MOHAMED ELAMIN

Risk analysis: Identifying the determinants of HORECA companies default risk

UNIVERSITY OF ALGARVE

FACULTY OF ECONOMICS

2021

MOHAMED ELAMIN

Risk analysis: Identifying the determinants of HORECA companies default risk

Internship Report for obtaining Masters degree in Financial Economics

Work made under the supervision of:

Professor Jorge Miguel Lopo Gonçalves Andraz

Dr. David João Antunes Pedro

UNIVERSIDADE DO ALGARVE
FACULDADE DE ECONOMIA

UNIVERSITY OF ALGARVE

FACULTY OF ECONOMICS

2021

Authorship and Copyright Declaration

Risk analysis: Identifying the determinants of HORECA companies default risk

Work authorship declaration

I declare to be the author of this work, which is unique and unprecedented. Authors and works consulted are properly cited in the text and are included in the listing of references included.

MOHAMED ELAMIN

Copyright

Acknowledgement

This internship report would have never been completed without the help and support of my supervisor Professor Jorge Andraz. Throughout the time of building writing and rewriting of this report, he was always there to give me much needed information and support. Professor Andraz was the first faculty member I saw when I first arrived to Portugal, I was completely nervous and anxious on how to tackle the Masters program since I arrived later than usual for the beginning of the Master's Program. But he guided me and showed me the best way that was possible to handle the Master's program without any losses. Any time I had a question or doubt he was available to answer it, any time I had a problem he was there to help me fix it. Because of his constant kindness and because of his wonderful character he became like a second father to me. Professor, thank you will never be enough.

I also would like to thank in these few lines below, my supervisor at Yourdata Mr. David Pedro and the amazing folks there.

Being there daily in Yourdata felt like being amongst a family rather than work colleagues. This was the same feeling I felt being a student at University of the Algarve. From day one my colleagues at Yourdata were extremely helpful, supporting and friendly towards me and towards any doubts that I had, and I had plenty. My supervisor at the host company Mr. David. was extremally kind and patient with me in performing my duties and tasks particularly in the beginning of the internship. My experience in Yourdata when it came to asking for some clarification or a query, felt as if I'm asking a friend of mine for some help. The internship at yourdata was nothing short of pleasant, informative and it never felt dull, never felt like an obligation nor a forced duty. I truly can't explain and give proper rights and thanks to the wonderful people in there, it's an experience that I'm extremely grateful and glad that I have gone through.

It was a pleasure and an honor having both of my supervisors in my life.

RESUMO

As pequenas e médias empresas (PME) são extremamente importantes. Eles são o verdadeiro motor de qualquer economia, uma vez que são grande enpregadores da força de trabalho disponível e geralmente são consideradas um bom indicador da trajetoria da economia. Em Portugal, as PME dominam a quota de mercado no setor empresarial não financeiro e contratam a maior parte da força de trabalho disponível. A sua importância vital determina a importância de se conhecer a sua situação financeira.

Foi realizado um estágio na Yourdata (empresa de Business Intelligence localizada em Faro) com o objetivo de tentar construir um modelo de previsão da probabilidade de incumprimento de empresa pertencentes ao sector HORECA na região do Algarve e olhar para o bem-estar financeiro das PME's da região. O modelo foi construído exclusivamente a partir de dados e informações proporcionados pela Yourdata, e que posteriormente foram colocados num ficheiro de dados Excel. Durante o estágio, foram analisadas as informações financeiras de diferentes empresas (principalmente informações contabilísticas) pertencentes aos clientes da Yourdata.

A construção do modelo utilizou a metodologia de regressão multinomial devido à natureza multicategórica e não ordenada da variável dependente. O modelo, na sua forma final, baseia-se em três índices financeiros cobrindo as áreas da liquidez, rentabilidade e alavancagem.

Os resultados do modelo estimado estão de acordo com o que foi encontrado em estudos e artigos semelhantes na literatura. Os resultados do modelo estimado mostraram que um aumento nos índices de rentabilidade e liquidez resulta na diminuição da probabilidade de incumprimemto das empresas. E um aumento nos índices de alavancagem, gera um aumento na probabilidade de incumprimento.

ABSTRACT

Small and medium enterprises (SME's) are extremely important. They are the true engine of any economy. They hire the majority of the available workforce and usually considered to be an index to gauge the performance of an economy. In Portugal SME's dominate the market share in the non-financial business sector and they hire the most of the available workforce. And because of their vital importance it is very important to know about their financial wellbeing.

An internship was undertaken at Yourdata (a Business Intelligence firm located in Faro) with a goal to try and build a model that is capable of predicting the default probability of a firms based in the HORECA (the food and Service) sector in the Algarve region. The model was built solely out of data and information that the host company had, and that was later placed in a single Excel database file. During the internship the trainee examined different companies' financial information (mainly accounting information) belonging to YourData clients.

The estimation of the model was based on the multinomial regression methodology due to the multi-categorical nature and unordered nature of the dependent variable. The model in its final shape consisted of three financial ratios addressing the areas of liquidity, profitability and leverage.

The results of the estimated model came in accordance with what have been found in similar studies and articles throughout the literature. The results of the estimated model showed that an increase in profitability and liquidity ratios, naturally resulted in a decrease in the default probability of a firm. And an increase in ratios addressing leverage, generates an increase in the estimated default probability of a firm.

Key Words: SME's, HORECA, Business Intelligence, Financial Ratios, Default Prediction.

GENERAL INDEX

CHAPTER 1: INTRODUCTION

Worldwide small and medium enterprises play a major role in most economies, both in developed and developing countries. SME's account for the majority of businesses in the world. According to the World Bank, 600 million jobs will be needed by 2030 to absorb the growing global workforce, and most of them are SME positions. (World Bank, 2020). Therefore, it is important to further develop and invest in SME's growth and stability, and it should be a major priority for any government around the world.

Given their significance, SME's are considered to be an indicator for the general health of an economy. They perform an unquestionable role in supporting the performance of an economy by being the main sources of entrepreneurship abilities, innovation and jobs creation in an economy.

In Portugal, SME's dominate the non-financial business market with 68.3% of value added, 77.4% of employment, and with 99.9% of the market share of the non-financial business. In addition, the value added in Portuguese SME's increased in the period of 2014 and 2018 to 15.2% (European Commission, Portugal SBA factsheet, 2019).

Given the relevance of their role worldwide, the assessment of their financial robustness, i.e. their financial health and wellbeing, their credit risk level and consequently their default probability is very important. SME's are often very exposed to economic downturns and financial crisis. In addition, SMEs are less likely to obtain bank loans compared to larger firms. Instead, often SME's rely on internal funds or cash from friends and family to launch and continue their business (World Bank, 2020).

In order to get a better understanding of risk analysis in general and default prediction in particular, I did an internship in the credit risk field to conclude the Master in Financial

Economics at the Faculty of Economics of the University of Algarve. The internship took place in Yourdata Analytics, Lda, which provides business intelligence, data analytics and artificial intelligence solutions at the service of a business model that innovates the way SMEs work with their business data.

Yourdata's direct clients are whole-sellers/suppliers who deal with their own customers, usually small establishments such as restaurants, snack bars, small shops, etc. Normally, these smaller establishments often ask the whole-sellers (Yourdata's clients) for credit and those in turn ask Yourdata to facilitate the assessment of their clients' (smaller establishment) financial risk by looking into their financial information.

The purpose of the internship was twofold. First, to assist Yourdata in building a database with the clients' financial and accounting information that could later be used for default risk analysis. Second, to use the database to build and estimate a model that could assess the probability of default. Concerning the second goal, it is intended to estimate a model that allows the identification of the determinants of risk default that Yourdata's clients are exposed to.

This internship report follows a conventional structure. After this brief introduction, Chapter 2 introduces the Yourdata Analytics, Lda, the host internship firm, and a picture of the internship project at the host company. Chapter 3 presents describes the internship project the HORECA database. Chapter 4 presents the model´s estimation and discusses the results. Chapter 5 states the main conclusions.

CHAPTER 2: THE HOST FIRM: YOURDATA AND THE INTERNSHIP PROJECT

This section addresses the host company Yourdata. It’s objectives both strategic and S.M.A.R.T, it’s characterization, it’s goals, a glance at the industry of business intelligence. And finally, the internship at Yourdata.

2.1 Characterization of Yourdata

Yourdata was established on May 1, 2015 in the form of a sole shareholder company. On August 5, 2015, the company saw the entry of new members, some of which are actually family members which reflected the interest, commitment and faith of the partners in the success of the firm.

Yourdata strives to be a hub for data and a source of information business intelligence in the Algarve region, and to be a leader in the field of business intelligence in the Portuguese market. One of its goals is to increase the productivity and the profitability of their clients. This is achieved with the help of Yourdata’s knowledge, experience and the use of powerful business intelligence softwares. It is also achieved with the help and support of Yourdata strongest assets - it’s employees. Yourdata aims to help their clients to have a simplified, intuitive and quick way to view, and adjust their performance accordingly.

Yourdata is small in size and young in age, but it’s growing and expanding rather rapidly as the demand and the space for its services is high in the Algarve region. The company comprises of 8 employees at the time writing this report. With two CO-founders at the top handling the procurement of new clients and other main duties, and the rest of the employees handling its daily operations such as working on “Dashboard” (one of Yourdata’s products explained below)”, meeting with current and potential clients and following up with them, in addition to performing other tasks as required.

One of the more important services/products offered by Yourdata and worth mentioning is the “Dashboard”. The “Dashboard” consists of graphs, charts and other easy to understand visuals showing the end users (Yourdata’s clients) how they are doing in terms of profitability, liquidity, risk avoidance and other crucial financial areas. Knowing where they stand, clients can then make the necessary changes to work on areas in which they have difficulties and emphasize on areas where they are strong. The “Dashboard” is fully adaptable and customizable to meet any requirements of the client, and it’s also one of Yourdata’s further important services/products in terms sales.

2.2 Yourdata’s goals, services, the industry of Business Intelligence in general

2.2.1 S.M.A.R.T. and Strategic objectives of Yourdata

S.M.A.R.T objectives stands for Specific, Measurable, Achievable, Relevant and Time-oriented. S.M.A.R.T Objectives are used to guide the development of achievable goals rather than a generalized ones. A simple example of a S.M.A.R.T objective would be to increase sales by 10% in two months this quarter. On the other hand, Strategic objectives are the broad approach on how the company wants to accomplish its goals. They frame what the company wants to do to achieve its mission. An example would be, to launch a new product, to increase the overall profitability, or grow market share for a specific product.

Yourdata S.M.A.R.T objectives are:

- Achieve a turnover of 125 thousand euros in 2018 and 230 thousand euros by 2021.
- Conversion of 50% of clients through follow-ups.
- Achieve a follow-up ratio of 4 companies per data analyst from the beginning of 2018.
- To hire 2 additional Business Analysts and 3 Data Analysts in the year 2018 and hire / subcontract a data scientist.

Contrarily, Yourdata strategic objectives are to be the first business intelligence company in the Algarve region in terms of market share and sales. And to be well known for its innovations in terms of products and solutions offered in the business intelligence sector for the needs of the Portuguese market.

2.2.2 Yourdata's services

Yourdata services comes in three levels. The first level, consists on the Implementation of business intelligence solutions which consists on (1) the standardization and integration of databases. (2) the Introduction of statistical models and predictive analysis; and (3) the offer of consultancy support on the identification of business models and organizational structure as well as training of clients and their staff.
The second level consists of, business monitoring – (Control and Reporting)- through controllers via online Dashboard.

The third level consists on separate works, by transforming data from any database belonging to a client and enabling visual analysis, by designing dashboards from Excel (.xls and .xlsx) files.

2.2.3 The industry of Business Intelligence:

Business Intelligence is a specific area of information systems. It deals with the process of collecting, organizing, analyzing, sharing and monitoring information that supports business management. It is a set of techniques and tools that support the transformation of raw data into meaningful and useful information that is easy and understand by the end user. Business intelligence technologies are able to support a large amount of unstructured raw data to help to identify, develop, and create new business strategy opportunities. The goal of business intelligence is to enable easy interpretation of immense volumes of data. Business intelligence is a relatively new area especially when it's framed as self-service business intelligence, that is, when the final or end user is allowed to construct their own analysis from a visual perspective.

In Portugal, many companies have invested and continue to invest in management softwares. Such as ERP (Enterprise resource planning), CRM (Customer relationship management), website, online stores, communication platforms, online sales, through social networks, etc.. (Yourdata business plan,2018).

Therefore, it is increasingly important to be able to relate data from all of these platforms that generate huge amount of data on a daily basis into actions and decisions that can be implemented quickly. And since business intelligence allows you to be able to understand vast amounts of data intuitively and quickly, investing in business intelligence should to be the next logical step.

Yourdata intends to take advantage of this national context, by introducing existing business intelligence tools in the perspective of Self-Service business intelligence, cloud-based software with existing apps that provide Yourdata clients with high mobility, flexibility, and with the ability to be used on various media such as mobile phone, PC, and tablets. Just to mention, Yourdata mainly implements and uses two BI softwares the PowerBI (Microsoft) and the Tableau Software (Salesforce). These two softwares represent excellent choices for Yourdata, because of their outstanding support (after sales, trouble shooting, upgrades, etc.), quality and for their relative affordability.

Normally there are two types of organizations providing BI solutions, business consultants and software vendors. By extending their value chains both of them may in the future be competitors to Yourdata. First, consultants by integrating programming and data integration skills. Second, software vendors by integrating business management skills.

For these reasons, there are two main differentiating factors that differentiate Yourdata from the rest of the competition. These factors generate competitive advantages because they do not exist in other competitors. These factors are:

- The creation of a team with certified Business Intelligence skills and certificates. A specialization that does not exist in the national market;
- The introduction of the predictive component in data analysis. That is, the introduction of statistical models that provide forecasts of future developments.

2.3 The internship project at Yourdata

The plan for the trainee during the internship was to focus on building a database based of companies in the food and service sector in the Algarve (HORECA). As it would serve later as the source from which the expected model can locate the components necessary to estimate the default probability. The main goal behind the internship was to turn a subjective evaluation into a criterion based objective definer of what credit conditions a client should have based of the data available in the database.

The beginning of the internship focused around heavy data entry and data mining. Then gradually evolved to the analysis of those data to try and get to the empirical proof needed to produce functioning prediction model. The collection of the required financial data and indicators took place from the clients' financial reports, or any other financial data that was obtained by Yourdata.

To get used to the process and as a way to break the ice, the trainee was tasked to build a database for different clients of Yourdata that are based in the construction sector in the Algarve. The database consisted of the names of the companies, the year of founding, EBIDTA, cashflow, sales, purchases, and other financial/ accounting data. The process in the beginning was difficult but after the second week the process became familiar. While building the construction database a problem presented itself which was encountered again later. It is the issues of missing data, misplaced data and data that is not accurately presented.

After finishing the construction sector database, it was the turn to move to the HORECA sector database. Some difficulties were encountered such as identifying the relevant data and where to locate them. Bearing in mind that all of the data were in Portuguese and the trainee didn't understand the language very well. Once again, the same problem of missing data was encountered whilst building the database, but regardless the database was completed in the end.

CHAPTER 3: THE FIRST TASK OF THE INTERNSHIP: THE HORECA DATABASE

The first task of the internship at Yourdata consisted of gathering statistical information on financial indicators of companies located in the HORECA sector, which will later be used to estimate the final model. Given this ultimate goal, the objective of this chapter is to identify and analyze a complete set of financial ratios linked to Portuguese small and medium enterprises (SMEs) and found in the built database.

From the above mentioned purposes, this chapter is structures as follows. First, a set of financial ratios is identified based on a brief literature review. Second, and based on that set of ratios, the database containing information of set of firms belonging to the HORECA sector in Portugal is presented together with some preliminary statistical analysis.

3.1 Identification of financial ratios to support the credit default risk: a brief literature review

In order to identify the adequate and relevant financial ratios, we resort to the literature on default risk assessment which is vast. The seminal works in this field are Beaver (1967) and Altman (1968) which estimated univariate and multivariate models to predict companies default risk using a set of financial ratios. Beaver (1967) analyzed 14 financial ratios, whereas Altman (1968) analyzed 22 ratios but ended up with 5 ratios which were classified into five categories: profitability, leverage, liquidity, activity and solvency. This categorization of financial ratios has been adopted in subsequent studies (e.g., Deakin, 1972; Edmister, 1972; Blum, 1974; Eisenbeis, 1977; Taffler and Tisshaw, 1977; Altman et al., 1977; Bilderbeek, 1979; Micha, 1984; among others) and more recently in Lehmann (2003), Grunet et al. (2004), Keasey and Mcguinness (1990), Altman and Sabato (2005). The mostly considered financial ratios and their categorization are presented in Table 1 as well as their expected effect on the probability of default.

Table 1: Financial ratios and their expected effects of probability of default

Category	Financial ratios	Definition	Expected impact on probability of default
Leverage	Equity to Assets	[Equity] ÷ [Total assets]	-
	Debt to Equity	[Debt] ÷ [Equity]	+
	Liabilities to Assets	[Total liabilities] ÷ [Total assets]	+

	Equity to Liabilities	[Equity] ÷ [Total liabilities]	-
Liquidity	Cash to Assets	[Bank deposits and Cash] ÷ [Total assets]	-
	Capital to Assets	[Working capital] ÷ [Total assets]	-
	Cash to Sales	[Bank deposits and Cash] ÷ [Total sales]	-
	Liquidity to Liabilities	[Bank deposits & Cash Marketable securities] ÷ [Short-term liabilities]	-
Profitability	EBITDA to Sales	[EBITDA] ÷ [Total sales]	-
	Earnings to Assets	[Retained earnings] ÷ [Total assets]	-
	Income to Assets	[Net income] ÷ [Total assets]	-
	EBITDA to Assets	[EBITDA] ÷ [Total assets]	-
	Income to Sales	[Net income] ÷ [Total sales]	-
	Results to Assets	[Net profit & loss] ÷ [Total assets]	-
Activity	Interest to Sales	[Interest costs] ÷ [Total sales]	+
	Sales to Assets	[Total sales] ÷ [Total assets]	-
Solvency	Debt Service	[Current earnings depreciation] ÷ [Interest & similar costs]	-
	Results LT Liabilities	[Net profit & Loss+Depreciation+Provisions] ÷ [Long-term liabilities]	-
	Earnings to Debt Serv.	[Operating earnings] ÷ [Interest & Similar costs]	-

Source: Own elaboration.

To summarize, a consensus seems to exist in the literature that financial ratios covering the areas of leverage, liquidity, profitability, activity and solvency contribute to produce a working model capable of predicting a firm's default probability with a certain degree of confidence.

Given the above identified categories and the correspondent financial ratios used in the literature, the HORECA database was prepared. In its raw form, the database contained data for 1732 firms, covering the period from 2012 to 2016. After much consideration for quality and to avoid missing data as much as possible, 2015 was identified as the year with the least amount of missing data. Therefore, the analysis that follows is based on the information of 105 firms for the year 2015. The raw information collected for each firm is described in Table 2.

Table 2: The available information in the HORECA sector Database

Number of years of operating activity
Total Assets
Total Liabilities
Total sales
Working capital
Cash-flow
Total purchases
Average payment period
Total revenues
Net income
Profit margin ratio (net income ÷ revenues).
Gross margin
Earnings before interests, taxes, depreciations and amortizations (EBITDA)

Source: Own elaboration.

Based on the raw financial data available, several financial ratios were computed in order to cover the categories of financial information often used in default prediction literature. The ratios computed are reported in Table 3.

Table 3: Computed financial ratios of firms in the HORECA sector

Category	Financial ratios	Description
Profitability Ratios	EBITDA to Sales	[EBITDA] ÷ [Total sales]
	Earnings to Assets	[Retained earnings] ÷ [Total assets]
	Income to Assets	[Net income] ÷ [Total assets]
	EBITDA to Assets	[EBITDA] ÷ [Total assets]
Leverage Ratios	Equity to Liabilities	[Equity] ÷ [Total liabilities]
	Liabilities to Assets	[Total liabilities] ÷ [Total assets]
	Cash to Assets	[Bank deposits and Cash] ÷ [Total

Liquidity Ratios

		assets]
	Capital to Assets	[Working capital] ÷ [Total assets]
	Cash to Sales	[Bank deposits and Cash] ÷ [Total sales]
Activity Ratio	Sales to Assets	[Total sales] ÷ [Total assets]

Source: Own elaboration.

In summary, the analysis considers 4 of the 6 profitability ratios, 2 of the 5 leverage ratios, 3 of the 4 liquidity ratios and 1 of the 2 activity ratios identified in the literature. Unfortunately, it was not possible to compute solvency ratios from the available information.

3.2 The HORECA database: Preliminary statistical analysis

Basic descriptive statistics for the selected financial ratios of 105 HORECA sector firms for the year 2015 are reported in Tables 4 to 7 covering profitability, leverage, liquidity and activity ratios.

All profitability ratios report negative average values and relatively low standard deviations, with the exception of the ratio of EBITDA to sales.

The negative average ratio of EBITDA to sales suggests that the average firm is extremely poor at producing more earnings by keeping the costs low. At the same time there are no earnings remaining after operating expenses are paid.

The negative average ratio of retained earnings to total assets suggests that, on average, firms are more reliant on common debt financing and less retaining their profits and using them to finance assets.

The negative average value for the net income to total assets ratio suggests that the average firm is less than efficient in utilizing its assets in order to generate revenues.

The mean for the EBIDTA to total assets ratio is also negative, which suggests that, on average, firms are not using efficiently their assets to generate earnings.

TABLE 4: PROFIABILITY RATIOS

	Minimum	Maximum	Mean	Std. Deviation
EBITDA to Sales	-37000.00	620.69	-359.6173	3612.01330
Retained Earnings to Total Assets	-31.76	8.45	-.1054	3.65645
Net Income to Total Assets	-25.39	2.03	-.2281	2.56533
EBITDA to Total Assets	-11.17	2.27	-.0490	1.26769

Source: Own elaboration.

The leverage ratios report relatively low average values and also low standard deviations, with the exception of the debt to equity ratio.

The mean for the debt to equity ratio is 1.14, which suggests that the average firm has a good equity (owners) to debt ratio allowing it to cover outstanding debt. And that the average firm is financing its operations through wholly owned funds rather than debt. However, the relatively high standard deviation suggests different situations for the firms included in the database.

The negative average value for the equity to total liabilities ratio indicates that the average firm has more liabilities than assets.

The average value for the liabilities to total assets ratio is only marginally different from zero, which suggests that, on average, firms have funded their assets through equity rather than debt.

TABLE 5 - LEVERAGE RATIOS

	Minimum	Maximum	Mean	Std. Deviation
Debt to Equity Ratio	-100.00	70.92	1.1409	13.11803
Equity to Total Liabilities Ratio	-12.830	.940	-.30098	1.738418
Liabilities to Total Assets Ratio	-5.88	5.56	.6415	1.23496

Source: Own elaboration.

The liquidity ratios report relatively low average values and also low standard deviations, with the exception of the cash to sales ratio. The average values of the cash to total assets ratio and the working capital to total assets ratio are only marginally positive. These values suggest that the average firm has a little bit of liquidity to manage its immediate responsibilities and is barely liquid enough in the short term a realistic commonality amongst SMEs. Additionally, the expressive negative average value of the cash to sales ratio suggests that, on average, firms are extremely poor in generating cash from their sales.

TABLE 6 - LIQUIDITY RATIOS

	Minimum	Maximum	Mean	Std. Deviation
Cash to Total Assets	-4.03	2.05	.0746	.59067

WC to Total Assets	-4.03	2.05	.0746	.59067
Cash to Sales	-38816.50	262.38	-379.4547	3788.63041

Source: Own elaboration.

The activity ratios are considered in this analysis by the sales to total assets ratio. This ratio ranges from -2.09 to 3.63, corresponding to a relatively reduced standard deviation and a reduced average value, which suggests that, on average, firms are not as efficient as possible in generating sales from as little assets as possible.

TABLE 7 – ACTIVITY RATIO

	Minimum	Maximum	Mean	Std. Deviation
Sales to Total Assets	-2.09	3.63	.1325	.54897

Source: Own elaboration.

The bivariate correlation analysis between the financial ratios is reported below in Table 8. In general, it reveals no high significant values between the ratios, which suggests no serious multicollinearity in considering them as possible predictors of firms' credit default risk.

Table 8: Bivariate correlations between financial ratios

Ratios	EBITD to Sales	Earnings to Assets	Income to Assets	EBITDA to Assets	Debt to Equity	Equity to Liabilities	Liabilities to Assets	Cash to Assets	Capital to Assets	Cash to Sales	Sales to Assets
EBITDA to Sales	1.00										
Earnings to Assets	-0.018	1.00									
Income to Assets	-0.046	0.105	1.00								
EBITDA to Assets	-0.074	0.154	0.960**	1.00							
Debt to Equity	0.011	0.028	-0.072	-0.061	1.00						
Equity to Liabilities	0.279**	0.115	-0.062	-0.078	0.059	1.00					
Liabilities to Assets	0.076	-0.552**	-0.160	-0.192	-0.014	0.259**	1.00				
Cash to Assets	-0.148	0.299**	0.225*	0.445**	-0.002	-0.140	-0.361**	1.00			
Capital to Assets	-0.159	0.218	0.122	0.406**	-0.004	-0.123	-0.396**	0.535**	1.00		
Cash to Sales	1.00**	-0.018	-0.048	-0.075	0.011	0.278**	0.075	-0.148	-0.148	1.00	
Sales to Assets	0.024	0.064	0.008	0.029	0.022	-0.014	-0.076	-0.080	-0.080	0.024	1.00

Source: Own elaboration.

*Correlation significant at the 5% level (2-tailed); ** Correlation significant at the 1% level (2-tailed)

CHAPTER 4: THE SECOND TASK OF THE INTERNSHIP: BUILDING THE DEFAULT RISK MODEL

Given the information of several financial ratios reported in the database, the next task outlined in the internship work plan is the construction of a model capable of predicting companies' default risk.

In this chapter the data described previously is used in a model to predict the default risk and at the same time to identify the most important determinants.

To measure the firms' credit default risk, an additional indicator was used. This indicator is collected from external audit reports conducted by INFORMA Portugal. This indicator, gives each firm a score reflecting its risk level. This indicator takes discrete values, ranging from 1 to 20, with higher values corresponding to lower credit default risk and vice versa.

For ease of analysis, the variable is recoded with values from 1 to 4. Level 1 corresponds to low risk and is given to firms with INFORMA scores between 15 and 20; the level 2 corresponds to medium risk and is given to firms with INFORMA scores between 11 and 14; the level 3 corresponds to high risk and is given to firms with INFORMA scores between 6 and 10; and finally, the level 4 corresponds to very high risk and is given to firms with a INFORMA scores between 1 and 5.

4.1 Econometric model specification

A multinomial unordered logistic regression approach is adopted given, on the one hand, the multi-categorical nature of the dependent variable and, on the other hand, the unordered nature of the categorical dependent variable (confirmed by the Brant's test, Brant (1990)). This setting allows the estimation of the likelihood of each level of risk (low level/medium level/high level/very high level).

A stepwise regression is first conducted to eliminate statistically non-significant explanatory variables, using a stopping rule based on a p-value below 0.05. The results of this analysis suggest the exclusion of several financial ratios as predictors. For the model specification purposes, the selected satisfactory predictors are specified as follows:

$EQUITY\_LIAB$ - ratio of Equity to Total Liabilities (Leverage ratio);
$CASH\_SALES$ – ratio of Cash to Sales (Liquidity ratio);
$EBITDA\_SALES$ – ratio of EBITDA to Sales (Profitability ratio)

Given the set of covariates, the multinomial model for the risk default level, composed of three equations is defined as follows:

$$\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i = \gamma_0 + \gamma_1 EQUITY\_LIAB_i + \gamma_2 CASH\_SALES_i + \gamma_3 EBITDA\_SALES_i + \varepsilon_i \quad (1)$$

$$\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i = \delta_0 + \delta_1 EQUITY\_LIAB_i + \delta_2 CASH\_SALES_i + \delta_3 EBITDA\_SALES_i + \vartheta_i \quad (2)$$

$$\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i = \alpha_0 + \alpha_1 EQUITY\_LIAB_i + \alpha_2 CASH\_SALES_i + \alpha_3 EBITDA\_SALES_i + \varphi_i \quad (3)$$

where Y represents the variable "risk default"; γ_0, δ_0, and α_0 are the equations' intercepts; γ_j, δ_j and α_j(for $j = 1, 2, 3$) are the predictors' coefficients; and ε, ϑ and φ are the usual stochastic error terms; $\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i$, $\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i$ and $\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i$ represent the logarithm of the odds ratio of the different risk levels, specifically, "low level", "medium level", and "high level", for firm i, respectively.

Using some algebraic manipulation, we arrive at the expressions of the estimated probabilities of each level of default risk, given the values of the covariates, which are given by:

$$P(Y=1|X_i) = \frac{e^{\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i}}{1+e^{\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i}} \quad (4)$$

$$P(Y=2|X_i) = \frac{e^{\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i}}{1+e^{\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i}} \quad (5)$$

$$P(Y=3|X_i) = \frac{e^{\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i}}{1+e^{\ln\left(\frac{P(Y=1)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=2)}{P(Y=4)}\right)_i}+e^{\ln\left(\frac{P(Y=3)}{P(Y=4)}\right)_i}} \quad (6)$$

$$P(Y=1|X_i) + P(Y=2|X_i) + P(Y=3|X_i) + P(Y=4|X_i) = 1 \quad (7)$$

where X_i represents the independent variables for firm i.

4.2 Estimation results and discussion

This section is devoted to the presentation and interpretation of the estimation results and its structure is as follows. First, we report the coefficients' estimates of each equation together with the diagnostic tests related to the models' quality and adequacy. Second, we assess the relative importance of each covariate to determine the level of default risk. All estimation and diagnostic testing procedures are performed using the software IBM SPSS.

The estimation results for Equations (1) to (3) are reported in Table 9 The reference category of the dependent variable was set at 4. Table 9 is horizontally divided into three parts, each reporting the coefficients' estimates of each equation of the multinomial logistic model.

The upper part reports the estimates of the coefficients for Equation (1); the central part reports the estimates for Equation (2); and, finally, the lower part reports the estimates for Equation (3).

The estimation results provide useful and interesting information insights on the importance of each covariate on the logs of the odds ratios and therefore on the relative likelihood of each level of default risk.

The estimation procedure guarantees that all the estimated coefficients are statistically significant at a 5% level of significance or lower. All coefficients are positive, accordingly to what is found in the literature. Thus, the results suggest that higher levels of the ratio of Equity to Total Liabilities, the ratio of Cash to Sales, and the EBITDA to Sales increase the probabilities associated to lower levels of default risk.

Table 9: Multinomial model estimates

Risk		Coefficients	Std. Errors	Wald statistic	Degrees of freedom	P-values
1.00	Intercept	-0.275	1.077	0.065	1	0.799
	Equity to Liabilities	28.489	10.975	6.738	1	0.009
	Cash to Sales	0.679	0.293	5.364	1	0.021
	EBITDA to Sales	0.306	0.143	4.599	1	0.032
2.00	Intercept	0.298	1.049	0.081	1	0.776
	Equity to Liabilities	26.335	11.030	5.701	1	0.017
	Cash to Sales	0.653	0.270	5.866	1	0.015
	EBITDS to Sales	0.365	0.141	6.703	1	0.010
3.00	Intercept	0.087	1.069	0.007	1	0.935
	Equity to Liabilities	25.435	10.855	5.490	1	0.019
	Cash to Sales	0.672	0.279	5.798	1	0.016
	EBITDA to Sales	0.289	0.123	5.529	1	0.019

Note: The reference category is: 4.00

Source: Own calculation using SPSS.

A set of diagnostic tests were considered to assess the quality of the estimation results and they are reported in Table 10.

Briefly, they suggest joint statistical significance of the covariates in the model, as the p-values reported in the models' fitting information confirm the rejection of the null hypothesis of joint non-relevance. The goodness-of-fit tests exhibit high p-values associated to the Pearson and Deviance test statistics, leading to the non-rejection of the null hypothesis of data good fit. These results of the diagnostic tests are not independent from the fact that all variables exhibit statistically significant Wald statistics at the significance level of 5% or lower. Additionally, the Pseudo R-squares assume very interesting values, in particular the Nagelkerke R-square according to which about 80% of the variability of the dependent variable is explained by the predictors' joint variability.

Table 10: Diagnostic tests and Pseudo R-squares

	Test statistic	P-value	
Diagnostic tests			
LR test	141.869	0.000	
Pearson goodness of fit	225.620	0.991	
Deviance goodness of fit	136.586	0.992	
Pseudo R-Squares			
	Cox and Snell		0.741
	Nagelkerke		0.797
	McFadden		0.509

Source: Source: Own calculation using SPSS.

The good performance of the model is also confirmed in Table 11. This table crosses the number of firms positioned in each risk level in the sample with the number of firms that is predicted by the model in each risk level. The results reveal that the 88.2% of the firms with low risk level, 45.0% of the firms with medium risk level, 13.3% of the firms with high-risk level and 100.0% of the firms with very high-risk level are correctly classified by the model. On average, the model predicts correctly 73.3% of the cases. Moreover, given the predictors considered, the model reveals particularly good performance in predicting very high-risk levels and very low risk levels.

Table 11: Classification table

	Predicted cases				
Observed cases	1.00	2.00	3.00	4.00	Percent Correct
1.00	30	4	0	0	88.2%
2.00	11	9	0	0	45.0%
3.00	10	3	2	0	13.3%
4.00	0	0	0	36	100.0%
Overall Percentage	48.6%	15.2%	1.9%	34.3%	73.3%

Source: Own calculation using SPSS

Using Expressions (4) to (7) it is possible to compute the probabilities associated to each credit default risk, for particular values of the predictors. According to the estimation results, and considering the predictors at their sample average values, the probabilities of each risk level are $P(Y = 1) = 0.000431, P(Y = 2) = 0; P(Y = 3) = 0.999 + P(Y = 4) = 0$. Therefore, firms with predictors at the average values will be according to the model, considered high risk. Those with predictors below the average values will be according to the model, considered of very high risk; finally, those firms with predictors above the average values will be classified either as average or low risk.

Chapter 5: CONCLUSIONS

The internship efforts personified in this report aimed to find a possible link between the data that was collected during the course of the internship, and the default probability of firms in the HORECA sector in the Algarve region. The goal was to build a model that is able to some extent of predicting the default probability of firms in the HORECA sector solely based of the data collected during internship. To go about it, the multinomial regression methodology was the best fit for the task, first, because of the multi-categorical nature of the dependent variable. And second, the unordered nature of the categorical dependent variable.

The literature on similar endeavors such as the ones perused in this report, specifies that for a model estimating the default risk of a firm, the presence of financial ratios covering the five categories of liquidity, profitability, activity, leverage and solvency is crucial in order to have adequate results.

Form the data collected, only four ratio categories out of five were studied, due to the limitations of the data that was available. After further testing only three out of the four ratios proved to be of significance. This paved the way to finalize the model in its final shape which contained three ratios.

The results of the estimated model came in agreement with similar studies and articles found in the literature, demonstrating that firms showing an increase in ratios covering the area of liquidity and profitability would result in a decrease in the probabilities associated with default. And firms with an increase in ratios covering the area of leverage would result in an increase in the probabilities associated with default.

This signifies that having a model with three ratios covering the areas of leverage, profitability and liquidity should present results that are acceptable.

REFRENCES:

- E.C., 2019, SBA Fact Sheet PORTUGAL, The European Commission Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. Available at: https://ec.europa.eu/docsroom/documents/38662/attachments/23/translations/en/renditions/native [Accessed 23 rd June 2020].
- Altman, E.I. & Sabato, G., 2005. Modeling Credit Risk for SMEs: Evidence from the US Market. SSRN Electronic Journal.
- Andreeva, G., Calabrese, R. and Osmetti, S., 2016. A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models. European Journal of Operational Research, 249(2), pp.506-516. Anon, SMEs in Portugal – a brief fact check.
- Ciampi, F. & Gordini, N., 2012. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. Journal of Small Business Management, 51(1), pp.23–45.
- Ergeshidze, A., 2017. Credit Risk Model: Assessing Default Probability of Mortgage Loan Borrower. SSRN Electronic Journal.
- Figini, S., Bonelli, F. and Giovannini, E., 2017. Solvency prediction for small and medium enterprises in banking. Decision Support Systems, 102, pp.91-97.
- Figini, S., Bonelli, F. & Giovannini, E., 2017. Solvency prediction for small and medium enterprises in banking. Decision Support Systems, 102, pp.91–97.
- World Bank. 2020. World Bank SME Finance. Available at: https://www.worldbank.org/en/topic/smefinance. [Accessed 23 rd December 2020].
- Kousenidis, D., Ladas, A. and Negkakis, C., 2019. Aggregate Accounting Data and the Prediction of Credit Risk. The International Journal of Accounting, 54(01), p.1950001.
- Lucic, L., 2014. Financial ratios in the function of business risk assessment. Iiakm.org. Available at: http://www.iiakm.org/ojakm/articles/2014/volume2_3/OJAKM_Volume2_3pp21-34.pdf [Accessed 08 th April 2019].

- Motta, V. and Sharma, A., 2020. Lending technologies and access to finance for SMEs in the hospitality industry. International Journal of Hospitality Management, 86, p.102371.
- Malakauskas, A., 2017. MODELLING CREDIT RATING OUTLOOK FOR SME ENTITIES IN THE BALTIC STATES. Available at: https://vb.ism.lt/object/elaba:22780253/ [Accessed 09th June 2020].
- Memic, D., 2015. Assessing Credit Default using Logistic Regression and Multiple Discriminant Analysis:Empirical Evidence from Bosnia and Herzegovina. Interdisciplinary Description of Complex Systems, 13(1), pp.128–153.
- Soares, João & Pina, Joaquim & Ribeiro, M.S. & Catalão-Lopes, Margarida, 2011. Quantitative vs. Qualitative Criteria for Credit Risk Assessment. Frontiers in Finance and Economics. 8. 69-87. 2011. SSRN Electronic Journal.
- Stephanie Glen, 2017. "Multinomial Logistic Regression: Definition and Examples" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/multinomial-logistic-regression/ [Accessed 13th July 2020]
- World Bank, 2020. Small and Medium Enterprises (SMEs) Finance Improving SMEs' access to finance and finding innovative solutions to unlock sources of capital. Available at: https://www.worldbank.org/en/topic/smefinance [Accessed 13 th March 2020].