing of the estimated models, Model 1 and Model 2. N-1 is a data set made up of insolvent companies`data one year prior to insolvency (same insolvent companies used for models estimation). N-2 contains insolvent companies`data from two years prior to insolvency and active companies`data from the same respective years. N-3 contains insolvent companies`data from three years prior to insolvency and active companies`data from the same respective years. N-1&2&3 contains insolvent companies`data from 1, 2 and 3 years prior to insolvency, all together, and active companies`data from the same respective years. N-1to6 contains all insolvent companies`data available for this study, up to 6 years prior to insolvency and active companies`data from the same respective years. N-1 , N-1&2&3 and N-1to6 all contain data for insolvent companies used for models estimation (for year N-1, one year prior to insolvency). N-2 and N-3 are out-of-sample data sets.

Table 23 presents the forecast accuracy of Model 1 and Model 2, with both 0.5 cutoff and re-estimated cutoffs.

Table 23: Forecast Accuracy for Models 1 and 2 (%)

		N-1		**N-2**		**N-3**		**N-1&2&3**		**N-1to6**	
	True Status	Distressed	Active	Distressed	Active	Distressed	Active	Distressed	Active	Distressed	Active
MODEL 1	Distressed	78.62	21.38	58.77	41.23	51.32	48.68	59.28	40.72	56.31	43.69
	Active	19.53	80.47	20.21	79.79	21.27	78.73	19.53	80.47	19.51	80.49
Cutoff = .5	**Accuracy**	80.46		79.72		78.60		80.30		80.16	
	Distressed	75.99	24.01	56.10	43.90	48.30	51.70	56.63	43.37	53.84	46.16
	Active	19.53	80.47	20.21	79.79	21.27	78.73	19.53	80.47	19.51	80.49
Cutoff = .52	**Accuracy**	80.46		79.71		78.58		80.29		80.12	
MODEL 2	Distressed	72.70	27.30	54.98	45.02	50.32	49.68	64.09	35.91	81.29	18.71
	Active	2.01	97.99	2.09	97.91	2.21	97.79	2.02	97.98	2.47	97.53
Cutoff = .5	**Accuracy**	97.96		97.76		97.56		97.72		97.31	
	Distressed	97.37	2.63	94.67	5.33	93.06	6.94	95.08	4.92	96.19	3.81
	Active	13.44	86.56	13.96	86.04	14.68	85.32	13.45	86.55	13.53	86.47
Cutoff = .13	**Accuracy**	86.57		86.07		85.36		86.62		86.61	
Source: Author											

This table presents the composition of the data sets used for testing of the estimated models, Model 1 and Model 2. N-1 is a data set made up of insolvent companies`data one year prior to insolvency (same insolvent companies used for models estimation). N-2 contains insolvent companies`data from two years prior to insolvency and active companies`data from the same respective years. N-3 contains insolvent companies`data from three years prior to insolvency and active companies`data from the same respective years. N-1&2&3 contains insolvent companies`data from 1, 2 and 3 years prior to insolvency, all together, and active companies`data from the same respective years. N-1to6 contains all insolvent companies`data available for this study, up to 6 years prior to insolvency and active companies`data from the same respective years. N-1 , N-1&2&3 and N-1to6 all contain data for insolvent companies used for models estimation (for year N-1, one year prior to insolvency). N-2 and N-3 are out-of-sample data sets.

Despite the fact that the goodness-of-fit statistics showed a good fit only for Model 1, the testing results suggest that Model 2 is able to predict with better overall accuracy and also with lower Type I errors than Model 1.

Overall predictive accuracy of Model 1 is almost identical with either of the cuttofs, .5 and optimal calculated of .52. Model 1 with the cutoff of .05 performs slightly better than same Model 1 with the cutoff calculated at .52, being able to correctly classify over 50% of the insolvent companies across all data sets. Model 1 with .52 cutoff can classify only 48% of the insolvent companies at year N-3.

Overall predictive accuracy of Model 2 with the cutoff set at .5 is higher than at the optimal calculated cutoff of .13 but, at .13 cutoff Model 2 correctly classifies over 93% of the insolvent companies across all data sets, as opposed to only over 50% across all data sets with the cutoff set at .5.

Model 2 with with the cutoff of .5 performs better than Model 1 for all data sets of Portuguese companies, being able to correctly classify over 50% of the insolvent companies across all data sets, as well as better classify the active companies at over 97%, compared to Model 1 which classifies active companies with an accuracy of 80% for year N-1, 79% for year N-2 and 78% for year N-3.

We can conclude that most appropriate cutoffs are .5 for Model 1 and .13 for Model 2. For each one of them, the overall forecast accuracy level is quite similar through all the periods analysed (years N-1, N-2, N-3, N-1&2&3, and N-1 to 6), of around 80% for Model 1 and 86% for Model 2. However, Type II errors increase the further in time we go from the time of insolvency, as expected. For Model 1, for all periods besides N-1, Type II errors are over 40%. With the adjusted cutoff, Type II errors are very small for Model 2 with the adjusted cutoff of .13, which could be explained by the larger total number of observations used.

5.1 Considerations on Results for Project Hosting Company nBanks

Considering the above, we conclude that the combination of variables presented in this study, which are Current Assets to Total Assets, Operating Cash Flow to Total Assets, Operating Cash Flow to Debt, Retained Earnings to Total Assets and Equity to Debt, can be used in order to timely detect a possible insolvency situation for Portuguese SMEs, or to assess the risk of insolvency of a SME at a specific point in time.

Model 1, obtained from the PSM 1 to 1 matching method, with cutoff set at .5, presents the best GOF and a forecast accuracy of around 78% for one year prior, 58% two years prior and 51% three years prior to insolvency, as well as 59% for years 1 to 3, and 56% for years 1 to 6 prior insolvency. Model 2, obtained from the PSM 1 to 10 matching method, with optimal cutoff set at .13, presents an even better forecast accuracy than Model 1, which may be due to the larger sample of control (active) companies, with a forecast accuracy of around 97% for one year prior, 94% two years prior and 93% three years prior to insolvency.

The same Model 2 with the cutoff set at .5 presents a forecast accuracy similar to Model 1, of over 50% overall capacity of correctly classifying insolvent Portuguese SMEs. Forecast accuracy is of 72% one year prior, 54% two years prior and 50% three years prior to insolvency, as well as 64% for years 1 to 3, and 81% for years 1 to 6 prior to insolvency.

Considering the above, project hosting company may consider Model 2 with .13 cutoff to assess the probability of default of customer SMEs. The estimated logistic function is:

P = $\frac{1}{1+e^{-Z}}$

In which:

P = probability of company failure

$$Z = -1.467 + 4.511 \cdot X5 - 10.388 \cdot X7 - 0.005 \cdot X12 - 0.336 \cdot X13 - 1.528 \cdot X17$$

and X5 = Current Assets/Total Assets; X7 = Operating Cash Flow/Total Assets; X12 = Operating Cash Flow/Total Liabilities; X13 = Retained Earnings/Total Assets; X17 = Shareholders` Funds/Total Liabilities.

Further analyses that include qualitative variables such as activity sector or dummy variables for export activity might help to further improve the forecast accuracy of the model.

6 Conclusions

This study has the objective of estimating by logistic regression and testing a financial distress prediction model for the SMEs of Portugal.

The financial information for both the insolvent companies and the active companies used as control group was collected from AMADEUS database. Four models were estimated using the logistic regression method, with the same ratios but with different estimated coefficients, out of which Models 1 and 2 present the most accurate discriminating results. The propensity score matching method was also used with the aim of reducing the unbalance in terms of number of observations between the insolvent and the active companies and making them more comparable in terms of common criteria such as size, sector of activity and year of the financial information.

Regarding the size of the Portuguese insolvent SMEs of the data set analysed, most belong to the micro and small-sized categories, representing 89% of the total. Medium-sized companies account for the remaining 11%. In terms of localisation, the insolvent companies are situated mainly in the Northern Portugal, followed by Lisbon metropolitan area and Central Portugal, in this order. The main activity sectors which concentrate the insolvent companies are Wholesale and retail trade; repair of motor vehicles and motorcycles (G), Manufacturing (C) and Construction (F).

The results of this study show that the predictive capacity of the estimated models is in general over 80%, with the propensity score matched data sets being more robust than the data sets combined by random selection. The model estimated with 1 to 1 matching appears more robust in terms of goodness-of-fit, while the model estimated with 1 to 10 matching shows a predictive ability superior by around 6% for all the data sets tested.

The overall accuracy for classifying insolvent Portuguese SMEs is above 50% even for three years prior to insolvency for Model 1, and above 90% for Model 2.

The models were estimated using data from all sectors of activity. More accurate models could be estimated using data from companies belonging to the same sector of activity for greater consistency of the behaviour of the financial ratios.

To further improve the model the introduction of qualitative variables, such as existence of export activity or location, could be tested, since some studies (such as Lehmann, 2003) indicate that this improves the predictive power in the case of SMEs. These data were not available for this study. Other studies indicate that the introduction of macroeconomic variables such as inflation, GDP, unemployment, among others, may improve the predictive accuracy (Ptak-Chmielewska and Matuszyk, 2019).

Also, although in recent years access to SMEs financial data has become easier due to increasingly digitised financial data, the reliability of these data still remains uncertain since most SMEs are not legally obligated to present audited financial statements, which can lead to distorted or biased information.

References

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94, 164–184.

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

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589.

Altman, E.I., Haldeman, R. G., Narayan, P. (1977). Zeta Analysis: A New Model to Identify Bankruptcy. *The Journal of Banking and Finance*, June, 29-54.

Altman, E. I. (1983). A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy. *John Wiley & Sons, Inc.*, 5-7.

Altman, E. I. (2002). Revisiting Credit Scoring Models in a Basel 2 Environment. *Unpublished Working Paper, New York Stern University,* (May), 1–37.

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 and Accounting*, 28(2), 131–171.

Altman, E. I., Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U.S. Market. Abacus, 43(3), 332–357

Armour, J. (2001). The Law and Economics of Corporate Insolvency: A Review. *Working Papers* 197*, Centre for Business Research, University of Cambridge*, 3-4.

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

Barros, G. C. O. de. (2008). Modelos de Previsão da Falência de Empresas, Aplicação Empírica ao Caso das Pequenas e Médias Empresas Portuguesas. (Master thesis, *ISCTE, Instituto Universitário de Lisboa)*. Retrieved from https://repositorio.iscte-iul.pt/handle/10071/1462.

Basel Commission on Banking Supervision. (2001). Overview of The New Basel Capital Accord. *BIS*, January. Retrieved from https://www.bis.org/publ/bcbsca02.pdf.

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

Bellovary, J. L., Giacomino, D. E., Akers (2007), M. D. A Review of Bankruptcy Prediction Studies. *Journal of Financial Education*, 33, 1-42.

Black, F., Scholes, M. (1973). The Pricing of Options and Corporate Liabiliites. *The Journal of Political Economy,* 81(3), 637-654.

Butera, G, Faff, R. (2006). An Integrated Multi-model Credit Rating System for Private Firms. *Review of Quantitative Finance and Accounting* , 27(3), 311-340.

Bisogno, M., Restaino, M., & Di Carlo, A. (2018). Forecasting and preventing bankruptcy: A conceptual review. *African Journal of Business Management*, *12*(9), 231–242.

Capitalizar, O. P. (2017). *Diário da República, 1.ª série — N.º 125 — 30 de junho de 2017*. Retrieved from https://www.iapmei.pt/getattachment/PRODUTOS-E-SERVICOS/Qualificacao-Certificacao/Certificacao-PME/Decreto-Lei-81-2017.pdf.aspx

Charalambakis, E. C., Garrett, I. (2018). On Corporate Financial Distress Prediction: What Can We Learn From Private Firms in a Developing Economy? Evidence From Greece. *Review of Quantitative Finance and Accounting*, 52(2), 467-491.

Cochran, W. G., Rubin, D. B (2004), Controlling Bias in Observational Studies: A Review. *The Indian Journal of Statistics*, Series A (1961-2002), 417-446.

Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167–179.

Deheija, R. H., Wahba, S (1998). Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs. *Journal of the American Statistic Associatio*n, 94(448), p.1053-1062.

Dimitras, A.I., Zanakis, Z. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research* , 90, 487-513.

Doumpos, M.; Kosmidou, K.; Baourakis, G. and Zopoundis, C. (2002): “Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis”, European Journal of Operational Research, 138, 392-412.

Du Jardin, P. (2010). Predicting Bankruptcy Using Neural Networks and Other Classification Methods: The Influence of Variable Selection Techniques on Model Accuracy. *Neurocomputing* , 73, 2047-2060.

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

Huberty, C. J. (1994). Applied Discriminant Analysis. *New York: Wiley*

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

Lacerda, A, Moro (2008), R. Analysis of the Predictors of Default for Portuguese Firms. *Banco de Portugal - Estudos e Documentos de Trabalho.*

Lehmann, B (2003). Is It Worth the While? The Relevance of Qualitative Information in Credit Rating. *Working Paper*. Available at http://dx.doi.org/10.2139/ssrn.410186.

Merton, R.C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), 449-470.

NBanks. (2019). *New Banking Relationship Business Model*. Retrieved from https://www.nbanks.net.

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18* (1), 109.

Regulation (EC) No. 1059/2003 of the European Parliament and of the Council of 26 May 2003. *Official Journal L154*, 1-41.

Regulation (EC) No. 1893/2006 of the European Parliament and of the Council of 20 December 2006. *Official Journal L393*, 1-39.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70(1), 41–55.

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

Vieira, E. S., Pinho, C., Correia, C. (2013). Insolvency Prediction in the Portuguese Construction Industry. *Marmara Journal of European Studies*, 21(2), 143-164.

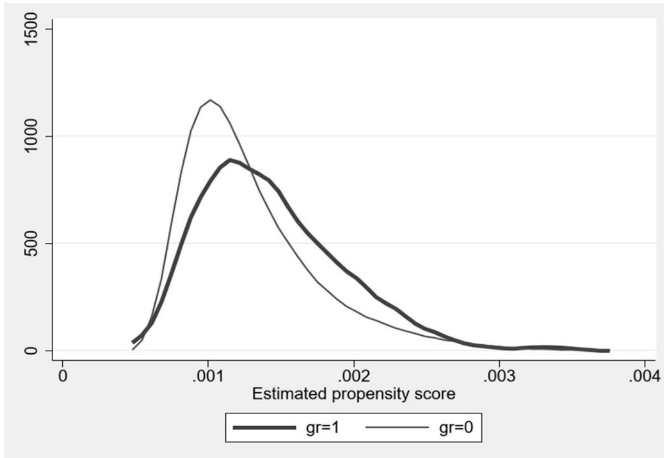
Zavgren, C. V. (1985). Assessing the Vulnerability to Failure of American Industrial Firms: a Logistic Analysis. Journal of Business Finance & Accounting, 12(1), 19–45.

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

Weiss, L. A. (1996). The Impact of Incorporating the Cost Errors Into Bankruptcy Prediction Models. *INSEAD Working Paper Series*, 2.

Appendices

Appendix 1: Kernel Density Plot of the Propensity Score



Appendix 2: Difference in Means After PSM 1 to 5

Variable	Sample	Mean		%bias	t-test	
		Treated	Control		t	p>|t|
Log (Total Assets)	Unmatched	13.757	13.37	29.2	5.29	0.000
	Matched	13.757	13.732	1.9	0/23	0.821

Source: Author

Appendix 3: Data Set Composition After PSM 1 to 5

After PSM 1 to 5	Insolvent	Active
Companies	289	1,083
Observations	289	1,100

Source: Author

Appendix 4: Data Set Composition by Year of Financial Statement After PSM 1 to 5

Year	Insolvent	PSM 1 to 5
2013	34	228
2014	27	99
2015	48	110
2016	60	179
2017	69	233
2018	51	251
Total	289	1,100

Source: Author

This table presents the number of observations by year of financial statement. Column Year represents the year of the financial statement. Columnt Insolvent represents the number of observations for insolvent companies at one year prior to insolvency. PSM 1 to 5 represents number of observations for active companies matched to the insolvent by PSM 1 to 5.

Appendix 5: Model Estimation Data Sets - Distribution by Region (PSM 1 to 5)

Region-NUTS II	Insolvent	PSM 1 to 5
PT11 - North	117	472
PT17 - Area Metropolitana de Lisboa	69	211
PT16 - Centro	61	278
PT18 - Alentejo	16	56
PT15 - Algarve	13	48
PT20 - Regiao Autonoma dos Acores	11	19
PT30 - Regiao Autonoma da Madeira	2	16
TOTAL	289	1100
Source: Author		

This table presents the distribution of the model estimations data sets by region. Insolvent represents the number of insolvent companies per region, PSM 1 to 5 represents the number of matching active companies selected by this matching method, by region.

Appendix 6: Descriptive Statistics for the Financial Ratios before PSM

Ratios	Active Companies				Insolvent Companies			
	Obs	Mean	St. Dev.	Median	Obs	Mean	St. Dev.	Median
X1	235,711	3.415	5.417	1.823	3,282	1.974	3.442	1.256
X2	235,711	0.233	0.258	0.201	3,282	0.214	0.289	0.202
X3	235,711	1.728	2.670	0.958	3,282	0.881	1.402	0.553
X4	235,711	0.739	1.701	0.188	3,282	0.215	0.704	0.040
X5	235,711	0.046	0.101	0.038	3,282	0.544	0.152	0.630
X6	235,711	0.087	0.106	0.076	3,282	-0.035	0.117	0.003
X7	235,711	0.064	0.097	0.056	3,282	-0.032	0.119	0.004
X8	235,711	0.027	0.192	0.038	3,282	-0.117	0.314	0.005
X9	235,711	0.022	0.088	0.018	3,282	-0.063	0.121	-0.015
X10	235,711	3.073	9.735	1.612	3,282	4.162	14.979	2.513
X11	235,711	8.098	27.311	6.039	3,282	5.787	41.981	6.606
X12	235,711	12.342	37.914	7.205	3,282	-0.033	0.249	0.005
X13	235,711	0.231	0.385	0.248	3,282	-0.171	0.512	-0.013
X14	235,711	0.679	0.305	0.674	3,282	0.960	0.373	0.882
X15	235,711	78.045	451.215	4.138	3,282	-3.953	189.776	0.198
X16	235,711	141.205	740.875	9.030	3,282	12.926	256.511	1.249
X17	235,711	0.947	1.503	0.483	3,282	0.209	0.579	0.134
X18	235,711	1.173	0.899	0.972	3,282	1.029	0.847	0.821
X19	235,711	1.175	51.191	2.965	3,282	0.315	40.735	1.867
Source: Author								

This table presents the number of observations, mean, standard deviation and median for the variables, for the initial sample composed of insolvent companies and active companies. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 7: Descriptive Statistics for the Financial Ratios After PSM 1 to 5

Ratios	Active Companies				Insolvent Companies			
	Obs	Mean	St. Dev.	Median	Obs	Mean	St. Dev.	Median
X1	303	2.861	4.032	1.719	1,100	1.736	3.189	1.018
X2	303	0.261	0.260	0.241	1,100	0.132	0.360	0.127
X3	303	1.365	1.506	0.968	1,100	0.679	0.969	0.411
X4	303	0.473	0.861	0.151	1,100	0.122	0.310	0.029
X5	303	0.040	0.110	0.038	1,100	-0.150	0.233	-0.073
X6	303	0.080	0.116	0.075	1,100	-0.115	0.222	-0.042
X7	303	0.059	0.112	0.055	1,100	-0.139	0.226	-0.064
X8	303	0.019	0.249	0.036	1,100	-0.296	0.507	-0.129
X9	303	0.018	0.105	0.020	1,100	-0.173	0.236	-0.093
X10	303	3.487	12.170	1.667	1,100	1.727	20.361	-1.374
X11	303	11.009	40.376	6.174	1,100	-1.209	62.727	-4.959
X12	303	12.744	37.489	7.450	1,100	-7.793	62.263	-5.948
X13	303	0.217	0.422	0.246	1,100	-0.567	0.843	-0.258
X14	303	0.685	0.320	0.670	1,100	1.201	0.591	1.021
X15	303	61.350	413.009	4.213	1,100	-44.975	184.872	-4.549
X16	303	104.254	518.921	9.480	1,100	-18.108	85.313	-3.085
X17	303	0.806	1.028	0.492	1,100	0.009	0.458	-0.020
X18	303	1.203	0.880	1.048	1,100	0.897	0.848	0.715
X19	303	4.645	39.377	3.220	1,100	-1.685	31.340	0.970
Source: Author								

This table presents the number of observations, mean, standard deviation and median for the variables, for the data set after PSM 1 to 5. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 8: Descriptive Statistics for the Financial Ratios After PSM 1 to 10

Ratios	Active Companies				Insolvent Companies			
	Obs	Mean	St. Dev.	Median	Obs	Mean	St. Dev.	Median
X1	303	2.861	4.032	1.719	2,171	1.736	3.189	1.018
X2	303	0.261	0.260	0.241	2,171	0.132	0.360	0.127
X3	303	1.365	1.506	0.968	2,171	0.679	0.969	0.411
X4	303	0.473	0.861	0.151	2,171	0.122	0.310	0.029
X5	303	0.040	0.110	0.038	2,171	-0.150	0.233	-0.073
X6	303	0.080	0.116	0.075	2,171	-0.115	0.222	-0.042
X7	303	0.059	0.112	0.055	2,171	-0.139	0.226	-0.064
X8	303	0.019	0.249	0.036	2,171	-0.296	0.507	-0.129
X9	303	0.018	0.105	0.020	2,171	-0.173	0.236	-0.093
X10	303	3.487	12.170	1.667	2,171	1.727	20.361	-1.374
X11	303	11.009	40.376	6.174	2,171	-1.209	62.727	-4.959
X12	303	12.744	37.489	7.450	2,171	-7.793	62.263	-5.948
X13	303	0.217	0.422	0.246	2,171	-0.567	0.843	-0.258
X14	303	0.685	0.320	0.670	2,171	1.201	0.591	1.021
X15	303	61.350	413.009	4.213	2,171	-44.975	184.872	-4.549
X16	303	104.254	518.921	9.480	2,171	-18.108	85.313	-3.085
X17	303	0.806	1.028	0.492	2,171	0.009	0.458	-0.020
X18	303	1.203	0.880	1.048	2,171	0.897	0.848	0.715
X19	303	4.645	39.377	3.220	2,171	-1.685	31.340	0.970
Source: Author								

This table presents the number of observations, mean, standard deviation and median for the variables, for the data set after PSM 1 to 10. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 9: Test of Equality of Means Between Active and Insolvent Companies - Before PSM

Ratios	Difference between means	t value	Pr(|T| > |t|)
X1	1.125	4.487	0,000***
X2	0.130	7.035	0,000***
X3	0.687	7.519	0,000***
X4	0.351	6.971	0,000***
X5	0.190	20.065	0,000***
X6	0.196	20.696	0,000***
X7	0.198	21.205	0,000***
X8	0.314	15.037	0,000***
X9	0.191	20.519	0,000***
X10	1.761	1.893	0.059*
X11	12.218	4.083	0,000***
X12	20.537	7.190	0,000***
X13	0.784	22.327	0,000***
X14	-0.516	-20.155	0,000***
X15	106.325	4.362	0,000***
X16	122.362	4.088	0,000***
X17	0.796	13.133	0,000***
X18	0.306	5.400	0,000***
X19	6.329	2.582	0.001***
Source: Author			

This table presents the number of observations, mean, standard deviation and median for the variables, for the initial sample composed of insolvent companies and active companies. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix 10: Test of Equality of Means Between Active and Insolvent Companies - After PSM 1 to 5

Ratios	Difference between means	t value	Pr(|T| > |t|)
X1	-1.442	8.773	0,000***
X2	-0.019	6.195	0,000***
X3	-0.847	11.482	0,000***
X4	-0.524	10.591	0,000***
X5	0.498	-320.000	0,000***
X6	-0.122	67.668	0,000***
X7	-0.095	62.425	0,000***
X8	-0.144	22.190	0,000***
X9	-0.085	60.521	0,000***
X10	1.089	-2.899	0,003***
X11	-2.312	1.718	0,085*
X12	-12.374	7.127	0,000***
X13	-0.402	59.042	0,000***
X14	0.281	-51.662	0,000***
X15	-81.998	6.049	0,000***
X16	-128.280	6.112	0,000***
X17	-0.738	21.403	0,000***
X18	-0.144	8.442	0,000***
X19	-0.861	0.821	0.007***
Source: Author			

This table presents the number of observations, mean, standard deviation and median for the variables, for the data set after PSM 1 to 5. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

*** p<0.01, ** p<0.05, * p<0.1

Appendix 11: Pearson`s Correlation Coefficients for the Financial Ratios - Before PSM

Before PSM	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X16	X17	X18	X19	X20
X1	1.0000																		
X2	0.1605*	1.0000																	
X3	0.7677*	0.1295*	1.0000																
X4	0.6686*	-0.0984*	0.8365*	1.0000															
X5	0.0167*	0.0148*	0.0824*	0.0925*	1.0000														
X6	0.0104*	-0.0752*	0.1129*	0.1328*	0.7101*	1.0000													
X7	0.0172*	-0.0763*	0.1182*	0.1358*	0.6997*	0.9853*	1.0000												
X8	0.0411*	0.0113*	0.1050*	0.1082*	0.5021*	0.6045*	0.6062*	1.0000											
X9	0.0509*	0.0311*	0.1271*	0.1332*	0.7622*	0.9020*	0.9116*	0.6691*	1.0000										
X10	-0.0520*	-0.0089*	-0.0670*	-0.0606*	-0.0142*	-0.0276*	-0.0252*	0.0107*	-0.0176*	1.0000									
X11	-0.0229*	0.0405*	-0.0408*	-0.0470*	0.0180*	-0.0043*	0.0114*	0.1483*	0.0565*	0.0813*	1.0000								
X12	-0.0214*	0.0738*	-0.0468*	-0.0577*	-0.0092*	-0.0365*	-0.0270*	0.1198*	0.0371*	0.0911*	0.4417*	1.0000							
X13	0.2118*	0.1083*	0.2820*	0.2667*	0.4284*	0.5356*	0.5564*	0.4689*	0.6289*	-0.0588*	0.0011	-0.0115*	1.0000						
X14	-0.2792*	-0.1332*	-0.3397*	-0.3035*	-0.2546*	-0.3466*	-0.3737*	-0.3330*	-0.4202*	0.1033*	0.0344*	0.0473*	-0.8646*	1.0000					
X16	0.0631*	-0.0004	0.0943*	0.1144*	0.1779*	0.1913*	0.1945*	0.1550*	0.2210*	-0.0236*	-0.0022	-0.0141*	0.1733*	-0.1384*	1.0000				
X17	0.0629*	-0.0132*	0.0895*	0.1090*	0.1246*	0.1491*	0.1579*	0.1072*	0.1636*	-0.0220*	-0.0002	-0.0141*	0.1363*	-0.1159*	0.9378*	1.0000			
X18	0.4553*	0.0264*	0.4829*	0.4749*	0.1301*	0.1634*	0.1799*	0.1827*	0.2140*	-0.1299*	-0.0735*	-0.0849*	0.5555*	-0.6990*	0.1694*	0.1562*	1.0000		
X19	-0.1370*	-0.0416*	-0.0705*	-0.0537*	0.1920*	0.2558*	0.2303*	0.0801*	0.1956*	0.0069*	-0.0243*	-0.0313*	0.0322*	0.0395*	0.0478*	0.0299*	-0.1050*	1.0000	
X20	-0.0152*	0.0688*	0.0009	-0.0234*	0.0038	0.0010	0.0020	-0.0063*	0.0076*	-0.0016	0.0001	0.0027	0.0067*	-0.0101*	-0.0006	-0.0024	-0.0145*	0.0465*	1.0000

Source: Author

This table presents Pearson`s correlation coefficients for the initial sample composed of insolvent companies and active companies. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

* denotes statistical significance at 5% or inferior.

Appendix 12: Pearson`s Correlation Coefficients for the Financial Ratios After PSM 1 to 5

PSM 1 to 5	wX1	wX2	wX3	wX4	wX5	wX6	wX7	wX8	wX9	wX10	wX11	wX12	wX13	wX14	X15	X16	X17	X18	X19
X1	1.0000																		
X2	0.2784*	1.0000																	
X3	0.5650*	0.2031*	1.0000																
X4	0.4876*	-0.0533*	0.7500*	1.0000															
X5	0.1293*	0.2202*	0.2117*	0.1760*	1.0000														
X6	0.0992*	0.1550*	0.2172*	0.1887*	0.9643*	1.0000													
X7	0.1074*	0.1624*	0.2155*	0.1845*	0.9603*	0.9930*	1.0000												
X8	0.1459*	0.1305*	0.1704*	0.1387*	0.5984*	0.5824*	0.5730*	1.0000											
X9	0.1371*	0.2284*	0.2092*	0.1706*	0.9924*	0.9544*	0.9648*	0.5879*	1.0000										
X10	-0.0391	0.0077	-0.0544*	-0.0419	0.0388	0.0434	0.0463	0.0575*	0.0420	1.0000									
X11	0.0823*	0.0814*	-0.0295	-0.0120	0.0782*	0.0634*	0.0710*	0.1472*	0.0862*	0.1106*	1.0000								
X12	0.0520	0.1485*	0.0037	-0.0085	0.1044*	0.0807*	0.0899*	0.1478*	0.1141*	0.0832*	0.2880*	1.0000							
X13	0.2218*	0.3305*	0.3274*	0.2777*	0.7040*	0.6697*	0.6891*	0.4923*	0.7224*	0.0390	0.0410	0.1359*	1.0000						
X14	-0.2676*	-0.3348*	-0.3525*	-0.2948*	-0.5228*	-0.4998*	-0.5219*	-0.3870*	-0.5444*	-0.0231	-0.0195	-0.1131*	-0.9138*	1.0000					
X15	0.0584*	0.0248	0.1345*	0.1347*	0.2441*	0.2250*	0.2140*	0.1673*	0.2320*	0.0076	0.0570*	0.0372	0.1784*	-0.1316*	1.0000				
X16	0.0552*	-0.0183	0.1206*	0.1397*	0.1825*	0.1754*	0.1684*	0.1289*	0.1748*	0.0002	0.0412	0.0226	0.1367*	-0.1086*	0.8806*	1.0000			
X17	0.4464*	0.1732*	0.4889*	0.4680*	0.3235*	0.3111*	0.3204*	0.2730*	0.3317*	-0.0693*	-0.0160	0.0194	0.5772*	-0.6793*	0.1586*	0.1617*	1.0000		
X18	-0.1685*	-0.1285*	-0.0207	0.0164	0.1165*	0.1491*	0.1157*	0.1793*	0.0822*	0.0125	-0.0392	-0.0213	0.0226	0.0522	0.0054	0.0162	-0.0697*	1.0000	
X19	-0.0129	0.0336	0.0376	0.0152	0.0904*	0.1030*	0.0991*	0.0281	0.0867*	-0.0219	-0.0271	0.0114	0.0824*	-0.0507	0.0138	-0.0193	0.0334	0.1048*	1.0000

Source: Author

This table presents Pearson`s correlation coefficients for the data set composed of the insolvent companies with data from one year prior to insolvency, and the active companies matched to these insolvent companies by PSM 1 to 5. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

* denotes statistical significance at 5% or inferio

Appendix 13: LDA - Ranking of the Standardised Coefficients

Rank	PSM 1 to 1	Standardised Coefficients (absolute values)	PSM 1 to 5	Standardised Coefficients (absolute values)	PSM 1 to 10	Standardised Coefficients (absolute values)	Without PSM - random 80% of active companies	Standardised Coefficients (absolute values)
1	**X9**	0.6173	**X7**	1.5629	**X5**	2.3057	**X6**	3.2732
2	**X7**	0.5320	**X6**	0.7632	**X9**	2.1858	**X7**	2.9294
3	**X14**	0.5184	**X9**	0.4128	**X6**	0.9968	**X5**	2.2761
4	**X12**	0.3952	**X14**	0.3854	**X7**	0.3840	**X9**	1.7331
5	**X16**	0.3897	**X18**	0.2635	**X14**	0.3771	**X2**	0.0784
6	**X15**	0.3072	**X12**	0.2230	**X18**	0.2022	**X17**	0.0463
7	**X18**	0.2667	**X2**	0.1378	**X2**	0.1761	**X12**	0.0387
8	**X6**	0.2371	**X8**	0.1284	**X8**	0.1664	**X4**	0.0368
9	**X17**	0.2095	**X4**	0.1220	**X12**	0.1401	**X18**	0.0362
10	**X8**	0.1569	**X16**	0.1220	**X15**	0.1263	**X16**	0.0268
11	**X13**	0.1270	**X17**	0.1171	**X16**	0.1196	**X14**	0.0265
12	**X4**	0.1135	**X15**	0.1013	**X13**	0.1004	**X3**	0.0257
13	**X5**	0.1041	**X11**	0.0958	**X4**	0.0986	**X8**	0.0239
14	**X2**	0.1013	**X13**	0.0642	**X3**	0.0740	**X13**	0.0212
15	**X11**	0.0933	**X5**	0.0610	**X11**	0.0702	**X10**	0.0175
16	**X3**	0.0823	**X1**	0.0584	**X10**	0.0568	**X1**	0.0135
17	**X1**	0.0691	**X3**	0.0447	**X17**	0.0468	**X11**	0.0069
18	**X19**	0.0493	**X10**	0.0330	**X19**	0.0316	**X15**	0.0042
19	**X10**	0.0031	**X19**	0.0249	**X1**	0.0303	**X19**	0.0005

Source: Author

This table presents the LDA standardised coefficients in absolute values for the data sets composed as follows: PSM 1 to 1 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 1; PSM 1 to 5 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 5; PSM 1 to 10 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 10; Without PSM – random 80% of active companies = insolvent companies with data from one year prior to insolvency and randomly selected 80% of the active companies from the same years as the insolvent. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

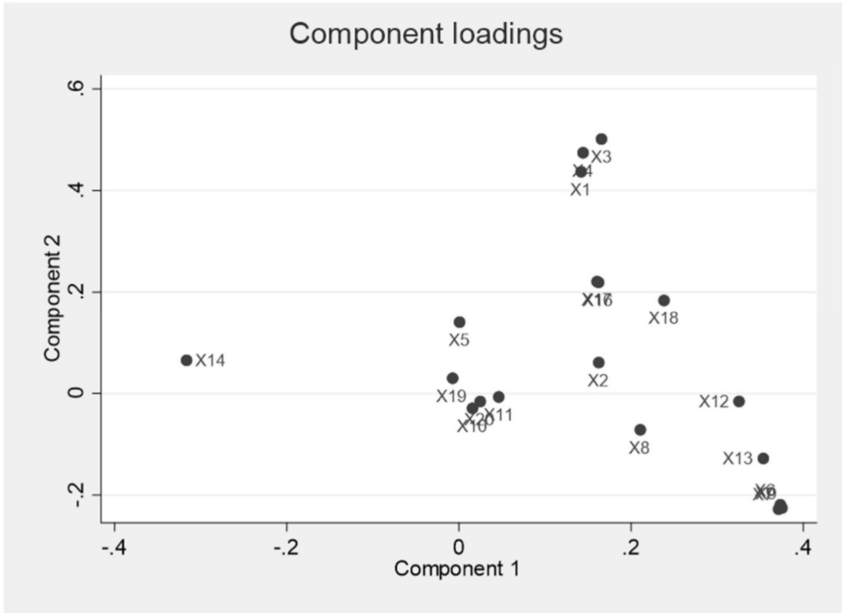
Appendix 14: LDA - Ranking of the Structure Coefficients

Rank	PSM 1 to 1	Structure Coefficients (absolute values)	PSM 1 to 5	Structure Coefficients (absolute values)	PSM 1 to 10	Structure Coefficients (absolute values)	Without PSM - random 80% of active companies	Structure Coefficients (absolute values)
1	X12	0.7015	wX13	0.8092	X13	0.8133	X5	0.3741
2	X14	0.6641	wX7	0.7685	X9	0.8038	X6	0.0645
3	X13	0.6531	wX6	0.7501	X7	0.7981	X7	0.0516
4	X7	0.6409	wX9	0.7437	X5	0.7641	X13	0.0513
5	X6	0.6181	wX14	0.7305	X6	0.7587	X9	0.0499
6	X9	0.6136	wX5	0.7272	X14	0.7149	X14	0.0462
7	X17	0.6122	wX8	0.5450	X8	0.6029	X8	0.0344
8	X8	0.4185	wX17	0.4760	X17	0.4072	X17	0.0299
9	X3	0.3486	wX3	0.2725	X3	0.2272	X12	0.0215
10	X4	0.3304	wX12	0.2606	X2	0.2270	X4	0.0209
11	X1	0.2781	wX2	0.2550	X12	0.2159	X3	0.0203
12	X2	0.2520	wX4	0.2527	X4	0.2076	X1	0.0174
13	X15	0.2303	wX18	0.1957	X18	0.1586	X15	0.0115
14	X16	0.2125	wX1	0.1626	X15	0.1471	X16	0.0114
15	X18	0.1518	wX15	0.1581	X1	0.1397	X18	0.0098
16	X11	0.1219	wX16	0.1482	X11	0.1376	X10	0.0094
17	X5	0.0523	wX11	0.1480	X16	0.1264	X2	0.0037
18	X19	0.0511	wX19	0.0936	X19	0.0828	X11	0.0028
19	X10	0.0300	wX10	0.0686	X10	0.0597	X19	0.0007

Source: Author

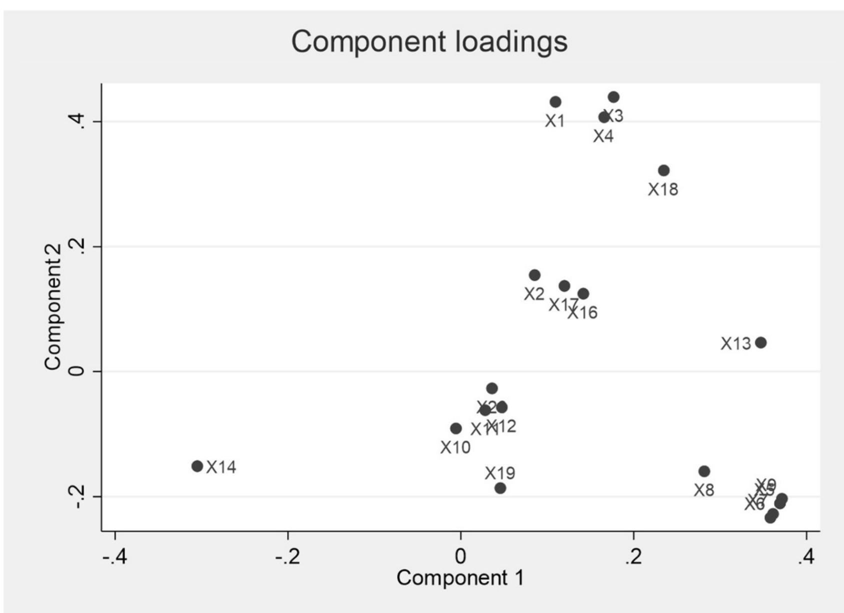
This table presents the LDA structure coefficients in absolute values for the data sets composed as follows: PSM 1 to 1 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 1; PSM 1 to 5 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 5; PSM 1 to 10 = insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 10; Without PSM – random 80% of active companies = insolvent companies with data from one year prior to insolvency and randomly selected 80% of the active companies from the same years as the insolvent. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 15: Component Loadings - PSM 1 to 1



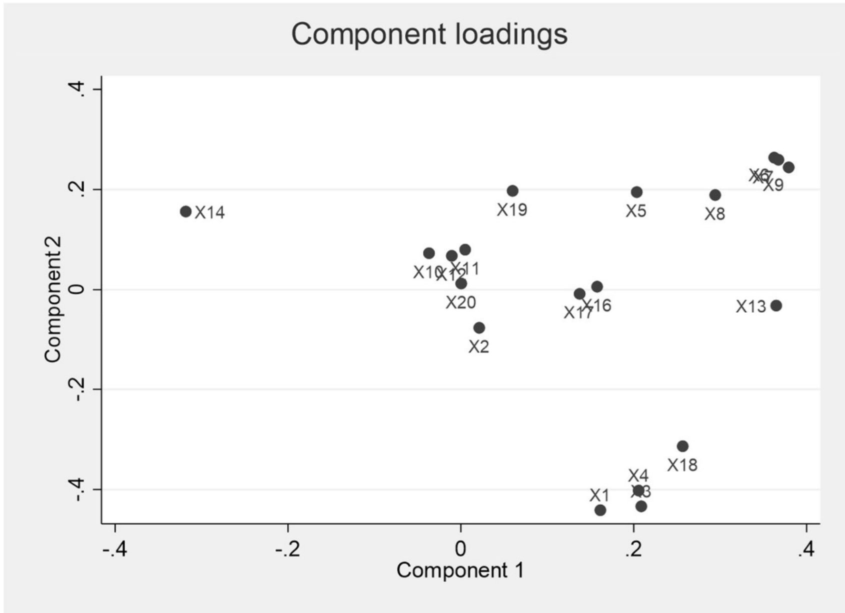
X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 16: Component Loadings - PSM 1 to 10



X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 17: Component Loadings - Random 80% Selection of Active Companies



X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 18: Logit regression - X-standardised coefficients (in log-odd units) PSM 1 to 1

PSM 1 to 1	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
X1	-0.0903	0.1669	-0.54	0.589	-0.4175	0.2369
X2	-0.2426	0.1807	-1.34	0.179	-0.5967	0.1115
X3	-0.0836	0.2730	-0.31	0.759	-0.6186	0.4514
X4	-1.1815	0.5702	-2.07	0.038	-2.2990	-0.0640
X5	0.3738	0.1545	2.42	0.016	0.0711	0.6766
X6	1.1911	1.9261	0.62	0.536	-2.5839	4.9662
X7	2.1956	1.2866	1.71	0.088	-0.3261	4.7173
X8	-0.0980	0.3138	-0.31	0.755	-0.7131	0.5172
X9	-2.4383	2.2737	-1.07	0.284	-6.8947	2.0180
X10	0.0297	0.1078	0.28	0.783	-0.1816	0.2410
X11	-0.0748	0.1150	-0.65	0.515	-0.3002	0.1506
X12	-2.3255	0.6165	-3.77	0.000	-3.5337	-1.1173
X13	-1.5076	1.0146	-1.49	0.137	-3.4963	0.4810
X14	1.2932	0.9292	1.39	0.164	-0.5280	3.1143
X15	-0.1651	0.9585	-0.17	0.863	-2.0437	1.7134
X16	-0.7349	1.6031	-0.46	0.647	-3.8769	2.4071
X17	-0.4213	0.3941	-1.07	0.285	-1.1936	0.3511
X18	-0.5017	0.1920	-2.61	0.009	-0.8780	-0.1254
X19	-0.1033	0.1172	-0.88	0.378	-0.3330	0.1263
Source: Author						

This table presents the X-standardised coefficients of logit regression for the data set composed of insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 1. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 19: Logit regression - X-standardised coefficients (in log-odd units) PSM 1 to 5

PSM 1 to 5	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
X1	0.0287	0.0835	0.34	0.731	-0.1350	0.1924
X2	-0.2168	0.0827	-2.62	0.009	-0.3789	-0.0548
X3	0.1663	0.1133	1.47	0.142	-0.0558	0.3885
X4	-0.2028	0.1159	-1.75	0.08	-0.4300	0.0244
X5	-11.8913	5.5498	-2.14	0.032	-22.7687	-1.0140
X6	14.3146	5.5992	2.56	0.011	3.3403	25.2889
X7	-15.2865	5.6609	-2.7	0.007	-26.3816	-4.1914
X8	-0.1482	0.1157	-1.28	0.201	-0.3750	0.0787
X9	11.4884	5.5968	2.05	0.04	0.5189	22.4578
X10	-0.0553	0.0643	-0.86	0.39	-0.1814	0.0708
X11	-0.0806	0.0673	-1.2	0.231	-0.2125	0.0514
X12	-0.1353	0.0672	-2.01	0.044	-0.2669	-0.0036
X13	-0.0462	0.2386	-0.19	0.846	-0.5139	0.4215
X14	1.1033	0.2519	4.38	0.000	0.6096	1.5970
X15	0.2969	0.2153	1.38	0.168	-0.1251	0.7188
X16	-0.2643	0.2159	-1.22	0.221	-0.6874	0.1589
X17	0.2849	0.1153	2.47	0.014	0.0588	0.5110
X18	-0.2293	0.0769	-2.98	0.003	-0.3799	-0.0786
X19	-0.0201	0.0656	-0.31	0.759	-0.1487	0.1085
Source: Author						

This table presents the X-standardised coefficients of logit regression for the data set composed of insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 5. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 20: Logit regression - X-standardised coefficients (in log-odd units) PSM 1 to 10

PSM 1 to 10	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
X1	-0.0330	0.0578	-0.57	0.568	-0.1463	0.0803
X2	-0.1525	0.0547	-2.79	0.005	-0.2597	-0.0453
X3	0.0799	0.0765	1.04	0.296	-0.0700	0.2298
X4	-0.0838	0.0751	-1.12	0.264	-0.2310	0.0634
X5	-1.6318	2.9039	-0.56	0.574	-7.3233	4.0597
X6	3.6044	3.0449	1.18	0.237	-2.3634	9.5722
X7	-4.0250	3.0094	-1.34	0.181	-9.9232	1.8733
X8	-0.2168	0.0832	-2.61	0.009	-0.3798	-0.0537
X9	1.4127	2.8730	0.49	0.623	-4.2182	7.0436
X10	-0.0628	0.0446	-1.41	0.159	-0.1503	0.0246
X11	-0.0439	0.0471	-0.93	0.352	-0.1363	0.0485
X12	-0.0625	0.0466	-1.34	0.179	-0.1539	0.0288
X13	-0.0726	0.1458	-0.5	0.619	-0.3583	0.2132
X14	0.5646	0.1550	3.64	0.000	0.2607	0.8685
X15	0.1554	0.1276	1.22	0.223	-0.0947	0.4055
X16	-0.1247	0.1255	-0.99	0.320	-0.3706	0.1212
X17	0.2220	0.0772	2.88	0.004	0.0708	0.3732
X18	-0.1187	0.0493	-2.41	0.016	-0.2153	-0.0222
X19	-0.0162	0.0439	-0.37	0.712	-0.1022	0.0698
Source: Author						

This table presents the X-standardised coefficients of logit regression for the data set composed of insolvent companies with data from one year prior to insolvency and corresponding active companies matched to them by PSM 1 to 10. X1, Current Ratio; X2, Working Capital to Total Assets; X3, Quick Ratio; X4, Cash Ratio; X5, Current Assets to Total Assets; X6, EBIT to Total Assets; X7, Cash Flow to Total Assets; X8, Operating Profit Margin; X9, ROA; X10, Debt to Equity; X11, Debt to EBITDA; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X14, Debt to Asset;; X15, Interest Coverage; X16, EBITDA to Interest Coverage; X17, Equity to Debt; X18, Total Assets Turnover; X19, Working Capital Turnover.

Appendix 21: Probit regression - Marginal Effects PSM 1 to 1

VARIABLES	Marginal effects
X1	-0.003
	(0.006)
X2	-0.136
	(0.103)
X3	-0.010
	(0.017)
X4	-0.217**
	(0.090)
X5	0.265**
	(0.118)
X6	0.752
	(1.372)
X7	2.690***
	(0.977)
X8	0.012
	(0.048)
X9	-1.947
	(1.643)
X10	0.000
	(0.001)
X11	-0.000
	(0.000)
X12	-2.435***
	(0.402)
X13	-0.252
	(0.170)
X14	0.145
	(0.229)
X15	0.000
	(0.000)
X16	-0.001**
	(0.000)
X17	-0.189***
	(0.065)
X18	-0.094***
	(0.036)
X19	-0.001
	(0.000)
Observations	578

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 22: Probit regression - Marginal Effects PSM 1 to 10

VARIABLES	Marginal effects
X1	-0.000 (0.002)
X2	-0.051*** (0.018)
X3	0.003 (0.005)
X4	-0.027* (0.015)
X5	0.527 (3.901)
X6	-0.005 (3.925)
X7	-0.436 (3.898)
X8	-0.009 (0.016)
X9	-0.352 (3.886)
X10	-0.000 (0.000)
X11	-0.000 (0.000)
X12	-0.000** (0.000)
X13	-0.024 (0.019)
X14	-0.008 (0.033)
X15	0.000 (0.000)
X16	-0.000 (0.000)
X17	-0.062*** (0.017)
X18	-0.026*** (0.007)
X19	-0.000 (0.000)
Observations	2,474

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 23: Coefficient Estimates for Model 3 and Model 4

VARIABLES	Model 3	Model 4
X5	3.632	43.515***
	(2.251)	(2.409)
X7	-9.365***	-34.257***
	(2.256)	(2.263)
X12	-0.006***	-0.021***
	(0.002)	(0.006)
X13	-0.358	-0.851
	(0.240)	(0.541)
X18	-1.646***	-0.889
	(0.300)	(0.543)
Constant	-0.771***	-10.417***
	(0.116)	(0.526)
Observations	1,403	188,896
Pseudo R-squared	0.335	0.948
Prob > chi2	0.000	0.000
Source: Author		

This table contains the estimation results for the logit Models 3 and 4 .The dependent variable equals zero if the firm is not financially distressed and one otherwise. The column Model 3 contains the results of the estimation using the data set composed by insolvent companies with data one year prior to insolvency and active companies matched to the insolvent ones by PSM 1 to 5. The column Model 4 contains the results of the estimation using the data set composed by insolvent companies with data one year prior to insolvency and a random selection of 80% of all active companies from the same years as the insolvent ones. X5, Current Assets to Total Assets; X7, Cash Flow to Total Assets; X12, Cash Flow to Debt; X13, Retained Earnings to Total Assets; X17, Equity to Debt. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1