U.PORTO

FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

Start-ups: A study to test survivorship and bankruptcy forecasting models

Carlos Manuel Peixoto da Silva

Dissertation
Master in Finance

Supervised by
Professor Miguel Augusto Gomes Sousa, PhD

2018

Biography

Carlos Manuel Peixoto da Silva was born in 1990 in Fafe, Portugal. He holds a degree in Economics from Universidade Lusíada do Porto, and a Master's degree in Economics and Innovation Management from Faculdade de Economia da Universidade do Porto. As a student, he has held several important positions in academic life, such as the positions of President of the Assembly and President of the Board of the Student Group stand out, giving him the responsibility of managing a department and coordinating several events.

During the period in which he completed his academic training, he completed a part-time internship in the commercial area of ActivoBank. During his academic career, he was a full-time analyst in the Business Intelligence & Development department at SONAE. In the exercise of these last functions, he worked directly with the direction of MO and ZIPPY and with its various departments, such as logistics, markets, operations and quality-control, among others, which allowed him to acquire solid knowledge on various business areas. Together with the aforementioned functions, he has managed a rural tourism accommodation in the city of Fafe, Casa de Fora, where he performed management functions, namely definition of promotional policies and prices, management of reservations, and contact with customers and suppliers.

Finally, it is worth mentioning the collaboration he has been providing in a family business, Gomes da Silva - Mediação de Seguros, where he performs management and monitoring of a portfolio of clients, as well as contract management.

Given his interest in the areas of corporate finance, investments, venture capital, investment strategies, he decided to apply for the Master's Degree in Finance in order to leverage his analytical skills.

Abstract

Early stage companies are widely recognized in the literature due to their high failure rate (Huynh et al., 2017). According to Instituto Nacional de Estatística (INE), about 91% of companies which were established one year before, managed to survive and remain in the market, and by the end of five years, only about half of the companies subsisted. It is possible to conclude that, if it is possible to understand why start-ups fail, it may be possible to increase their survival rates.

Having established that, in this investigation 59 independent variables were analysed, with information about coverage, profitability, liquidity, leverage and sector. In total, there were 207 companies and through a logit regression technique a model to evaluate the probability of survival of each company according to their ratios was suggested. The results of the model can correctly classify about 94% of Start-ups, which is almost 16% larger than a generic corporate Z-score model of predicting bankruptcy companies.

The conclusions that have been reached allow us to point out that from the financial point of view, the higher the account receivable to total liabilities, the higher the ROCE and the greater the solvency ratio, the greater the probability of survival of a company.

On the other hand, start-ups whose volume of intangible assets to total assets was higher, including the sectors of human health and social support activities or manufacturing industries, presented statistical evidence of greater risk of bankruptcy.

Keywords: Start-ups, entrepreneurship, success, failure, prediction model, credit risk model, SMEs; corporate bankruptcy prediction, risk assessment, forecasting models.

JEL Codes: G11, L25, L26, M13

Resumo

As empresas em estágio inicial são amplamente reconhecidas na literatura devido a sua alta taxa de falência (Huynh et al., 2017). Segundo o Instituto Nacional de Estatística (INE), cerca de 91% das empresas que nasceram um ano antes conseguiram sobreviver e manter-se no mercado um ano após a sua criação e no final de cinco anos de atividade, apenas cerca de metade das empresas subsistiam. É possível concluir que, se pudermos entender melhor por que razão as start-ups falham, devemos ser capazes de aumentar a taxa de sobrevivência.

Urge, por isso, a necessidade de analisar e compreender o impacto de fatores explicativos no sucesso ou insucesso destas novas firmas. Nesta investigação, foram analisadas 59 variáveis independentes, com informações sobre cobertura, rentabilidade, liquidez, alavancagem e setor de 207 empresas e por meio de uma técnica de regressão logística propomos um modelo para avaliar a probabilidade de sobrevivência de cada empresa segundo os rácios e sector onde a empresa se insere. Os resultados do modelo podem classificar corretamente cerca de 94% das start-ups, o que é uma previsão cerca de 16% superior ao modelo de Z-score para empresas não listadas.

As conclusões a que foi possível chegar permitem evidenciar que do ponto de vista financeiro, quanto maior a rubrica contas a receber sobre total do passivo, quanto maior o ROCE e maior o índice de solvabilidade, maior a probabilidade de sobrevivência de uma empresa.

Por outro lado, start-ups cujo volume de ativos intangíveis sobre ativos totais foram superiores, bem como empresas dos setores de atividade saúde humana e apoio social ou indústrias de transformação apresentaram evidências estatísticas de maior risco de falência.

Acknowledgments

This thesis is the result of a great commitment from several people, whom I would like to thank. Surely, I cannot thank all the people (who in different ways were involved in this project) enough. Nevertheless, I would like to express my deepest gratitude.

To Professor Miguel Sousa for supervising this dissertation, with rigor, constant availability and precious knowledge transmitted.

I would like to thank my whole family, especially my parents, brothers and godparents, for the support they gave me, their actions and encouragements were crucial.

To my girlfriend for the ideas and contributions to the writing of the present dissertation.

To my friends that supported, motivated and encouraged me to pursue this project.

Last but not least, to all the Masters in Finance's professors for the wise guidance, availability and for the valuable assistance throughout this journey.

Content Index

Figures and Table index

Glossary of Terms, Abbreviations and Acronyms

- Academic Spin-Off (ASO)
- Accelerated failure time (AFT)
- Best linear unbiased estimator (BLUE)
- Classificação Atividades Económicas (CAE's)
- Comissão do Mercado de Valores Mobiliários (CMVM)
- Earnings Before Interest, Taxes (EBIT)
- Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA)
- Gross national product (GNP)
- Instituto Nacional de Estatística (INE)
- Multiple discriminant analysis (MDA)
- Natural Logarithm (ln)
- Net Present Value (NPV)
- Ordinary least squares (OLS)
- Ordinary Least Squares (OLS)
- Percentage (%)
- Probability of default (PD)
- Probability of survival (PS)
- Research and development (R&D)
- Retained Earnings (RE)
- Return on Capital Employed (ROCE)
- Return on Equity (ROE)
- Small and Medium-size Enterprises (SMEs)
- Total Assets (TA)
- Venture Capital (VC)

1. Introduction

The scenario of failure of new companies is not new, or a temporary effect. A number of prior studies have demonstrated that failure rate has been keeping high. The findings of Mata and Portugal (1994) report that 20% of new firms born one before died during the first year of their activity and that only about 50% keep active in the fourth year after the foundation. Some year later, Cabral and Mata (2003) found that from 2651 firms established in 1984 only 1031 were still active in 1991.

Therefore, understanding the economic and financial phenomena and their importance to explain the survivorship of early-stage companies are the main objectives of this investigation.

More specifically, a set of financial and economic ratios will be investigated and applied to a set of Portuguese firms based on science and technology parks and incubators in order to understand key factors that promote the success or failure of new companies and how they contribute to redefining eligible investment criteria. At the end of this study, a set of answers are expected:

- Measures, control and the improvement of start-ups' survival;
- Identify financial factors that are predictors of a start-up's bankruptcy risk;
- Selection of target companies to invest taking into account its determinants.

As far as this matter is concerned, several papers have been conducted on the bankruptcy of companies. It has recently become a "hot topic" because researchers have failed to understand, explain, and predict why some businesses succeed and others fail due difficulty in predicting which ventures will succeed or fail (Olaison and Sorensen, 2014). First studies came from Altman (1968) who investigate default prediction methodologies in manufacturing companies. Since then, publications in this field have been multiplied in order to understand which variables cause an impact on the success and failure of companies.

Although it has been investigated, the topic deserves even greater focus. Firstly, there are only a few number of studies in this field in Portugal and most of them are mainly focused on human capital determinants. The application of the methodology bankruptcy forecasting models can be seen as the major contribution of the study. Moreover, many of those studies

have not been updated, so the characteristics of companies may have changed, and the conclusions may not be valid.

Furthermore, the motivation for this study is due to the lack of consensus regarding the determinants that influence the business success and failure worldwide, and the great discrepancy in the literature as to which variables do in fact lead to success, reinforce this idea, stating that there is a need to test models in multiple countries to assess the robustness of the findings (Lussier and Halabi, 2010).

This information may be useful for potential and current entrepreneurs to assess their competitive position and design smarter strategies. It may also be valuable for other stakeholders in the context of new companies, particularly venture capital, private equity funds, business angels, and for those responsible for science and technology parks and business incubators, to take strategies in order to increase the survivorship of new companies.

Regarding structuring, this thesis will be divided into several modules. In the first part after the summary, index, introduction, a literature review will be made, in which the subject's studies in Portugal, the most well-known bankruptcy forecasting models, as well as the determinants at the contextual, human and finance levels will be presented, in order to understand the reasons for the different success rates among new companies.

Section 3 presents an overview of the characterization of the sample, the criteria to be considered in the analysis, and a descriptive analysis. Section 4 presents the methodology, namely the model selection. Section 5 presents the results, validation of the model and their discussion. Finally, the last section summarizes the main conclusions of this investigation.

2. Literature Review

The role of early-stage companies has been highlighted in economics and management since the first studies. Schumpeter (1934) has informed us about how creative destruction introduced by new companies or innovative ones are a driving force of economic growth. The author informs us about new companies' importance, namely in the creation of new jobs and generation of new wealth for society and point this as a reason why institutions and governments should care about new companies.

The role of new companies is not limited to academic concern. Even governments and institutions are aware of the importance of new companies. For instance, Basel II agreement, whose main objective is to ensure the stability of the global financial system, in particular, how they make their risk assessment in lending activities, imposes very rigid criteria how banks should calculate credit risk and start-up rating before granting any credit.

Over the last decades, several studies have been developed in order to understand which determinants predict the success or failure of enterprises and evaluate their performance, but there is no generally accepted list of variables. Numerous explanatory variables for business success or failure were studied, which include human capital determinants, characteristics of companies and financial prospects. To be able to understand the topic, it's important to understand how researchers define start-ups, the limits of the concept and the difference to other companies.

2.1 Start-up concept

The concept of start-up has been defined according to a set of several parameters: age, business model, sector, profitability, growth metrics and other categories. Blank and Dorf (2012) defined a start-up as a temporary organization formed to search for a repeatable and scalable business model. Ries (2011) defines a start-up as an institution designed to deliver a new product or service under conditions of extreme uncertainty. Based on this, it seems that Start-ups have several types of different characteristics of all new ventures, characterized by the exploration of innovative business models or explore new products and services that can cause an impact on the level of risk. To incorporate this, companies will be consider according to its primary industry code, based on "Classificação das Atividades Económicas" (CAE's) that will be later detailed in Data.

2.2 Studies in Portugal

Teixeira and Grande (2013) look into 101 Academic Spin-off (ASO), which are quite similar to Start-up[1], both in terms of risks, industry, and profile. The working paper found that in terms of human capital more qualified labours, and founders with economics or managerial degree, support significant and positive relation to performance. In terms of start-up profile, older firms and exporting firms tend to present a stronger positive impact on their economic performance. Nevertheless, the number of founders, Research and development (R&D), and patents have no statistically significant impact on spin-off performance.

The approach of Gonçalves et al. (2014) was based on human, financial and industry determinants considering a sample of 1430 Start-ups. The findings suggest that will lower the default probability, greater company's ability to generate earnings before interest, taxes, depreciation and amortizations (EBITDA) that covers net debt. Furthermore, founders with a high level of education and management and business experience prove to be more likely to succeed.

Silva, et al. (2016) studied 50 start-ups established over the period from 2003 to 2015 and investigate the business success according to management experience, industry experience, marketing skills, age, education, capital, existence of professional advisors, partners, product or service timing and economic timing. Regarding the final out, the authors found that start-ups with young founders, founders with basic or less education, and with marketing skills are more likely to be unsuccessful cases. However, the findings that founders with marketing skills are more likely to be unsuccessful cases are opposing the previous literature. For instance, Lussier (1995) found that companies whose founders have marketing skills should have better chances to survive. Furthermore, Silva, et al. (2016) claims that Portuguese start-ups which have professional advisors have a greater chance of being well succeeded than other companies.

These two working papers are very relevant to this investigation because both of them follow bankruptcy prediction models. The former one used several variables from Altman (1968)

[1] Both a spin off and a start-up develops from an innovative idea, the difference is where it comes from. In the case of spin off, the project is derived from an organization already formed, be it a company (corporate spin-off) or university (academic spin-off).

and the latter used some of Lussier's (1995) variables, who are quite relevant on literature. Nevertheless, in both studies, the methodology of data collection was done by the survey.

2.3 Financial Determinants

A lot of researchers demonstrate that financial determinants play an important role in the assessment of the risk of new companies.

One of the first known studies on default prediction models came from Beaver (1966). The study had a sample of 158 firms, of which 50% failed and 14 financial ratios were investigated. After two years, Altman (1968) analysed the same topic in a sample based on 66 manufacturing firms, of which 33 failed and 33 did not fail. In total, 22 potentially explanatory financial ratios were analyzed and classified into five standard ratios categories, including liquidity, profitability, leverage, solvency and activity ratios. Overall, this was well-accepted because it was a measure of the net liquid asset, cumulative profitability over time, the productivity of the firm's assets, abstracting from any tax or leverage factors and ability to generate sales per asset.

A few years later Ohlson (1980) applied the model to the default prediction's study in a study of 2163 companies, composed by 105 bankrupt firms. He based the analysis on nine predictors, 7 financial ratios and 2 binary variables, which include:

- Size Total assets to Gross national product (GNP) price-level index;
- Total liabilities to total assets;
- Working capital to total assets;
- Current liabilities to current assets;
- Net income to total assets;
- Funds provided by operations to total liabilities;
- $\frac{Net\ Income\ t - Net\ Income\ t-1}{(Net\ Income\ t)+(Net\ Income\ t-1)}$; (2.1)
- Control variable: One if total liabilities exceeds total assets, zero otherwise;
- Control variable: One if net income was negative for the last two years, zero otherwise.

However, the models from Altman and Ohlson were valuable predictors, these models were estimated mainly for large companies. More recently, the methodology of default prediction model was adapted by the hand of Altman and Sabato (2007) to SMEs.

Altman and Sabato (2007) considered that in credit risk management, different methodologies should be implemented according to the type and the dimension of the company. These authors used a sample consisting of 2010 American SMEs with sales less than $65 million worth, including 120 defaults from 1994 to 2002. From a set of 22 financial ratios, the authors selected the ones with highest default event predictive accuracy, and adapted the model to new firms, with a degree of confident almost 30% higher than Z-Score. To increase the accuracy of the model, the authors propose some adjustments to prior models. For instance, due to the high variability among sectors in which companies operate, the authors advocated the use of logarithmic transformations to increase the prediction power[2]. With this adjustment, the accuracy ratio of the model jumped from 75% to 87% and decreased the type I error. In this particular case, reduced from 21% to about 12% the probability of attributing credit to a firm with a non-reliable credit risk. From the initial variables, the authors observed the ones with high accuracy and five ratios were selected [3]:

- Short-term debt to equity book value;
- Cash to total assets;
- EBITDA to total assets;
- Retained earnings to total assets;
- EBITDA to interest expenses.

In practical terms, the model predicts the probability of default (PD), specifically the one year PD, which can be applied in essential credit risk analysis under Basel II requirements. In addition, the authors conclude that the presence of qualitative variables, generally known as non-financial factors in SMEs, in general, improve the statistical accuracy of those models[4].

---

[2] For EBITDA/Total Assets (TA) the transformations was: -ln(1-EBITDA/TA); For Retained Earnings (RE) / TA was: -ln(1-RE/TA)

[3] **Leverage** [Short Term Debt/Equity (Book Value); Equity (Book Value)/Total Liabilities; Liabilities/TA**]**; **Liquidity** [Cash/Total Assets; Working Capital/TA; Cash/Net Sales; Intangible/TA**]**; **Profitability** [EBIT/Sales; EBITDA/ TA; Net Income/TA; Retained Earnings/ TA; Net Income/ Sales**]**; **Coverage[** EBITDA/ Interest Expenses; EBIT/Interest Expenses]; **Activity [** Sales/TA; Account Payable/Sales; Account Receivable/Liabilities**]**

[4] Notwithstanding this, qualitative information may not be available in our database, therefore only will be explored the variables accessible in that database.

From Cooper et al. (1994)'s point of view, the initial financial capital is a significant predictor of subsequent company's performance. According to them, the level of capitalization may act as a barrier against random shocks, allow companies to meet the financing demands, and allow to take more ambitious strategies.

Likewise, Huynh et al. (2012) propose that larger the initial size better should be the firm operations, in the sense that financial resources provide the necessary resources to develop the firm into a fully functioning establishment. Another idea supported by the researchers is that larger the initial size allows the firm to survive even with temporary negative profits. However, these conclusions are not valid for all studies. From the point of view of Astebro and Bernhardt (2005) was found negative and Manigart et al. (2002) found no correlation.

Another possible explanation pointed by Huynh et al. (2012) to the clarify the success or failure of new companies, is the debt-to-asset ratio. According to authors, more debt enables to increase size and take advantage of scale, therefore companies located in the lowers quintiles seems to have a negative relationship between initial debt-to-asset ratio and default rates, while companies with high initial debt-to-asset ratios, especially those in the located in fifth quintiles of the ratio may face higher financial constraints, which increased the chance of bankruptcy.

To identify the most important financial ratios Araghi and Makvandi (2013) investigated 189 companies during 2000 until 2010 and considering the correlation between variables, build a model with variables with the highest contribution to explanatory power. From 22 ratios 7 ratios were selected: return on equity (ROE), debt ratio, debt cover ratio, collection period, inventory turnover, and debt to equity, and product to working capital.

2.4 Company Determinants

The start-up's determinants are perceived as another's important predictors of business performance. At this matter, Baum and Silverman (2004) found a positive relationship between a venture capital (VC) investment in a start-up and its subsequent performance. Based on this theory, identify whether VC's are good at identifying promising start-ups, or instead of this, if they are good as coach, namely through the injection of knowledge and management skills on new companies can be a hypothesis to investigate.

According to Lee et al. (2001) innovation remained a powerful predictor of future growth. Intellectual property, i.e. patents or trademarks, allow new ventures to solely commercialize and differentiate themselves from incumbents, and therefore companies without intellectual property are more vulnerable to imitation and replication by competitors. According to Baum and Silverman (2004) firms with a patent are in a more favorable position to obtain VC financing, since a patent offers a temporary monopoly over the exploration of revenues. Considering this, even if temporary limited over time, in the same circumstances, companies with patents can be offered differentiation which, at least in theory can reduce the default rates.

Besides that, Baum and Silverman (2004) pointed out that ownership differences may also influence start-ups' performance, in particular, when start-ups are subsidiaries or spin-off of other companies. The authors support that in these cases, start-ups may have access to financial resources of their parent firm(s), as well as enjoy the reputation of their mother companies, and this may affect both their survival likelihood and their early rate of growth.

The size of the founding team is another variable and widely recognized determinant with potential impact on start-up's performance. From Miloud et al. (2012)'s point of view more founders allow more specialization in decision making, which can contribute to increasing the efficiency. To support this idea, Eisenhardt and Schoonhoven (1990) pointed out that the size of the founding team is positively related to new venture revenues growth. More founders can bring essential skills, however, it should be considered that a company with more founders may result in additional conflicts (Brinckmann and Hoegl, 2011).

Companies' size is another variable evoked by several authors as a predictor of company performance. According to Mata and Portugal (1994), larger firms have been founded to experience higher survival probabilities than smaller firms. One reason to justify its higher survival rates is due to the higher diversification or due to the lower quality managers in small companies (Geroski et al., 2009). There seems to be evidence that survival rates are likely to be lower for young firms, since they take time to develop specific knowledge and trust of the customers. In addition, Mata, et al. (1995) suggested that current size should be a better predictor than initial size. However, the authors consider that there are good reasons to include both, initial and current size in survival models. One reason to use both is that measuring initial size to current size as it can be applied as a measure of performance.

About the market dynamics, several studies indicate that companies founded during the years of economic recession have higher chances to goes to bankruptcy. According to Sikomwe et al. (2014) companies established during recession years have a greater chance to fail than the ones which were established during expansion periods. These findings are similar to the idea of the relationship between industry growth and a start-up survival (Mata and Portugal, 1994). Indeed, those companies that enter in growing industries may expect less competitive pressure, and may growth without inflicting competitor's market share, so it would be expected lower likelihood of aggressive reactions, and therefore it should be easier to survive. However it can be argued that recession can decrease the rate of creation of new firms, and in this sense, alleviate the pressure exerted upon established firms (Caballero and Hammour, 1994). Nevertheless, can be argue that it is not fair to compare companies created during the recession and the others created in expansion or growing industries.

2.5 Human Capital Determinants

Several authors argue that beyond financial and company determinants, start-ups firms are an extension of the founder's characteristics and debated the characteristics of the entrepreneurs as predictors of the performance (Chandler and Jansen, 1992).

In what concern to human capital determinants the experience and academic education are commonly cited as the most important predictors of companies performance. For instance, Peterson et al. (1983) argued that the major cause for small business failures is the lack of management expertise. According to Lee and Zhang (2010) more experienced founders may have higher survival chances, as well as better access to financial capital. Prior management experience on the establishment of the company provides to entrepreneur more skills, namely the interaction with different stakeholders, covering customers, investors and suppliers.

In addition, Lee and Zhang (2010) supported that founders with higher education can have better knowledge about specific and broad topics, which are recognized in literature as solid gains in terms of learning ability, business detection opportunities and organizational skills. According to them, the level of academic education may influence the access to information, which entrepreneurs are able to obtain and use to execute their ideas.

Another view is supported by Lussier and Pfeifer (2001), which argued that making a start-up to be successful cannot be necessarily obtained through formal education.

2.6 Research methodologies in literature

Depending on whether the objective is to explain or predict the probability of a company going bankrupt, different approaches can be adopted to create a bankruptcy-forecasting model. Among the most used approaches are Discriminant Analysis, Logistic Regression, Neural Networks and Decision Trees (Thomas, 2009).

Starting from the first studies on bankruptcy forecasting models, different methodologies were applied over time. Beaver (1966) classified companies in bankrupt/active as a function of cash-flow to total debt ratio. In this case, a univariate model was implemented to explain the dependent variable from a single independent variable.

However, researchers such as Altman (1968) criticized the univariate models with the argument that a ratio, analyzed in isolation, does not contain enough information. According to Altman (1968) point of view, different factors in interacting simultaneously is what leads to a corporate bankruptcy. Therefore, the author proposes a multivariate model to explain the dependent variable as a function of several independent variables and use this methodology in order to predict corporate bankruptcy. One advantage mentioned by the author of this type of analysis is to identify characteristics which distinguish the financial profile of bankrupt companies compared to active companies so that once you know the characteristics of a company it will be possible to predict to which group the company belongs. Since then MDA was applied by many authors (e.g. Lussier, 1995).

However, Altman and Sabato (2007) pointed out that the basic assumptions of MDA are often violated when applied to the default prediction models. One reason mentioned by the authors was that if model about the failure is considered, the dependent variable should consider two possible situations, failure or success. In this case, because the dependent variable is binary, the Ordinary Least Squares (OLS) method can no longer produce the best linear unbiased estimators (BLUE). Thus, in this case, the authors suggest the use of the Logit or Probit models, because they are more suitable to be used if the dependent variable is dichotomous. In addition to that, Lennox (1999) reported that Logit and Probit models showed higher accuracy than discriminant analysis.

Considering these advantages, Ohlson (1980) applied for the first time a Logit model to forecast the default in a study of 105 bankrupt firms and 2058 successful firms. Since then, Logit and Probit have been widely used in the investigation of business failure (e.g Lennox, 1999; Araghi and Makvandi, 2013; Altman and Sabato, 2007).

However, the magnitude of the accuracy between MDA and Logit and Probit seems to vary considerably across studies. For instance, Araghi and Makvandi (2013) compare MDA, with Logit and Probit. Discriminant analysis showed better performance than Logit and Probit Models in predicting bankruptcy of companies, but lower accuracy in predicting non-bankruptcy companies, and consequently lower general accuracy.

Another group of researchers has used the duration models to investigate the failure of the companies (e.g. Mata and Portugal, 1994; Shumway, 2001; Morris, 1997; Bellotti and Crook, 2007).

Duration models encompass a set of models and methods for the statistical analysis of data representing the lifespan from an initial state to the occurrence of a particular event. Survival analysis, or duration models, was initially designed in order to anticipate the duration of determining scenario until its occurrence, and the probability of determining the event to occur (Kleinbaum and Klein, 2002). The utility of the model is because it measures not only the probability of a business surviving through the "survival function", but also measures the probability of bankruptcy at any given time by the "hazard function" (Morris, 1997). Thus, analysis based on survival models allows predicting not only the probability of occurrence of a particular event but also the probability in several moments, and which are the ones more likely to occur. Assuming that behavior is dynamic, rather than estimating a start-up likelihood of going bankrupt or defaulting, this approach allows us to look in a deeper perspective. In this case, the objective is not only to estimate the probability of bankruptcy but to understand, for example, how long a certain company can overcome adversity before starting to fail to meet its financial obligations.

Stepanova and Thomas (2002) compared the results obtained through the survival models with other more traditional methods, such as logistic regression and neural networks, and concluded that models presented very similar performances. Although the results were similar, Stepanova and Thomas (2001) emphasized that another advantage of survival models is that they allow us to calculate profitability per operation, since the survival function allows

us to determine the probability of each client failing in several moments. Additionally, Mata and Portugal (1994) point out that to analyze the bankruptcy of new companies, typical statistical methods such as OLS are inadequate. One of the reasons pointed out by the authors is that during the research period a series of companies that have not failed yet, but will fail in the future. Hence, they argue that normal statistical methods can present biased and questionable estimates.

One of the drawbacks considered by Bellotti and Crook (2007) of simple duration analysis models is that it assumes that sooner or later all observations shift to non-survival or bankruptcy if applied to the business analysis. However, depending on the analysis, the proportion of bankruptcy can be greatly reduced. So, the assumptions of the model may be questionable.

As previously noted, there are several methodological approaches to attempting to predict corporate bankruptcy. Each predicting bankruptcy methodology can present its advantages and disadvantages, and therefore the most important is to select the most suitable methodology, taking into account the specific objectives of the study.

Taking into account what was mentioned, by the existence of information the methodology to be used will be the logistic regression. For a purely comparative effect, and not because it is considered that the proper methodology, a model of multiple-discriminant analysis will be presented a model.

3. Sample

This research uses a dataset of Start-up firms reported on the network Portus Park, which is an association that represents the Science and Technology Parks and Incubators in Portugal[5].

To avoid the survivorship bias and the simple fact of choosing companies that were disclosed in the Portus Park publications, would be, in essence, choosing companies that had survived in the prior years and thus had a better chance to persist than those who were not presented, it was sought on incubators' websites by the historical reports, which are generally disclosed once a year, to find out bankrupt companies that were no longer disclosed as soon as they have become inactive. Afterwards, all the financial information was extracted from the Amadeus database.

Initially, the sample consisted of 404 distinct companies, with 4040 observations related to the last 10 years of activity (even for those companies with less than 10 years of existence). Considering the diversity of companies extracted from the dataset, several criteria were applied to be considered in the analysis. Given the existence of a number of missing information or not applicable ratios in some of the variables, 1745 observations from firms with absent information and from companies which were missing financial information for ten or more ratios were erased. In addition, 40 observations concerning 4 companies that were branches of foreign companies were as well excluded from the sample. Finally, like Ohlson (1980) utilities, transportation, and financial services, such as banks, insurance brokers, and investment firms were excluded 1330 observations concerning 132 companies do not fall into the aforementioned sector of activity.

The final sample end up with 925 observations concerning 207 companies, considering financial information between 2005 and 2017. Default rates of the final sample were about 7% of the sample, close to the results presented in the literature. These results are close to the Ohlson's (1980), which presented a sample default rate of 4.87%, and Altman and Sabato's (2007) of 5.97%.

[5]More information about the database is available here: http://www.portuspark-network.org/pt/empresas

The following table shows how many companies end up with, the number of companies in default and non-default, the number of companies per sector, the average of total assets and the operating revenue per sector.

Table 1 - Number of companies per sector in total, in default and non-default situation. Average of total assets and operating revenues per sector.

Table 1 presents the number of companies in default and non-default, the number of companies per sector, the arithmetic average of total assets, and operating revenues. Total assets and the operating revenues are represented in thousands of euros.

Sector	Companies	Non-Default	Default	Average Total assets th EUR	Average Operating revenue th EUR
Agriculture, livestock, hunting, forestry and fishing (1)	3	3	0	208.76	141.28
Manufacturing (2)	28	27	1	1032.15	479.38
Wholesale and retail trade (3)	6	6	0	1668.70	2030.57
Information and communication activities (4)	118	109	9	1971.42	1538.33
Real estate activities (5)	1	1	0	631.20	869.15
Consulting, scientific, technical and similar activities (6)	44	41	3	506.99	425.06
Administrative and support services activities (7)	5	5	0	305.46	497.31
Human health activities and social support (8)	2	0	2	1890.03	2052.59
Total	**207**	**192**	**15**	**1457.63**	**1144.66**

In the following table will be summarized the net income, net income, EBITDA, total liabilities, ROE, EBIT margin per sector.

Table 2 - Average per sector of net income, ROE, EBITDA, EBIT margin and total liabilities.

This table reports the arithmetic average of net income, EBITDA, total liabilities, ROE and EBIT Margin. Total assets, operating revenues and total liabilities values are represented in thousands of euros. ROE and EBIT Margin are represented in %.

Sector	Average Net Income th EUR	Average EBITDA th EUR	Total Liabilities th EUR	Average ROE	Average EBIT Margin
Agriculture, livestock, hunting, forestry and fishing (1)	7.75	22.64	148.53	3.65%	-1.04%
Manufacturing (2)	-38.56	5.45	614.18	-26.40%	-8.81%
Wholesale and retail trade (3)	90.96	171.71	1121.28	21.96%	3.18%
Information and communication activities (4)	15.78	199.41	1648.42	-11.10%	0.33%
Real estate activities (5)	253.35	336.97	188.31	50.65%	26.64%
Consulting, scientific, technical and similar activities (6)	33.90	52.99	187.08	0.11%	2.38%
Administrative and support services activities (7)	13.01	30.51	228.37	1.42%	-6.12%
Human health activities and social support (8)	-73.60	48.47	1688.98	9.84%	-4.96%
Total	**15.91**	**135.79**	**1125.67**	**-8.27%**	**-0.17%**

As previously mentioned, the sample is mainly composed of non-default firms, which is most common in this type of studies. The average seniority of the companies on the sample was 3.87 years and the oldest company had at the time of the analysis 11 years.

The companies were grouped by primary industry code according to INE classification[6]. The number of start-ups in the information and communication activities' sector was 118. Consulting, scientific, technical and similar activities covered 44 firms. Manufacturing represents 28 start-ups, wholesale and retail trade, administrative and support services activities 5, agriculture, livestock, hunting, forestry and fishing, human health activities and social support 2, and real estate 1.

From the agriculture, livestock, hunting, forestry, and fishing sector, three companies integrate the sample, all of which survived. The sector presented a lower average of assets and revenues compared to its peers. On average companies presented a positive net income, EBITDA, and ROE.

Manufacturing sector includes 28 companies, one of which went bankrupt. The sector presents a low average of revenues per company, just ahead sector 1. On average the group presented a negative net income, an ROE of -26.40% and an EBIT Margin of -8.81%.

Wholesale and retail trade comprises six companies, all of them non-default during the analysis. Its average revenue was 2.031 thousands of euros, only surpassed by human health and social support sector. On average the sector generates a net income and positive EBITDA, ROE of 21.96% and an EBIT Margin of 3.18%.

Information and communication activities sector encompass 118 companies, 9 of which have failed. The sector presented the highest average of total assets. The average of net income and EBITDA is positive, although it is noteworthy the difference between both metrics, which suggest these companies spent a large amount of their earnings on interest, depreciations and amortizations.

[6] The industry to sector classification were done according to the same criteria followed in CAE-Rev.3d by INE on page 39
https: //www.ine.pt/ine_novidades/semin /cae/CAE_REV_3.pdf

From real estate activities, one survival company integrates the sample of the group. ROE reached almost 51% and EBIT Margin of 27%, with net profit and a profitability of 253 and 337 thousand euros.

Consulting, scientific, technical and similar activities contained a sample of 44 companies, 3 of which did not survive. The sector is characterized as a low value of assets, and revenues. The average earnings in the sector was around 34 thousand euros per company.

Administrative and support services activities comprise 5 different companies, with no default observation. The sector presented on average a low value of assets and revenues when compared to other groups. The average net income was around 13000 thousand euros, with an ROE of 1.42% and a negative EBIT margin of 6.12%.

Finally, Human health activities and social support sector included 2 companies, both of them defaulted. Although the EBITDA of those companies was positive when tax, interest, depreciation and amortization were considered, the average net income of these companies was negative.

As it can be noted above, there are marked differences between sectors of activity. For example, although all of them had positive EBITDA on average, manufacturing and human health activities and social support showed a negative net income. Considering the significant differences among sectors a dummy variable will be tested.

4. Methodology

In this thesis, to predict start-ups default, two different statistical models were used, logistic regression model and OLS model.

For the logistic regression models the estimation was based on the following expression:

$$\text{Probability of survival} = \frac{e^{Zit}}{1+e^{Zi}}, \text{ being, } Zi = C + \beta Xi' + Ei \quad (4.1)$$

To describe the OLS model the mathematical equation is given by the following equation:

$$Zi = C + \beta Xi' + Ei. \quad (4.2)$$

Where C is the constant, β is the vectors of estimated coefficients, Xi' is the vector of ratios/variables of company i, and Ei is the error term.

Considering Hair et al. (1998) if the model predicted a probability greater than 0.5, then the prediction of the model classifies that the company remains active, otherwise the model predicts that the company is going bankrupt.

4.1 Definition of variables

As discussed in the literature review, there are a large number of possible ratios identified as important predictors of a firm's performance. After gathering all the variables and financial ratios, each independent variable was tested through a "t-test" and the P-value of the model was observed. All variables with the significance level (α) of 0.1 (10%) or lower were tested in the regression. Only those which showed an improvement in the predictive capability of the model were included.

Considering the high variability of the ratios used, both due to the seniority of the companies, and due to the sectors of activity or typology, were winsorized in order to reduce the variability and number of outsiders without reducing the sample. All observations whose values were below the 0.01 percentile of the variable/ratio or above the 0.99 percentile of the sample assumed the value at their respective percentile. Once control variable assumes binary values it was not winsorized.

In order to avoid multicollinearity issues on the model such as those described by Verbeek (2012), in which two or more variables suffer from high correlation[7] with a linear or approximate relationship between them, the ratio that did the best overall job together in the remaining model was considered. To avoid this issue, table 3 indicates the correlation the level between variables of model, which was important information to avoid to put high correlated variables in the same model.

Table 3 - Correlation matrix of variables in the models

This table presents the correlation matrix statistics of all variables used in the model. ROCE was computed by dividing net income by total assets minus current liabilities. Solvency ratio was computed by dividing net income plus depreciation by total liabilities.

Correlation Matrix										
Variable	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)
1) Account Payable/Sales	1.00									
2) Account Receivable/Total Liabilities	-0.10	1.00								
3) Control Variable 5	0.12	-0.09	1.00							
4) EBIT/Interest Expenses	-0.03	0.07	0.00	1.00						
5) Intangible/Total assets	0.23	-0.18	-0.06	-0.06	1.00					
6) LN (Working Capital/Total Assets)	-0.12	0.32	-0.06	0.08	-0.35	1.00				
7) ROCE	-0.16	0.13	-0.04	0.20	-0.19	0.11	1.00			
8) Sales/Total Assets	-0.23	0.08	-0.21	0.11	-0.29	0.18	0.25	1.00		
9) Solvency Ratio	-0.17	0.16	-0.14	0.17	-0.21	0.20	0.33	0.19	1.00	
10) Working Capital/Total Assets	-0.08	0.34	-0.06	0.06	-0.34	0.87	0.13	0.15	0.16	1.00

After excluding all those non-significant variables, excluding the highly correlated variables, the variables to explain measure start-ups risk and its correlation matrix is exposed above. From a statistical point of view, that the variables did not present a high degree of correlation among them, except for the working capital to total assets and its natural logarithm. This is why working capital to total assets and its natural logarithm was not included simultaneously in the same model.

Considering the paragraphs mentioned above, from the 59 initial variables described in the appendices of the document, the 10 independent variables are summarized below in table 4 with its descriptive statistics.

[7] Was considered high correlation equal to or above 80%.

Table 4 - Descriptive statistics of variables used in the model

This table presents the descriptive statistics of all variables used in the model. The descriptive reflects post-winsorized values at 1st and 99th percentiles, except for the control variable which was not winsorized. Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities industry or manufacturing sectors. ROCE was computed by dividing net income by total assets minus current liabilities. Solvency ratio was computed by dividing net income plus depreciation by total liabilities.

Variable	Averag	Std Error	Mean	Std Deviation	Kurtosis	Skewness	Minimum	Maximum	Observations
Account Payable/Sales	1.44	0.33	0.40	10.05	571.62	22.20	0.00	271.17	925
Account Receivable/ Total Liabilities	0.87	0.03	0.65	1.01	21.06	3.92	0.00	8.02	925
Control Variable5	0.13	0.01	0.00	0.34	2.76	2.18	0.00	1.00	925
EBIT/Interest Expenses	832.38	643.80	3.95	16701.66	28.73	1.73	-106089.45	95829.61	673
Intangible/Total Assets	0.11	0.01	0.01	0.19	3.27	2.00	0.00	0.87	907
Ln (Working Capital/Total Assets)	-1.32	0.03	-1.02	1.01	4.27	-1.79	-6.07	-0.06	838
ROCE	-0.12	0.04	0.06	0.90	15.17	-3.43	-5.26	1.22	640
Solvency Ratio	12.40	6.80	2.83	206.75	18.00	0.71	-1585.58	1233.29	925
Working Capital/Total Assets	0.32	0.01	0.33	0.27	1.06	-0.40	-1.24	0.92	919

As previously seen there are several methodologies to test the default risk. Taking into account the advantages of binary models over multiple-discriminant analysis, it will be proposed a logistic regression model that gives us a probability of default or non-default of each company. Only for comparative purposes, a multiple discriminant analysis model, estimated by a least squares regression, will be presented.

Regarding the binaries approach, there are several options to define the dependent variable. Altman (1968) considered defaulted firms the ones that are legally bankrupt, either in liquidation or under the supervision of the courts in a reorganization process. According to Lehmann (2003), the concept of default corresponds to any of internal credit overdue for more than 90 days. Similarly, Beaver (1966) considered a default any event in those companies with the inability to settle the financial obligations at maturity. Differently, Altman et al. (2008) considered default those companies that went bankrupt, all that entered into voluntary liquidation, administration or judicial liquidation.

In this thesis, the same method applied by Altman et al. (2008) will be used. The dependent variable (status) assume the value of 1 for start-ups which remained active or zero when companies became inactive, insolvent, and with insolvency proceedings.

4.2 Empirical model

In order to interpret and discuss the results to evaluate and measure the degree of survival of start-ups in Portugal, three different regressions were estimated.

The following table reports the regression results for models I, II and III. Each estimation was aimed at obtaining the maximum of several significant variables at a level of significance up to 10%.

Table 5 - Logistic regression for the model I and II and least square for model III

This table reports the results of the regression, the independent variables, the coefficient and the statistical significance according to Eviews information, imported in panel-data structure. Standard errors are reported under the coefficients in parenthesis. Models I and II were estimated according to logit and model III was estimated by the least squares method. The dependent variable is company status (1=Non-default; 0=Default). ***, **, * indicates the significance level at 1%, 5% and 10%, respectively. Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities or manufacturing sectors. Ln refers to the natural logarithm. ROCE was computed by dividing net income by total assets minus current liabilities.

Variable	Model I	Model II	Model III
Account Payable/Sales	-0.004	-0.008	-0.004
	(0.02)	(0.027)	(0.002)
Account Receivable/Total Liabilities	0.315	0.245	0.002
	(0.33)	(0.314)	(0.013)
Control Variable 5	-1.357 ***	-1.525 ***	-0.138 ***
	(0.39)	(0.436)	(0.031)
EBIT/Interest Expenses	0.000		
	(0)		
Intangible/Total Assets	-1.625 **	-1.737 *	-0.167 ***
	(0.749)	(0.941)	(0.063)
LN (Working Capital/Total Assets)		0.058	0.004
		(0.179)	(0.011)
ROCE	0.306 **	0.417 **	0.035 **
	(0.135)	(0.164)	(0.014)
Sales/Total Assets	-0.152	0.042	-0.002
	(0.196)	(0.271)	(0.012)
Solvency Ratio	0.002 ***	0.002 ***	0.000 ***
	(0.001)	(0.001)	(0)
Working Capital /Total Assets	0.597		
	(0.673)		
C	2.989 ***	3.365 ***	0.978 ***
	(0.456)	(0.649)	(0.03)
McFadden R-squared / R-squared	0.169	0.211	0.143
Prob(LR statistic) / Prob(F-statistic)	0.000	0.000	0.000
Mean dependent var	0.915	0.929	0.929
Total OBS	627	575	575

Table 5 presents the regression coefficients in the different models which allow describing the start-ups' default probability in accordance with activity, coverage, leverage, liquidity, profitability, and industry. According to the table, all models are generally significant, although it has not been possible to have more than four significant variables. Even though some of the variables were not significant, their inclusion in the regression has allowed

increasing the forecasting capacity of the model. It should also be noted that the dependent variable assumes a value above 0.9 in all models, which makes it clear that a large part of the sample is related to non-default companies.

5. Results

In this section, the results of estimation presented in table 5 are analysed in order to identify the model that best identifies the companies' status (default or not-default). In addition, the average per variable of non-defaults and defaults group is compared in table 6.

Table 6 - Mean per group of variables included in the models

This table report the variable average according to the group that the company belongs to (default or non-default). Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities industry or manufacturing sectors. LN refers to the natural logarithm. ROCE was computed by dividing net income by total assets minus by current liabilities.

Variable	Non-Default group average	Default group average
Account Payable/Sales	1.33	2.96
Account Receivable/Total Liabilities	0.90	0.51
Control Variable 5	0.12	0.31
EBIT/Interest Expenses	1187.43	-3157.05
Intangible/Total Assets	0.10	0.21
LN (Working Capital/Total assets)	-1.29	-1.74
ROCE	-0.06	-0.72
Sales/Total Assets	1.31	0.90
Solvency Ratio	24.64	-157.94
Working Capital/Total Assets	0.33	0.18

5.1 Activity ratios

To measure how efficiently start-ups are being run three different activity ratios were used. To measure how much suppliers' money companies use to finance its sales, account payable to sales ratio was used. This gives us an indication of the speed a company pays to its vendors concerning sales. By assessing the results it is clear that companies in default have more than twice as much of account payable to sales than the non-default group. In the regression, a negative coefficient is observed, which suggests that companies with a higher ratio have a lower probability of survival, notwithstanding the ratio was not statistically significant.

The second ratio to evaluate the companies' competitive position according to the activity was receivables to total liabilities. This gives us the idea of how much credit start-ups provide to its customers compared to the total debt. Similarly to the previous ratio, this one did not show significance.

Finally, sales to total assets ratio was computed. According to table 6, non-default companies are generating more revenues per firm's asset than the default companies. In line with this conclusion, the regressions indicate that firms with a higher ratio have a higher probability of surviving. Although the coefficient of regression follows the expected signal and the conclusion in line with other studies such as Altman's (1968), the variable was not significant for any of the models.

5.2 Coverage ratios

The coverage ratio chosen to measure companies' ability to repay its debt was EBIT to interest expenses. A lower or a negative ratio shows that companies have less capacity to meet interest payments and vice-versa. According to table 6, start-ups in non-default situation present EBIT cover the interest payments, while the insolvency start-ups group on average do not generate an enough EBIT to cover interest payments. Regarding the regression, it suggests that companies with higher ratios may present a higher probability of surviving, although the variable was statistically irrelevant for all of the models.

5.3 Liquidity Ratios

To measure the firms' liquidity, the working capital to total assets and the intangible to total assets ratios was used. The working capital to total assets measures a firm's ability to cover its short-term financial obligations, and therefore it is expected that companies with a higher ratio are more likely to survive. This is the case of this sample, whose active companies have a higher ratio than insolvent companies. Although the variable was not statistically significant, the results of estimation are aligned with Altman's (1968), suggesting that companies with a higher ratio are more likely to survive.

Afterwards, in order to measure the liquidity, the intangible assets to total assets was considered. A higher ratio will indicate that a large portion of a firm's total assets is comprised of assets in intellectual property, trademarks, and copyright. Strong evidence were found at a significant level of 1%, 5% and 10% (depending on the model) that companies with large portion of intangible assets may experience higher default probabilities.

In this context, these results are in conflict with what is suggested in the literature. For instance, Lee et al. (2001) suggest that intellectual property on companies remained a powerful predictor of future performance. Likewise, Baum and Silverman (2004) argue that patents offer a temporary monopoly and therefore are in a more favourable position

compared to companies in an opposite situation. In contrast with literature, our results may be due to companies with a higher amount of intangible assets, have more assets which are not easy to be liquidated for money.

5.4 Solvency ratios

In order to quantify the level of leverage of a firm and its ability to meet its long-term obligations, the solvency ratio was computed according to the following mathematical expression:

$$\text{Solvency ratio} = \frac{Net\ Income + Depreciations}{Total\ Liabilities} \tag{5.1}$$

Having in mind the results from table 5, the index of solvency suggests that there is strong evidence that the higher the company ratio, the greater the probability of not entering in default. This conclusions were valid all model with significance level of 1% and in line with Ohlson's (1980) conclusions, whose suggest that higher the earnings generated by operations to total liabilities, higher the survival probability of survival.

5.5 Profitability

To measure the profitability the return on capital employed (ROCE) was computed based on the following expression:

$$\text{ROCE} = \frac{Net\ Income}{Total\ Assets - Current\ Liabilities} \tag{5.2}$$

Generally speaking, ROCE measures how well a company is using its capital to generate profits. In line with the expectations for all models presented in table 5, there is evidence for a significant level of 5% for companies which companies that generate more earnings per capital invested, hence, increasing the chances of survival.

5.6 Industry

Finally, a control variable according to the industry which the company belongs to was used. Control variable 5 represents a variable dummy, which assumes value 1 when companies are integrated into the human health and social support activities or manufacturing sector. This variable was set after comparing the level of profitability in the different sectors in table 1 and attesting remarkable differences according to the sector to which the company belongs to.

The statistical results revealed evidence at a significant level of 1% that on average companies in the industries of the human health and social support activities or manufacturing presented of a lower probability of survival.

5.7 Validation of results

As a way to evaluate the accuracy of each model, the prediction of the model (active/inactive) versus its real status was compared to determine the number of situations in which the model succeeded and failed. As previously stated, if the probability predicted by the model is greater than 0.5, then the prediction of the model will be that the company remains active. In contrast, if the probability predicted by the model is equal or lower than 0.5, the prediction of the model is that the company remains inactive.

In order to assess the quality of the general prediction, the model was validated by the number of observations correctly classified, by the type I and II error as presented on table 7.

Table 7 - Accuracy per model, according to the number of observations

This table reports the accuracy of each model by the number of observations. For this propose was compared the prediction of the model to the real status of the company.

Accuracy of models	Model I - Logit	Model II - Logit	Model III - OLS	Z-SCORE
Non-defaulted firms on reality classified as non-default by model	854	854	633	692
Non-defaulted firms on reality classified as defaulted by model	1	1	230	34
Defaulted firms on reality classified as defaulted by the model	6	8	38	28
Defaulted firms on reality classified as non-defaulted by the model	56	54	24	171
Nº of well classified observations	860	862	861	720
Sample	925			

Four possible scenarios are represented in table 7. The model classifies the firm as active and the company is active, therefore a correct forecast. The firm is not in default but the model classifies it as a default company (type II error). The model classifies the firm as inactive and the firm is inactive, therefore a correct prediction. The firm is in default and the model classifies the firm as non-default (type I error).

By grouping the number of correctly classified observations, the number of errors of type I and type II error, the following table presents a summary of error and general accuracy in relative terms.

Table 8 - Type I error and Type II error per model

Overall accuracy refers to the percentage of correct observations predicted by the model. Type I error is the number of defaulted firms classified as non-defaulted by the model to total defaults firms. Type II error is the number of non-defaulted firms classified as defaulted by the model to total non-defaults firms.

Accuracy of Models	Model I	Model II	Model III	Z-score
Overall accuracy	93.84%	94.05%	93.51%	77.84%
Type I error	90.32%	87.10%	95.16%	54.84%
Type II error	0.12%	0.12%	0.12%	19.81%

Overall, the models predict the status of around 860 out of 925 companies correctly, which represent a general accuracy of 93%. Nevertheless, the models fail to correctly classify the companies that were going to go into bankrupt. The model with higher accuracy predicts 8 out of 59 observations of default firms, which means that the model correctly classified about 13% of the total default companies.

Although the results of type 1 error are not as complete as desired, these results are somewhat justified and in line with other studies in the literature. According to Lennox (1999), studies in which the sample comprise an approximate number of failing and non-failing companies have much smaller error rates than those in which the frequency of failure is relatively small, as in this study. The issue of type I error is also found in Zmijewski's (1984) study, whose model predicts 95.3% of the companies correctly, although their precise classification for failed companies was 0%.

Despite the aforementioned, considering that it correctly classifies the larger number of companies, and presents a lower rate of error type I, the most appropriate model to evaluate the probability of survival (PS) was model II, defined by:

$$P = \frac{e^z}{(1+e^z)}, \text{ being } Z = 3{,}365 - 0{,}008X1 + 0{,}245X2 - 1{,}525X3 - 1{,}737X4 + 0{,}058X5 + 0{,}417X6 + 0{,}042X7 + 0{,}002X8 \quad (5.3)$$

[8] X1 Account Payable / Sales; X2 Account Receivable / Liabilities; X3 Control Variable5; X4 LN (tangible / Total Assets) ; X5 LN(working Capital/ Total Assets); X6 ROCE; X7 Sales / Total Assets; X8 Solvency Ratio

In order to evaluate the model's results, model II was compared to the Z-score estimated by Altman for private firms. Therefore, the Z-score was applied over the sample in order to evaluate the correct number of well-classified firms and type I and II error.

Considering the results presented in tables 5 and 6, in general, Z-score has a lower overall accuracy. Notwithstanding, Z-score has presented a superior ability to predict bankruptcy companies, correctly classifying about 45 % of default companies.

Comparing logistic regression results (model II) with OLS model (model III), for models with the same variables, the results suggest that regression is in agreement with what is stressed in the literature. In line with previous findings, the logistic model presents a greater overall forecasting capacity and greater predictability of the active companies.

The percentage of correctly predicted observation may not be a sufficient criterion for assessing the predictive capabilities of a model. It becomes necessary to know the impact of classifying well or badly a company. In this way, it would be more hazardous to estimate a company as active when the actual situation is that of insolvency (type I error) than to predict that a company is insolvent when it is not (Type II error).

In this context, in an attempt to reduce type I error, what was verified is that changing the value of the variable constant to a lower value will lead to a smaller percentage of type I error.

Table 9 - Sensitivity analysis of the impact of the change in the value of the constant in overall accuracy, type II and II error.

The table above represents the impact of changing the value of the constant on model II and its subsequent impact on the general accuracy of the model. Type I error is the number of defaulted firms classified as non-defaulted by the model to total defaults firms. Type II error is the number of non-defaulted firms classified as defaulted by the model to total non-defaults firms.

Constant	Overall Accuracy	Type I error	Type II error
-5.00	6.76%	0.00%	100.00%
-4.00	6.87%	0.00%	99.88%
-3.00	7.42%	0.00%	99.30%
-2.00	8.07%	0.00%	98.60%
-1.00	14.29%	1.61%	91.81%
-0.50	25.95%	1.61%	79.30%
0.00	59.54%	17.74%	42.11%
0.25	66.19%	33.87%	33.80%
0.50	70.99%	40.32%	28.19%
0.75	75.14%	45.16%	23.39%
1.00	78.08%	51.61%	19.77%
2.00	90.51%	67.74%	5.26%
3.00	94.22%	80.65%	0.35%
3.37	94.00%	87.10%	0.12%

The initial estimation of the constant was a value of 3.37, and for this result the model predicts the status of 94% of companies correctly, although it classifies as survival start-ups about 87% of non-survival companies.

Reducing the constant value of 3.37 down to 0.25 the model decreases the type 1 error from 87% to 33%, that is, the capability of Model I* to correctly classified bankrupt companies increase from 13% to 67%.

Whereas this amendment is done there is an opportunity cost in lowering the constant value because the model will classify active companies as insolvent, consequently it can stop investing in financially healthy companies, but on the other hand, there is greater certainty in avoiding bad investment and being more selective. Another important note to consider is that classifying active companies as insolvent may be a more conservative procedure, offering the awareness that although the company is not yet bankrupt, this can happen in a short time.

6. Conclusion

Considering the importance of startups and the risk associated to this type of business, particularly as far as the access to public or private financing is concerned, there is a huge need to avoid the risk of bankruptcy, both in terms of credit risk and investment.

In this sense, the present thesis was intended to contribute to presenting an estimated model which allows us to diagnose the axes that lead to survivorship/bankruptcy, allowing the preparation of a plan that can be implemented in the company, with the respective priorities of action, mitigating the risk and enhancing the degree of survival of the start-ups.

For this study, three different models with from 207 start-ups incubated in different science and technology parks and incubators located in Portugal were applied. Through the analysis of the behaviour of 59 financial, demographic and industry indicators, a logistic regression and an OLS analysis together with several determinants representing the liquidity, profitability, leverage, activity, and coverage were chosen in order to realistically transmit the financial situation of the company.

A series of conclusions on what the impacts of the risk of bankruptcy are were reached. The first, and perhaps the most obvious, conclusion concerns the industry differences in the sample. A critical analysis of the differences between indicators, both in terms of descriptive analysis and statistical analysis allowed us to conclude that there are significant differences in the value of assets, liabilities, revenues, earnings, ROE, EBITDA and EBIT margin of companies according to the sector. It is not surprising that, for example, the human health activities and social support industry has a volume of assets quite different from real estate industry, when the former needs research assets such as laboratory equipment that can be naturally more expensive than equipping an office for services in the real state. Considering the remarkable differences at this level, there was statistical strong evidence that companies in human health and social support activities or manufacturing sectors presented greater risks of bankruptcy.

From the analysis of the activity indicators, and its impact on the success or failure of the start-ups, the results were in line with expectation, although not conclusive as they did not present statistical significance.

In an attempt to measure start-ups' ability to repay its financial obligations, the coverage ratio EBIT to Interest Expenses was investigated, and again, although the variable has followed the expected signal, from the statistical point of view, the variable was irrelevant.

From the analysis of the impact of liquidity on the survival rate of the new companies two variables were tested: working capital to total assets and intangible to total assets. The results of working capital to total assets are aligned with previous findings of Altman (1968), although the results have not been conclusive, since the p-value of the variable is relatively high and, therefore, although the result is in line with the position of the literature, it is not relevant from the statistical point of view.

Contrary to what is suggested in the empirical literature, there is statistical evidence that the higher volume of intangible assets, the greater their likelihood of bankruptcy. Although in contrast to the literature, the results may be explained because intangible assets are not easy to be liquidated for money, or it may be due to start-ups acquiring other companies and overpaying for them.

According to the sample results, strong statistical evidence was found that companies with a high ability to meet its long-term obligations proved to be related with a lower level of insolvency rate. Therefore, companies whose ratio of net income plus depreciation to total liabilities was superior, tended to present better performance than companies in reverse situations.

By the same logic, in line with the empirical evidence, companies with a greater return per capital employed, also performed better than those whose ROCE was lower, with evidence for a significance level of 5%.

Another conclusion is related to the differences between the predictive capacity of the logistic regression models and the least square model. As suggested in the literature, the logit model presented better results than OLS, being highlighted that the greater capacity of the state of the largest number of companies, the greater ability to predict the states of the companies that remain active.

6.1 Questions for further research and study limitations

This dissertation is not free from limitations, therefore more studies in this area that should be pursued in order to validate the conclusions, test the robustness of the results, and test the hypotheses with more updated data.

The results obtained were very similar to those found in the literature, although the insights generated and explored in this thesis should be further investigated in-depth in future studies.

As noted above, the number of significant variables did not exceed four, depending on the model. Perhaps the unfeasibility to obtain a greater number of significant variables may be due to the heterogeneity of the sample.

Because there is no historical database of start-ups, and although several precautions have been pursued to nor exclude default companies, there may be a large number of bankruptcy companies that were not considered. As it has been verified, only 7% of the companies failed, which, even though they are insolvency rates close to the values reported in the sample, it may not represent the totality of the bankrupt companies of the population.

7. References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance, 23*(4), 589-609. Retrieved from https://www.jstor.org/stable/2978933

Altman, E. I., & Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from US Market. *ABACUS, 43*(3), 332-357. Retrieved from https://www.ssrn.com/abstract=872336

Altman, E. I., Sabato, G., & Wilson, N. (2008). The Value of Non-Financial Information in SME Risk Management. *Journal of Credit Risk, 6*(2), 95-127. Retrieved from https://ssrn.com/abstract=1320612

Araghi, M. K., & Makvandi, S. (2013). Comparing Logit, Probit and Multiple Discriminant Analysis Models in Predicting Bankruptcy of Companies. *Asian Journal of Finance & Accounting, 5*(1), 132-143. Retrieved from https://doi.org/10.5296/ajfa.v5i1.2977

Astebro, T. B., & Bernhardt, I. (2005). The Winner's Curse of Human Capital. *Small Business Economics,, 24*(1), 63-78. Retrieved from https://www.jstor.org/stable/40229410

Baum, J. A., & Silverman, B. (2004). Picking winners or building them Alliance, intellectual,and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing, 19*(3), Pages 411-436. Retrieved from https://doi.org/10.1016/S0883-9026(03)00038-7

Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research, 4*, 71-111. Retrieved from http://www.jstor.org/stable/2490171

Bellotti, T., & Crook, J. (2007). Credit Scoring With Macroeconomic Variables Using Survival Analysis. *Credit Research Centre Management School and Economics University of Edinburgh.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.8675&rep=rep1&type=pdf

Blank, S., & Dorf, B. (2012). *The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company.* Pescadaro: K&S Ranch.

Brinckmann, J., & Hoegl, M. (2011). Effects of Initial Teamwork Capability and Initial Relational Capability on the Development of New Technology-Based Firms. *Strategic Entrepreneurship Journal, 5*(1), 37–57. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/sej.106/abstract

Caballero, R. J., & Hammour, M. L. (1994). On the Timing and Efficiency of Creative Destruction. *The American Economic Review, 84*(5), 1350-1368. Retrieved from https://www.jstor.org/stable/2117776

Cabral, L. M., & Mata, J. (2003). On the Evolution of the Firm Size Distribution: Facts and Theory. *The american economic review, 94*(4), 1095-1090. Retrieved from https://ssrn.com/abstract=292129

Chandler, G., & Jansen, E. (1992). The Founder's Self-Assessed Competence and Venture Performance. *Journal of Business Venturing, 7*(3), 223-236. Retrieved from https://doi.org/10.1016/0883-9026(92)90028-P

Cooper, A. C., Gimeno-Gascon, F., & Woo, C. Y. (1994). Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing, 9*(5), 371-395. Retrieved from https://doi.org/10.1016/0883-9026(94)90013-2

Eisenhardt, K. M., & Schoonhoven, C. B. (1990). Organizational Growth: Linking Founding Team, Strategy, Environment, and Growth Among U.S. Semiconductor Ventures, 1978-1988. *Administrative Science Quarterly, 35*(3), 504-529. Retrieved from https://www.jstor.org/stable/2393315

Geroski, P., Mata, J., & Portugal, P. (2009). Founding conditions and the survival of new firms. *Strategic Managment Journal, 31*(5), 510-529. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/smj.823/abstract

Gonçalves, V. S., Martins, F. V., & Brandão, E. (2014). The Determinants of Credit Default on Start-Up Firms. Econometric Modelling using Financial Capital, Human Capital and Industry Dynamics Variables. *Working Paper Faculdade de Economia do Porto, 534*. Retrieved from wps.fep.up.pt/wps/wp534.pdf

Huynh, K. P., J. Petrunia, R., & Voia, M. (2012). Duration of new firms: The role of startup financial conditions, industry and aggregate factors. *Structural Change and Economic*

Dynamics, 23(4), 354–362. Retrieved from https://doi.org/10.1016/j.strueco.2012.03.008

Huynh, T., Patton, D., Arias-Aranda, D., & Molina-Fernández, L. M. (2017). University spin-off's performance: Capabilities and networks of founding teams at creation phase. *Journal of Business Research*, 10-22. Retrieved from http://www.sciencedirect.com/science/article/pii/S0148296317301388

Kleinbaum, D. G., & Klein, M. (2002). *Survival Analysis: A Self-Learning Text, Third Edition.* New York: Springer Science. doi:10.1007/978-1-4419-6646-9

Lee, C., Lee, K., & Pennings, J. M. (2001). Internal capabilities, external networks, and performance: a study on technology-based ventures. *Strategic Management Journal, 22*(6-7), 615-640. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/smj.181/full

Lee, J., & Zhang, W. (2010). Financial capital and startup survival. *SSRN Electronic Journal.* Retrieved from https://ssrn.com/abstract=1659046

Lennox, C. (1999). Identifying Failing Companies: A Reevaluation of the Logit, Probit and. *Journal of Economics and Business, 51*, 347–364. Retrieved from https://www.researchgate.net/publication/4973411_Identifying_failing_companies_A_reevaluation_of_the_logit_probit_and_DA_approaches

Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management, 33*(1), 8-20. Retrieved from https://www.questia.com/library/journal/1G1-16787401/a-nonfinancial-business-success-versus-failure-prediction

Lussier, R. N., & Halabi, C. E. (2010). A Three-Country Comparison of the Business Success versus Failure Prediction Model. *Journal of Small Business Management, 48*(3), 360-377. Retrieved from http://dx.doi.org/10.1111/j.1540-627X.2010.00298.x

Lussier, R. N., & Pfeifer, S. (2001). A crossnational prediction model for business success. *Journal of Small Business Management, 39*(3), 228-239. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/0447-2778.00021/pdf

Manigart, S., Baeyens, K., & Hyfte, W. V. (2002). The survival of venture capital backed companies. *An International Journal of Entrepreneurial Finance, 4*(2), 103-124. Retrieved from https://doi.org/10.1080/13691060110103233

Mata, J., & Portugal, P. (1994). Life duration of new firms. *The Journal of Industrial Economics, 42*(3), 227-245. Retrieved from https://www.jstor.org/stable/2950567

Mata, J., Portugal, P., & Guimaraes, P. (1995). The Survival of New Plants: Start-up Conditions and Post-entry Evolution. *International Journal of Industrial Organization, 13*(4), 459-481. Retrieved from https://doi.org/10.1016/0167-7187(95)00500-5

Miloud, T., Aspelund, A., & Cabrol, M. (2012). Startup valuation by venture capitalists an empirical study. *Venture Capital: An International Journal of Entrepreneurial, 14*, 151-174. Retrieved from https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1080%2F13691066.2012.667907

Morris, R. C. (1997). *Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence.* Aldershot: Ashgate Publishing.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109-131. Retrieved from http://www.jstor.org/stable/2490395

Olaison, L., & Sorensen, B. M. (2014). The abject of entrepreneurship: failure, fiasco, fraud. *International Journal of Entrepreneurial Behaviour & Research, 20*(2), 193-211. Retrieved from https://www.researchgate.net/publication/263245485_The_abject_of_entrepreneurship_Failure_fiasco_fraud

Peterson, R. A., Kozmetsky, G., & Ridgway, N. M. (1983). Perceived Causes of Small Business Failures: A Research Note. *Journal of Small Business, 8*(1), 15-19. Retrieved from https://www.researchgate.net/publication/236965160

Ries, E. (2011). *The Lean Startup* (1 ed.). New York: Crown Publishing Group.

Schumpeter, J. A. (1934). *The Theory of Economic Development.* Cambrigde: Harvard University Press.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business, 74*(1), 101-124 . Retrieved from http://www.jstor.org/stable/10.1086/209665

Sikomwe, S., Kandufa, P., Giga, D., & Mudzurandende, F. (2014). Analyzing Business Failure in Relation to the Life Cycle of the Business and the Economic Cycle: A Study of Entrepreneurial Ventures in Hwange (Zimbabwe). *European Journal of Business and Management, 6*(7), 80-91. Retrieved from http://www.iiste.org/Journals/index.php/EJBM/article/view/11434

Silva, D., Cerqueira, A., & Brandão, E. (2016). Portuguese Startups: a success prediction model. *Working Paper Faculdade de Economia do Porto, 581*. Retrieved from http://wps.fep.up.pt/wps/wp581.pdf

Stepanova, M., & Thomas, L. (2002). Survival Analysis Methods for Personal Loan Data. *Operations Research, 50*(2), 277-289. Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/opre.50.2.277.426

Stepanova, M., & Thomas, L. C. (2001). PHAB Scores: Proportional Hazards Analysis Behavioural Scores. *The Journal of the Operational Research Society, 52*(9), 1007-1016 . Retrieved from http://www.jstor.org/stable/822780

Teixeira, A. A., & Grande, M. (2013). Determinants of the economic performance of Portuguese Academic Spin-offs: do Science & Technology infrastructures and support matter? *Working Papers Faculdade de Economia do Porto, 502*. Retrieved from http://wps.fep.up.pt/wps/wp502.pdf

Thomas, G. (2009). *How to Do Your Research Project: A Guide for Students in Education and Applied Social Sciences.* London: SAGE.

Verbeek, M. (2012). *A Guide to Modern Econometrics, 4th Edition.* Wiley.

Zmijewski, M. E. (Journal of Accounting Research). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research, 22*, 59-82. Retrieved from https://www.jstor.org/stable/pdf/2490859.pdf?seq=1#page_scan_tab_contents

8. Annex

8.1 Descriptive statistics of tested variables

Table 10 - Descriptive statistics of tested variables

This table presents the descriptive statistics (average, standard error, mean, standard deviation and kurtosis) of the variables and the final sample of 925 observations, concerning to different reported years of 207 companies. The descriptive reflects post-winsorized values at 1st and 99th percentiles, except for control variables which were not winsorized. Age refers to the seniority of the company. The book value of equity refers to shareholders funds. Cash flow equal to net income plus depreciation and amortization minus capital expenditures. Control variable 1 is a dummy variable that assumes the value of one if total liabilities exceed total assets, zero otherwise. Control variable 2 is a dummy variable that assumes the value of one if net income was negative for the last two years, zero otherwise. Control variable 3 is a dummy variable that assumes the value of one if the company belongs to the sector of activity of CAE 20592, 72110, 72190, 74900, 86906, zero otherwise. Control variable 4 is a dummy variable that assumes the value of zero if net income was negative for the last three years, one otherwise. Control variable 5 represents a variable dummy, which assumes value 1 when companies are integrated into the human health and social support activities sector or manufacturing sectors. Control variable 5 for companies integrated into agriculture, livestock, hunting, forestry and fishing, wholesale and retail trade, information and communication activities, real estate activities, consulting, scientific, technical and similar activities, administrative and support services activities sectors, the dummy assumes the value of zero. Current ratio was calculated by dividing current assets by current liabilities. Year of incorporation refers to the year on which the company was legally registered. Intangible assets refers to intangible fixed assets. Liquidity ratio was computed by dividing cash and cash equivalents by short-term liabilities. Ln refers to the natural logarithm.

Variable	Average	Std Error	Mean	Std Deviation	Kurtosis
Account Payable/Sales	1.44	0.33	0.40	10.05	571.62
Account Receivable/Total Liabilities	0.87	0.03	0.65	1.01	21.06
Age	3.74	0.08	3.00	2.54	-0.67
Book Value of Equity /Total Liabilities	1.84	0.18	0.55	5.43	104.75
Book Value of Equity plus Total Assets/Total Liabilities	4.84	0.45	2.11	13.68	257.25
Book Value of Equity/Total Assets	0.35	0.01	0.36	0.38	9.36
Cash/Sales	0.91	0.23	0.13	6.91	551.57
Cash/Total Assets	0.23	0.01	0.15	0.23	0.26
Cash-Flow/Operating Revenues	0.04	0.01	0.08	0.25	3.23
Cash-Flow/Total Assets	0.05	0.01	0.08	0.31	13.31
Control Variable1	0.09	0.01	0.00	0.29	6.28
Control Variable2	0.22	0.01	0.00	0.41	-0.11
Control Variable3	0.22	0.01	0.00	0.42	-0.20
Control Variable4	0.90	0.01	1.00	0.30	4.99
Control Variable5	0.13	0.01	0.00	0.34	2.76
Costs of Employees/Revenues	0.92	0.17	0.44	5.21	401.17
Current Ratio	2.90	0.13	1.65	3.80	18.29
EBIT Margin	0.00	0.01	0.04	0.25	2.75
EBIT/Interest Expenses	832.38	643.80	3.95	16701.66	28.73
EBIT/Sales	-1.12	0.39	0.03	11.78	368.47
EBIT/Total Assets	0.01	0.01	0.04	0.32	11.45
EBITDA/Interest Expenses	2512.83	928.58	10.77	24089.53	35.56
EBITDA/Total Assets	0.07	0.01	0.09	0.32	10.53
Intangible/Total Assets	0.11	0.01	0.01	0.19	3.27
Liquidity Ratio	2.69	0.12	1.51	3.55	16.74
Ln (Accounts Payables/Sales)	-0.89	0.04	-0.91	1.19	2.68
Ln (Accounts Receivables/Total Liabilities)	-0.61	0.04	-0.42	1.09	2.97
Ln (Age)	1.19	0.02	1.39	0.69	-0.88
Ln (Book Value of Equity/Total Liabilities)	-0.43	0.05	-0.49	1.50	0.66

Table 11 - Continuation of descriptive statistics of tested variables

This table presents the descriptive statistics (skewness, minimum, maximum and observations) of the variables and the final sample of 925 observations, concerning to different reported years of 207 companies. The descriptive reflects post-winsorized values at 1st and 99th percentiles, except for control variables which were not winsorized. Age refers to the seniority of the company. The book value of equity refers to shareholders funds. Cash flow equal to net income plus depreciation and amortization minus capital expenditures. Control variable 1 is a dummy variable that assumes the value of one if total liabilities exceed total assets, zero otherwise. Control variable 2 is a dummy variable that assumes the value of one if net income was negative for the last two years, zero otherwise. Control variable 3 is a dummy variable that assumes the value of one if the company belongs to the sector of activity of CAE 20592, 72110, 72190, 74900, 86906, zero otherwise. Control variable 4 is a dummy variable that assumes the value of zero if net income was negative for the last three years, one otherwise. Control variable 5 represents a variable dummy, which assumes value 1 when companies are integrated into the human health and social support activities sector or manufacturing sectors. Control variable 5 for companies integrated into agriculture, livestock, hunting, forestry and fishing, wholesale and retail trade, information and communication activities, real estate activities, consulting, scientific, technical and similar activities, administrative and support services activities sectors, the dummy assumes the value of zero. Current ratio was calculated by dividing current assets by current liabilities. Year of incorporation refers to the year on which the company was legally registered. Intangible refer to intangible fixed assets. Liquidity ratio was computed by dividing cash and cash equivalents by short-term liabilities. Ln refers to the natural logarithm.

Variable	Skewness	Minimum	Maximum	Observations
Account Payable/Sales	22.20	0.00	271.17	925.00
Account Receivable/Total Liabilities	3.92	0.00	8.02	925.00
Age	0.45	0.00	11.00	925.00
Book Value of Equity /Total Liabilities	9.01	-0.69	78.33	925.00
Book Value of Equity plus Total Assets/Total Liabilities	13.85	-0.38	299.30	925.00
Book Value of Equity/Total Assets	-1.84	-2.25	0.99	925.00
Cash/Sales	21.67	0.00	184.80	925.00
Cash/Total Assets	1.07	0.00	0.93	925.00
Cash-Flow/Operating Revenues	-1.50	-0.91	0.57	855.00
Cash-Flow/Total Assets	-2.34	-2.48	0.68	925.00
Control Variable1	2.88	0.00	1.00	925.00
Control Variable2	1.37	0.00	1.00	925.00
Control Variable3	1.34	0.00	1.00	925.00
Control Variable4	-2.64	0.00	1.00	925.00
Control Variable5	2.18	0.00	1.00	925.00
Costs of Employees/Revenues	18.76	0.01	125.45	905.00
Current Ratio	3.83	0.06	28.08	921.00
EBIT Margin	-1.48	-0.92	0.60	838.00
EBIT/Interest Expenses	1.73	-106089.45	95829.61	673.00
EBIT/Sales	-18.13	-271.83	0.60	925.00
EBIT/Total Assets	-2.19	-2.58	0.70	925.00
EBITDA/Interest Expenses	5.14	-93822.37	161536.17	673.00
EBITDA/Total Assets	-1.94	-2.47	0.76	925.00
Intangible/Total Assets	2.00	0.00	0.87	907.00
Liquidity Ratio	3.69	0.06	25.45	921.00
Ln (Accounts Payables/Sales)	0.69	-3.91	5.69	923.00
Ln (Accounts Receivables/Total Liabilities)	-1.08	-5.41	2.09	917.00
Ln (Age)	-0.44	0.00	2.40	852.00
Ln (Book Value of Equity/Total Liabilities)	-0.10	-4.95	4.83	842.00

Table 12 - Continuation of descriptive statistics of tested variables

This table presents the continuation of descriptive statistics (average, standard error, mean, standard deviation and kurtosis) of the variables and the final sample of 925 observations, concerning to different reported years of 207 companies. The descriptive reflects post-winsorized values at 1st and 99th percentiles, except for control variables which was not winsorized. Ln refers to the natural logarithm. Log refers to the common logarithm. Profit margin was calculated as net income divided by revenues. ROA was computed by dividing Net Income by total assets. ROCE was computed by dividing net income by total assets minus current liabilities. ROE was computed based by dividing net income by shareholder's equity. Shareholders liquidity ratio was computed by dividing shareholders' equity by total assets of the firm. Solvency ratio was computed by dividing net income plus depreciation by total liabilities. Year information refers to what year financial information is related.

Variable	Average	Std Error	Mean	Std Deviation	Kurtosis
Ln (Cash/Sales)	-2.11	0.06	-2.05	1.79	0.93
Ln (Cash/Total Assets)	-2.25	0.05	-1.91	1.61	1.21
Ln (EBIT/Interest Expenses)	3.75	0.13	3.38	2.74	1.41
Ln (EBIT/Sales)	-2.68	0.05	-2.51	1.13	-0.28
Ln (EBITDA/Interest Expenses)	4.21	0.12	3.81	2.54	1.89
Ln (Intangible/Total Assets)	-3.03	0.09	-2.46	2.21	0.00
Ln (Profit Margin)	-3.10	0.06	-2.75	1.54	0.30
Ln (Sales/Total Assets)	-0.14	0.04	0.10	1.10	6.82
Ln (Short-Term Debt/Book Value of Equity)	2.70	0.28	0.92	8.66	27.16
Ln (Working Capital/Total Assets)	-1.32	0.03	-1.02	1.01	4.27
Ln [1 - (EBITDA/Total Assets)]	0.12	0.01	0.10	0.33	2.76
Ln [1 - (Retained Earnings/Total Assets)]	0.63	0.02	0.44	0.73	3.59
Log (Book Value of Equity plus Total Assets/Total Liabilities)	0.40	0.01	0.33	0.42	2.65
Log [1 -(Intangible/Total Assets)]	4.09	0.10	3.46	2.46	6.27
Net Debt/Cash-Flow	2.65	0.52	1.03	15.74	18.38
Net Debt/EBITDA	1.38	0.49	0.90	14.97	20.05
Net Income/Total Assets	-0.01	0.01	0.02	0.31	14.39
Profit Margin	-1.12	0.39	0.02	11.78	373.50
Retained Earnings/Total Assets	0.35	0.01	0.36	0.38	9.36
ROA	-0.01	0.01	0.02	0.31	14.39
ROCE	-0.12	0.04	0.06	0.90	15.17
ROE	-0.07	0.03	0.10	0.99	11.58
Sales/Total Assets	1.28	0.03	1.10	1.05	6.82
Shareholders Liquidity Ratio	11.89	1.65	1.29	37.74	27.50
Short-Term Debt/Book Value of Equity	2.70	0.28	0.92	8.66	27.16
Solvency Ratio	12.40	6.80	2.83	206.75	18.00
Status	0.93	0.01	1.00	0.25	10.05
Working Capital/Total Assets	0.32	0.01	0.33	0.27	1.06
Year of Incorporation	2009.13	0.09	2008.00	2.63	-0.49
Year of Information	2012.87	0.08	2013.00	2.58	-0.68

Table 13 - Continuation of descriptive statistics of tested variables

This table presents the continuation of descriptive statistics (skewness, minimum, maximum and observations) of the variables and the final sample of 925 observations, concerning to different reported years of 207 companies. The descriptive reflects post-winsorized values at 1st and 99th percentiles, except for control variables which was not winsorized. Ln refers to the natural logarithm. LOG refers to the common logarithm. Profit margin was calculated as net income divided by revenues. ROA was computed by dividing Net Income by total assets. ROCE was computed by dividing net income by total assets minus current liabilities. ROE was computed based by dividing net income by shareholder's equity. Shareholders liquidity ratio was computed by dividing shareholders' equity by total assets of the firm. Solvency ratio was computed by dividing net income plus depreciation by total liabilities. Year information refers to what year financial information is related.

Variable	Skewness	Minimum	Maximum	Observations
Ln (Cash/Sales)	0.02	-7.27	5.26	924.00
Ln (Cash/Total Assets)	-1.14	-8.23	-0.07	924.00
Ln (EBIT/Interest Expenses)	1.07	-0.91	12.73	419.00
Ln (EBIT/Sales)	-0.53	-5.88	-0.50	615.00
Ln (EBITDA/Interest Expenses)	1.21	-0.56	12.91	448.00
Ln (Intangible/Total Assets)	-0.87	-9.38	-0.12	555.00
Ln (Profit Margin)	-0.90	-7.44	0.19	609.00
Ln (Sales/Total Assets)	-1.96	-6.91	1.98	925.00
Ln (Short-Term Debt/Book Value of Equity)	4.56	-18.47	59.99	925.00
Ln (Working Capital/Total Assets)	-1.79	-6.07	-0.06	838.00
Ln [1 - (EBITDA/Total Assets)]	0.56	-1.24	1.28	922.00
Ln [1 - (Retained Earnings/Total Assets)]	1.50	-1.18	4.37	925.00
Log (Book Value of Equity plus Total Assets/Total Liabilities)	0.62	-1.48	2.48	917.00
Log [1 -(Intangible/Total Assets)]	1.80	1.13	18.54	555.00
Net Debt/Cash-Flow	1.81	-76.58	115.15	925.00
Net Debt/EBITDA	-0.78	-90.65	94.17	925.00
Net Income/Total Assets	-2.64	-2.60	0.61	925.00
Profit Margin	-18.31	-271.83	0.99	925.00
Retained Earnings/Total Assets	-1.84	-2.25	0.99	925.00
ROA	-2.64	-2.60	0.61	925.00
ROCE	-3.43	-5.26	1.22	640.00
ROE	-2.92	-5.49	2.08	834.00
Sales/Total Assets	2.10	0.00	7.25	925.00
Shareholders Liquidity Ratio	5.09	-2.71	254.93	525.00
Short-Term Debt/Book Value of Equity	4.56	-18.47	59.99	925.00
Solvency Ratio	0.71	-1585.58	1233.29	925.00
Status	-3.47	0.00	1.00	925.00
Working Capital/Total Assets	-0.40	-1.24	0.92	919.00
Year of Incorporation	0.63	2005.00	2016.00	925.00
Year of Information	-0.56	2006.00	2017.00	925.00

Table 14 - Primary industry code and correspondent sector

This the table reports the primary industry code (CAE) and its correspondent sector of all companies analysed in the sample. The grouping criteria followed the same criteria of CAE-Rev.3d of INE.

Industry	Sector
01130	Agriculture, livestock, hunting, forestry and fishing
01210	Agriculture, livestock, hunting, forestry and fishing
01280	Agriculture, livestock, hunting, forestry and fishing
10860	Manufacturing
10893	Manufacturing
11021	Manufacturing
11050	Manufacturing
20120	Manufacturing
20144	Manufacturing
20592	Manufacturing
20594	Manufacturing
21100	Manufacturing
22292	Manufacturing
26120	Manufacturing
26512	Manufacturing
26701	Manufacturing
27122	Manufacturing
27900	Manufacturing
28293	Manufacturing
28992	Manufacturing
29310	Manufacturing
30300	Manufacturing
32300	Manufacturing
32502	Manufacturing

Table 15 - Continuation of primary industry code and correspondent sector

This the table reports the primary industry code (CAE) and its correspondent sector of all companies analysed in the sample. The grouping criteria followed the same criteria of CAE-Rev.3d of INE.

Industry	Sector
32996	Manufacturing
46520	Wholesale and retail trade
46690	Wholesale and retail trade
47910	Wholesale and retail trade
47990	Wholesale and retail trade
58210	Information and communication activities
58290	Information and communication activities
59110	Information and communication activities
61200	Information and communication activities
61300	Information and communication activities
62010	Information and communication activities
62020	Information and communication activities
62090	Information and communication activities
63110	Information and communication activities
63120	Information and communication activities
68311	Real estate activities
70220	Consulting, scientific, technical and similar activities
72110	Consulting, scientific, technical and similar activities
72190	Consulting, scientific, technical and similar activities
74900	Consulting, scientific, technical and similar activities
79900	Administrative and support services activities
80200	Administrative and support services activities
82110	Administrative and support services activities
86904	Human health activities and social support
86906	Human health activities and social support

8.2 Model I regression

Figure 1 - Model 1 regression, source: eviews

This figure reports the estimation output of Eviews of the model I regression, estimated through logit method. The dependent variable is company status (1=Non-default; 0=Default). ***, **, * indicates the significance level at 1%, 5% and 10%, respectively. X1 represents the account payable to sales, X2 account receivable to total liabilities, X3 is the control variable5, X4 is EBIT to interest expenses, X5 is intangible to total assets, X6 is ROCE ratio; X7 is sales to total assets; X8 is the solvency ratio and X9 is working capital to total assets. Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities or manufacturing sectors. ROCE was computed by dividing net income by total assets minus current liabilities.

Dependent Variable: STATUS
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 09/10/18 Time: 14:41
Sample: 1 925
Included observations: 627
Convergence achieved after 4 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1	-0.00392...	0.020084	-0.195539	0.8450
X2	0.314539...	0.330118	0.952809	0.3407
X3	-1.356587	0.390151	-3.477081	0.0005
X4	1.06E-05	9.17E-06	1.150457	0.2500
X5	-1.625099	0.749473	-2.168322	0.0301
X6	0.306439	0.134540	2.277684	0.0227
X7	-0.152065	0.195890	-0.776281	0.4376
X8	0.002338	0.000790	2.957681	0.0031
X9	0.597399	0.673262	0.887320	0.3749
C	2.989343	0.455729	6.559478	0.0000

McFadden R-squared	0.168771	Mean dependent var	0.915470
S.D. dependent var	0.278402	S.E. of regression	0.255108
Akaike info criterion	0.513504	Sum squared resid	40.15443
Schwarz criterion	0.584332	Log likelihood	-150.9834
Hannan-Quinn criter.	0.541021	Deviance	301.9668
Restr. deviance	363.2775	Restr. log likelihood	-181.6387
LR statistic	61.31063	Avg. log likelihood	-0.240803
Prob(LR statistic)	0.000000		

Obs with Dep=0	53	Total obs	627
Obs with Dep=1	574		

8.3 Model II regression

Figure 2 - Model II regression, source: eviews

This figure reports the estimation output of Eviews for model II regression, estimated through logit method. The dependent variable is company status (1=Non-default; 0=Default). ***, **, * indicates the significance level at 1%, 5% and 10%, respectively. X1 represents the account payable to sales, X2 is account receivable to total liabilities, X3 is control variable 5, X4 intangible to total assets, X5 represent the natural logarithm of working capital to total assets, X6 is ROCE, X7 is sales to total assets ratio and X8 it the solvency ratio. Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities or manufacturing sectors. ROCE was computed by dividing net income by total assets minus current liabilities.

Dependent Variable: STATUS
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 09/10/18 Time: 14:55
Sample: 1 925
Included observations: 575
Convergence achieved after 4 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1	-0.008271	0.027294	-0.303028	0.7619
X2	0.245459	0.314386	0.780759	0.4349
X3	-1.525301	0.435888	-3.499296	0.0005
X4	-1.737301	0.940941	-1.846344	0.0648
X5	0.057965	0.178807	0.324173	0.7458
X6	0.416963	0.163505	2.550154	0.0108
X7	0.042406	0.271346	0.156278	0.8758
X8	0.002495	0.000856	2.914047	0.0036
C	3.365375	0.649192	5.183941	0.0000

McFadden R-squared	0.211019	Mean dependent var	0.928696
S.D. dependent var	0.257557	S.E. of regression	0.228535
Akaike info criterion	0.436841	Sum squared resid	29.56122
Schwarz criterion	0.504996	Log likelihood	-116.5917
Hannan-Quinn criter.	0.463422	Deviance	233.1833
Restr. deviance	295.5499	Restr. log likelihood	-147.7749
LR statistic	62.36656	Avg. log likelihood	-0.202768
Prob(LR statistic)	0.000000		

Obs with Dep=0	41	Total obs	575
Obs with Dep=1	534		

8.4 Model III regression

Figure 3 - Model III regression, source: eviews

This figure reports the estimation output of Eviews for model III regression, estimated through least square method. The dependent variable is company status (1=Non-default; 0=Default). ***, **, * indicates the significance level at 1%, 5% and 10%, respectively. X1 represents the account payable to sales, X2 is account receivable to total liabilities, X3 is control variable 5, X4 intangible to total assets, X5 represent the natural logarithm of working capital to total assets, X6 is ROCE, X7 is sales to total assets ratio and X8 it the solvency ratio. Control variable 5 represents a variable dummy, which assumes the value of 1 when companies are integrated into the human health and social support activities or manufacturing sectors. ROCE was computed by dividing net income by total assets minus current liabilities.

Dependent Variable: STATUS
Method: Least Squares
Date: 09/10/18 Time: 15:02
Sample: 1 925
Included observations: 575

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	-0.003694	0.002462	-1.500661	0.1340
X2	0.002241	0.012625	0.177539	0.8591
X3	-0.138161	0.031006	-4.455949	0.0000
X4	-0.166846	0.063238	-2.638388	0.0086
X5	0.004459	0.011420	0.390462	0.6963
X6	0.034844	0.013938	2.499883	0.0127
X7	-0.002269	0.011711	-0.193716	0.8465
X8	0.000213	5.11E-05	4.178025	0.0000
C	0.977911	0.029732	32.89067	0.0000

R-squared	0.142506	Mean dependent var	0.928696
Adjusted R-squared	0.130386	S.D. dependent var	0.257557
S.E. of regression	0.240180	Akaike info criterion	0.000668
Sum squared resid	32.65039	Schwarz criterion	0.068824
Log likelihood	8.807863	Hannan-Quinn criter.	0.027250
F-statistic	11.75785	Durbin-Watson stat	0.294012
Prob(F-statistic)	0.000000		