

Credit risk evaluation and rating for SMES using statistical approaches: the case of European SMES manufacturing sector

Kyriazopoulos Georgios¹

Abstract

The prevention of financial losses is crucial for enterprises, especially in periods of market instability and uncertainty. Credit risk refers to the likelihood that a company will not be able to cover its liabilities and become insolvent and defaulted. Credit risk is of utmost importance not only for the enterprises but also for financial institutions (banks), which try to eliminate any possible losses from insolvent clients. Most of the enterprises in Europe are SMEs (Small and Medium Enterprises). Manufacturing sector is one of the most important, especially in Western Europe. The aim of the current study is to evaluate credit risk of European SMES manufacturing companies for the period 2012-2014 under different schemes, with the use of a popular statistical approach, namely logistic regression. The results of the analysis imply that even with a mixed and unbalanced data set with a small number of defaults, the applied method perform well and provide meaningful results. The results of this paper could help the owners and the financial managers of SMEs in European Union in their financial decisions and strategic investments so as to be able to avoid credit risk and future bankruptcy. More viable SMEs in European Union may mean more development and less unemployment.

JEL classification numbers: G30, G32, G33

Key Words: Credit risk, SMEs, Manufacture, Logistic Regression.

1 Introduction

¹ Technological Educational Institute of Western Macedonia, Greece

The granting of credit by a company is a crucial issue that require delicate care (Bohn & Stein, 2009). For both financial and nonfinancial corporations, it is very important to evaluate the risk profile of a debtor in a proper way. The ability to discriminate good customers from bad ones is crucial. Wrong credit decisions can have severe consequences: the refusal of a good credit can cause the loss of future profit margins, and the approval of a bad credit can cause the loss of the interest and the principal money. The necessity for reliable models that predict defaults accurately is imperative, in order to enable the interested parties to take either preventive or corrective action. Accurate risk assessment allows the financial institution to apply a correct request for collaterals in relation to the risk and with appropriate guarantees.

In an era of business market instability, with significant evolution of technologies and social demographics, a corporation has to deal with a very wide range of changing factors that creates many risks, hazard or unexpected losses (Boreiko, et al., 2016). Corporate financial management is important and have to be effectively insured in order to keep the corporation as healthy as possible.

Risk assessment and credit classification is based mainly in scoring models. A reason for this, is the humans' lack of capability to judge the worthiness of a loan and discover the useful relationships or patterns from the data (Saunders & Allen, 2002), together with the large volume of the data to be examined, and the nature of the relationships themselves that are not obvious. (Agrawal, et al., 2012). These models are constructed with the use of large number of credits and loans in the past and support the decision process consistently and efficiently. With the assistance of these models, loan applications can be categorized into good and bad applications.

The study starts with the clarification of terms of corporations in the instable and uncertain modern business market, following by a discussion of main risk categories which affect the corporations.

Due to the importance of credit risk analysis, we discuss some early empirical approaches (for example linear discriminant analysis (LDA)), and more modern such as support vector machine (SVM), that are used in the field of corporate credit rating, together with the introduction of some common known credit rating agencies.

Following into the analytical part, we used Logistic Regression method to predict and specify credit risk model predictability.

Regarding the significance contribution that the European SMEs provide to the European economy and in which it represents the largest portion of the European companies, the case of the European Manufacturing SMEs has been chosen to be examined in the research. A description regarding the European Manufacturing SMEs business environment, financial risks, and credit climate is introduced in section 3.

Section 4 describes the research design and methodology which illustrates the research process and the analytical flow of the research.

Section 5 includes the specifications regarding the obtain data design, description and statistics.

This study concludes with a discussion of the overall study results, with emphasis on the possible direction for future research that might be taken in this filed.

2 General Overview of the Corporation Environment

2.1 Corporations, and Business Market Instability

Corporations are the entities that operate in the business market seeking profits (Rottig, 2006; Vargo, 2011). There is a difference between the financial and the nonfinancial markets. The financial market is the market where to trade bonds, bills of exchange, commodities, foreign currency etc. (Bokpin, 2010) The non-financial market is the market that deals with the production of goods and nonfinancial transactions and services. (Verbeke, 2005).

The current marketplace is facing an increasing number of diversified problems. (Wickens, 2016), in his study of the market crisis in the euro zone, indicates an ongoing, and a higher level of market instability which requires attention by the corporations and the working businesses. Mouna & Anis, 2015, examine the effects of the economic crisis in different zones including Europe, USA and China. The studies raise many warning and critical issues that have to be considered by corporations to keep effective operations. Regarding the crisis and the market instability, many other studies, researches and tools have been introduced, trying to find a way to treat such a problematic market dynamics and fast-changing components.

2.2 Corporation and Risk

Derived from the uncertainty in the corporate markets, corporations have to deal with big difficulties related to the internal and external environment (Macro & Micro Environment). The major cause of the corporations' problems are issues related to the poor risk management. Risk is a future unexpected action that might affect the corporation and lead it to bankruptcy. Wherefore, corporation has to prevent itself from any lack of attention given to the surrounding circumstances and factors. Otherwise, the corporation will be in danger of bankruptcy.

Corporations set their strategies, procedures, plans and they follow many methodologies just to insure the perfect treatment of the future and unexpected risks. The lack of visionary of future events is a severe uncertainty. "Uncertainty is an elusive and immeasurable concept" (Salame, 2007). Since, the uncertainty is immeasurable, we, therefore, have to keep the environment as controlled as possible and setting strategies that doesn't have a wide gap of the real market and world. In the time of uncertainty, corporate have to deal with many types of risk and treat them according to their field of occurrence and burden.

"Major cause of serious and related systems problems continues to be directly related to negligent credit standards for borrowers and counterparties" (Salame,

2007). The credit risk is the risk associated with the customers' ability to pay their debts back which is the most severe risk in the matter of corporate monetary safety and the corporate solvency market stability (Gestel & Baesens, 2009).

2.3. Credit Risk

Credit risk is the risk associated with the corporation's ability to pay its debts back and the financial institution ability to get its money back (Hotchkiss & Altman, 2006). Alternatively, credit risk can be defined as the possibility of loss incurred as a result of a borrower or counterparty failing to meet its financial obligations. Credit risk and default, are similar terms in a way that the worst scenario that can occur in a company that has credit risk problems is to default.

Two main concepts of default can be distinguished (client oriented and transaction orientated). The first one, client oriented, focus on the client's likelihood of default. Here, all transactions done with the above client have the same probability of default, this means that are fully dependent to each other. In the second one, transaction oriented, default takes place when a contract is terminated. This is more likely to appear in cases when investors hold many financial products, with different characteristics. This means that default can occur, but in different time frame. (Wehrspohn, 2002).

In order to evaluate credit risk many researchers use credit scoring (Abdou & Pointon, 2011). Thomas, et al., 2002, comment about the philosophy behind credit scoring as "Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit". This depicts that the corporate credit rating or scoring is the system of choosing the appropriate techniques to assess the customers' probability to default or getting bankrupt. These techniques decide who will get credit, how much they should get, and what operational strategies will enhance the profitability of the borrowers to the lender (Siddiqi, 2006). Credit rating could be defined as a process in which the lender assesses the borrower's creditworthiness and reflects the circumstances that will occur for both sides, and defines the lender's view of potential future economic scenarios (Thomas, et al., 2002).

Eventually, after the assessment of the participants for credit by using different tools and techniques regarding the preference of the decision maker, the examined firm would be rated and divided into two groups (defaulted / non-defaulted).

3 A review of different approaches in the field of corporate credit rating and business failure prediction

3.1 Corporate Credit rating and Business Failure Research: Statistics, Methods, Models and Variables

The terms of failure, insolvency, default, and bankruptcy are major terms for discussion in the area of credit risk (Zopounidis & Dimitras, 1998). These terms are varying in definition regarding the condition of the firm. According to Altman,

et al., 1994 the term of failure means that the actual rate of return on the invested capital with the risk and unexpected events is significantly lower than the normal return of similar investments. The term of insolvency defines the situation of the liquidity problems or performance defect. The default is the term that deals with the firm that violates a condition of an agreement with a creditor and can make a legal action. Bankruptcy is the point when the business liquidates or make a reorganization program resulted from a severe loss of the net worth of the business.

Many methods, models and approaches have been used to evaluate the credit risk and the businesses' default. Some empirical methods have been introduced by American banks to assess and predict the businesses' failure. Methods like, "Five C" (Character, Capacity, Capital Condition, Coverage), The "LAPP" method of (Liquidity, Activity, Profitability, Potential), and the "Credit-Men" Method. (Zopounidis & Dimitras, 1998). Traditional methods of customers' evaluation depend mainly on the short-term condition of the participant, and it does not go deeper in the research and the analysis of the multivariate and long-term risks and default.

Following the traditional methods of default, ratios statistics, analysis, models started to be introduced as a way for better assessment of the creditworthiness and default prediction.

The early empirical approaches depended on the analysis of the financial ratios and the financial statements analysis. (Atiya, 2001). One of the first pioneers in the field of bankruptcy prediction was Altman with the use of multiple discriminant analysis (MDA) for the analysis of the financial statements data and the creation of the Z-Model. Another linear model has been introduced by Ohlson. Ohlson's model was used for bankruptcy prediction problems (Thomas, et al., 2002).

3.2. Logistic Regression

Logistic regression is a popular statistical method that examines and describes the relationship between a categorical response variable and a set of predictor variables. In the field of credit rating and corporate failure prediction, Logistic Regression works as a probabilistic indicator of the default dealing with binary or dichotomous variables. Logistic regression considers a predictive model for a qualitative response variable. One of the first logistic regression models has been introduced by Wiginton (1980). The model matches the probability odds by a linear combination of the characteristics variables. (Thomas, et al., 2002).

Wiginton 1980, introduced model formula, as following:

$$\log\left(\frac{P_i}{1-P_i}\right) = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p = y^* \quad (1)$$

This model is defined in term of convenient values to be interpreted as probabilities that the default might occur under different criteria. Also, the model

specifies that an appropriate function of the fitted probability of the event is a linear function of the observed values of the available explanatory criterions.

The left-hand side of the model defines the logit function of the fitted probability

$\log\left(\frac{P_i}{1-P_i}\right)$, as the logarithm of the odds for the event, namely the natural logarithm of the ratio between the probability of occurrence (Success), and the probability non-occurrence (Default).

The right-hand defines the normal linear model that concludes the variables that are used in the evaluation and their weights. i.e. $(X_1, X_2, X_3, \dots, X_p)$, are the representatives of the different factors that are significant for the discriminant process of the participant evaluation, and W_i representing the variable's effect in the participants' evaluation process.

To calculate the direct value of the probability, the probability formula can be derived as:

$$P_i = \frac{\exp(y^*)}{1 + \exp(y^*)} \quad (2)$$

The value that P_i takes must be between 0 and 1 because of that the $\frac{P_i}{1-P_i}$ takes the

value between 0 and ∞ , $\log\left(\frac{P_i}{1-P_i}\right)$ takes value between $-\infty$ and $+\infty$ (Thomas, et. al, 2002).

After the calculation of probability P_i , the value of each binary observation can range between 0 (minimum value) and 1 (maximum value). In most cases, there is also an error, where the target is to be as low as possible. In contrast to linear regression, here there is no option to decompose the observed values into the sum of the fitted value and an error term. (Salame, 2007).

A reason why to choose logit function towards linear function in order to link probability (P_i) to the linear combination of the explanatory variables, has to do with the fact that in the case of logit function probability tends toward 0 and 1 gradually. On the contrast, in linear function, probability can take values outside the interval, 0 to 1, which would be meaningless.

A logical S-shaped curve has been introduced by Giudici 2003, implies that the dependence of P_i on the explanatory variables is described by a sigmoid or S-shaped curve.

Different values of the unique explanatory variable, link to different range values of the success probability. Owing to the previous fact, the behavior of logistic curve can be visualized (Giudici, 2003).

A practical use of the logistic regression method has been made by Memić, 2015, assessing the default probability of 1196 different size Bosnian, Herzegovinian and Serbian companies (Memić, 2015).

3.3. Neural-Networks (NN)

The strength of the nonlinear and NN approaches derives from its ability to give a better problematic interpretation of the correspondence between the multivariate factors and the default (Gepp & Kumar, 2012).

A neural network consists of neurons which are organized in layers. Three types of layers can be found (input, output and hidden). The role of an input layer is to receive information from the external environment and transmit it to the next level. Output layer is the one that produces the final results. Hidden layers are the ones between input and output layers. Their role is only for analysis, converting input to output variables. The number of layers can vary dependent on the problem and its complexity. According to (Boguslauskas & Mileris ,2009), some authors count all the layers of neurons and others count the number of layers of weighted neurons.

The application of the Neural network in field of credit rating and default prediction can be reviewed in studies that have been done by, Handzic, et al., (2003), and Atiya, (2001).

3.4. Support Vector Machine (SVM)

Support vector machines (SVMs) use a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high-dimensional feature space (Min & Lee, 2005). SMV is a method uses for separable binary sets of ratios, and it goals to set a common hyperplane that classifies all training vectors in two classes. (Wu et al. 2004)

A study of bankruptcy prediction is done by Min & Lee, 2005. Min & Lee, 2005 used SVM method as a main prediction methodology of the bankruptcy prediction and compared the results of the model with other different methodologies of default prediction. The result shows that the use of the SVM in the bankruptcy prediction has better prediction results compared with other existing methods.

4 An overview of the European manufacturing sector

In this section we give a brief description of the European Manufacturing sector. We discuss, define, and analyze the main circumstances, surrounding influences, and the role-playing factors in this sector.

4.1 Manufacturing - Manufacturing in Europe

4.1.1 Manufacturing

The manufacturing sector is product oriented sector. Manufacturing is the process of transforming the form of raw materials in nature and their content to increase their value and using appropriate tools to make them satisfy a particular need, whether intermediate or final.

The manufacturing sector is an important pillar of long-term development in the economy as one of the most important sectors of diversifying sources of national income, reducing reliance on traditional sources and meeting the needs of civil

society in its continuous development and achieving greater value for natural resources through achieving value added (Sweeney, et al 2016).

4.1.2 Manufacturing SMEs and industrial growth

Manufacturing industries are flexible and one of the most responsive industry to benefit from (Bulak & Turkyilmaz, 2014). The benefits of manufacturing, seeking the satisfaction of the customers' needs by converting the materials and what is extracted from the land are crucial and are increasing day by day, taking into consideration the limitations of the resources (natural resources and human resources).

Humanity moved from the era of the industrial revolution to the age of scientific and technological revolution based on science and scientific research with discoveries in the science of mathematics and physics which are the basis of nuclear fission, nuclear industry, electronic computers as well as the discoveries of chemistry of different kinds, biology which is the basis of changes in agriculture and medicine, to accelerate modern manufacturing processes and very broad production and technical progress. This growth and change in the manufacturing sector have significantly affected the European SMEs either positively by creating more market chances or negatively by creating more severe challenges these SMEs need to deal with (Wilson, et al, 2006).

4.1.3 Manufacturing in Europe

In Europe, the manufacturing sector is a distinguished sector among the other market sectors in the union. European joint ventures appeared early in the European Union, and included many industrial and commercial fields. The most important industrial activities of the Union include the automobile industry, aircraft, heavy machinery and engines. Europe has many major industrial groups. The European Union ranks first in the automotive industry.

Many industries are in conflict with European laws that are bound to preserve the environment, European capital flows for investment and industrialization in other regions outside the EU or the continent as a whole (Scapolo, et al, 2003).

According to EU data, the average labor productivity was € 55.0 thousand per employed person (€46.9 thousand per working person). Regarding the labor cost, it was equivalent to € 38.3 thousand per employee. The value added per person was equivalent to 143.0% of the average staff costs per employee, close to the levels of the other sectors. Moving forward to further data, the overall gross operation rate was 7.9% and found to be the second lowest sector of profitability. (Source: NACE Rev2, May 2017).

5 The Research Design

5.1 The Goal of the Research Design

The research design and analysis will focus on testing the effectiveness and the efficiency of Logistic Regression approach for the sake of the corporate overall benefit and wealth maximization under different schemes. For the evaluation of credit risk, a multi criteria credit rating model will be developed. The model creation process will keep the connection between the operational tools usage (the use of the multi criteria approaches) and the core strategic goal of decreasing the financial and credit risk. The aim of this approach is the minimization of the corporate credit risk.

For building a harmonized model, we should start with the understanding of the strategic risk management process (Iazzolino & Laise, 2012).

The financial ratios that going to be used in the analysis belong to five main groups, similar to the ones found in literature review.

5.2. Data Description and Statistics

5.2.1 Data Description

The data used in the research analysis are obtained and collected from financial and accounting statements of European manufacturing SMEs. Each financial ratio in the data set describes different aspects of the overall financial situation of the examined firms. This study's data have been obtained from the ORBIS database of Bureau van Dijk (BvD). ORBIS database is a commercial database that contains administrative and financial information of over 50 Million European Companies. The data obtained from six European countries, namely United Kingdom, Germany, France, Belgium, Italy and Spain. The study period is from 2012 to 2014, including data of three years (2012, 2013, 2014) which have been split into two samples, training sample and testing sample. Companies data of 2012 and 2013 would be used as the training sample and 2014's data would be used as the testing sample. Training sample is the sample to be used for model building, and the testing sample is the data to be used for the model's validation and usability test. The total number of the companies that are going to be used in the analysis is 25875. The data obtained from unlisted firms which are companies with stocks that is not traded in the exchange market.

The data consists of two types of companies

1. Active / "Non Distressed Companies": The working companies in the manufacturing sector at the data collection period.
2. Distressed: Bankrupted or non- liquidation companies at the time of data collection.

Regarding the significance, 12 ratios have been chosen for the modeling process which are discussed below. The chosen ratios belong to 5 main categories which are:

1. Liquidity
2. Profitability
3. Leverage
4. Activity, and
5. Efficiency.

5.2.2. Data Statistics

Tables.1 to.4, explain and illustrate the overall statistics of the used data for the analysis and the models building.

Table 1: Total Number of companies Per Country and Year

Total Number of companies Per Country (Active + Distressed)				
Country/ Year	2014	2013	2012	Total
Belgium	441	535	559	1535
France	1189	1161	1108	3458
Germany	847	1016	1012	2875
Italy	3467	3380	3528	10375
Spain	1210	1375	1465	4050
United Kingdom	1221	1249	1112	3582
Total	8375	8716	8784	25875

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Table 1, depicts the total number of participating companies in the analysis. Noticeable, the Italian companies have the largest portion of the total data number with an intervention of 10375 companies, then it comes the United Kingdom with 3582 companies, France 3458, Germany 2875, Belgium 1535, and Spain with 4050 companies respectively. 8375 companies are observed in 2014, 8716 in 2013, and 8784 are observed in 2012.

Table 2: Total Number of Active companies per country year.

Total Number of Active companies per country year.				
Country/ Year	2014	2013	2012	Total
Belgium	434	519	549	1502
France	1140	1075	1062	3277
Germany	839	1000	1006	2845
Italy	3091	3245	3398	9734
Spain	1140	1288	1377	3805
United Kingdom	1210	1239	1102	3551
Totals	7854	8366	8494	24714

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Table2 shows the distribution of the Active observations across years and countries. The total active observation included in the sample is 24714 companies.

9734 out of 24714 (39.38%) are Italian active companies that belong to the manufacturing sector, 1502 out of 24714 (6.07%) are active Belgium companies, 3277 (13.25%) are French, 2845 (11.51%) German, 3805 (15.39%) are Spanish, and 2551 (14.36%) are English SMEs, Active and belong to the European Manufacturing sector. The sum of active observations per year are: 7854 in 2014, 8366 in 2013, and 8494 in 2012.

Table 3: Total number of Distressed companies per year and country.

Number of Distressed companies per year and Country.				
Country/ Year	2014	2013	2012	Total
Belgium	7	16	10	33
France	49	86	46	181
Germany	8	16	6	30
Italy	376	135	130	641
Spain	70	87	88	245
United Kingdom	11	10	10	31
Totals	521	350	290	1161

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Table 3 shows the distribution of the distressed observations across years and countries. The total distressed observation included in the sample is 1161 companies. 641 out of 1161 (55.2%) are Italian distressed (defaulted) companies that belong to the manufacturing sector, 33 out of 1161 (2.8%) are distressed Belgium companies, 181 (15.6%) are French, 30 (2.5%) German, 245 (21.10%) are Spanish, and 31 (2.67 %) are English SMEs, Distressed and belongs to the European Manufacturing sector. The sum of distressed observations per year are: 521 in 2014, 350 in 2013, and 290 in 2012.

Table 4: Total number of companies per country, Year and Group (Active "A", Distressed "D")

Total Number of companies Per Country, Group and Year.							
Year	2014		2013		2012		
Country/ Group (A or D)	A	D	A	D	A	D	Total
Belgium	434	7	519	16	549	10	1535
France	1140	49	1075	86	1062	46	3458
Germany	839	8	1000	16	1006	6	2875

Italy	3091	376	3245	135	3398	130	10375
Spain	1140	70	1288	87	1377	88	4050
United Kingdom	1210	11	1239	10	1102	10	3582
Total	8375		8716		8784		25875

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Table 4 shows the overall counting and statistics of the participating SMEs for the analysis regarding the year of observation, country of origin and the status of solvency. Although, the previous tables have shown precise details of the data statistics, Table 4 outlines the overall classification, counting and statistics in one table. As noticeable, Italy has the dominant observations number of both active and distressed companies among other countries and through the years precisely in the year of 2012. The variations between of the total numbers observed in each year are not large, although the number of distressed companies is not in balance with the number of active companies. Therefore, the weighting of the samples is applied to recover the unbalance.

5.2.3 Training and Testing Summary

5.2.3.1 Training Sample

As we mentioned in the introduction the obtain data would be split into two samples:

1. Training sample (the observations of 2012 and 2013)
2. The testing sample (the observations of 2014). Here we will start with discussion of the training sample.

Table 5: Training Sample the 2012 and 2013 years' data

Training Sample					
Year	2013		2012		
Country/ Group (A or D)	A	D	A	D	Total
Belgium	519	16	549	10	1094
France	1075	86	1062	46	2269
Germany	1000	16	1006	6	2028
Italy	3245	135	3398	130	6908
Spain	1288	87	1377	88	2840
United Kingdom	1239	10	1102	10	2361
Total	8716		8784		17500

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Table 5 shows the counting and statistics of the observations of the training sample that is going to be used in the models' development process. The financial ratios of the counted training sample companies are the independent variables and the predictors of each created and tested model of LR technique which will be discussed below.

The total number of training sample's companies is 17500 observed in two serial years (2012, 2013). Regarding the years' observations; 2013's companies are 8716 out of 17500, 8366 (96%) are active companies and 350 (4%) are distressed. 2012's companies present 8784 out of 17500, (96.69%) are active companies and (3.31%) are distressed. Belgium companies are 1094, (97.6%) active companies and (2.8%) are distressed. French companies are 2269, (94.20%) active and (5.80%) are distressed companies. German companies are 2028, (98.9%) active companies and (1.2%) are distressed. Italian companies are 6908, (96.10%) active and (3.90%) are distressed companies. Spanish companies are 2840, (93.8%) active and (6.2%) are distressed. The English companies are 2361, (99.10%) are active companies and (0.90%) are distressed.

In order to deal with the problem of class imbalance (different number of observations the two categories) a weighting process is implemented.

5.2.3.2 Validation Sample

The validation and testing sample is the set of data that is used to check the reliability of the created model (the model that has been created using the training sample). In this study, 2014's available data is the validation sample. Table 6 shows the number of companies that are included in the validation sample and it encapsulate the details regarding the data's country of origin and the group of solvency status. Regarding the year's observations; 2014's companies are 8375, 7854 (93.77%) of them are active companies and 521 (6.23%) are distressed.

Table 6: The Validation Sample

Validation Sample			
Year	2014		
Country/ Group (A or D)	Active (A)	Distressed (D)	Total
Belgium	434	7	441
France	1140	49	1189
Germany	839	8	847
Italy	3091	376	3467
Spain	1140	70	1210
United Kingdom	1210	11	1221
Total	8375		8375

Source: <https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis>

Belgium companies are 441, (98.4%) active companies and (1.6%) are distressed. French companies are 1189, (95.87%) active and (4.13%) are distressed companies. German companies are 847, (99%) active companies and (1%) are distressed. Italian companies are 3467, (89.15%) active and (10.85%) are distressed companies. Spanish companies are 1210, (94.2%) active and (5.8%) are distressed. The English companies are 1221, (99.10%) are active companies and (0.90%) are distressed.

5.3. Financial Ratios

As mentioned earlier, the financial ratios are an expression of the relationship between two items selected from the income statement or the balance sheet of a firm. Beaver, et al., 2005 state that the financial ratios are used to measure the relationship between two or more components of the financial statements and have greater meaning when the results are compared to industry standards for businesses of similar size and activity. According to the literature review and data availability, 12 ratios have been chosen for the analysis

5.3.1 Training sample's ratios statistics

As we mentioned before, a group of 12 financial ratios was selected to be calculated. Table 7 presents the selected ratios. Table 8 shows Calculated Ratios' averages for the Active (A) and the Distressed (D) firms. The next tables (Tables 9a - 9c) shows the total averages of the sample's calculated ratios per country of group of solvencies.

Table 7: The selected ratios

	Ratio	Equation Components
X ₁	Current Liquidity Ratio	Current assets / current liabilities
X ₂	Acid test	(Current assets-inventories) / current liabilities
X ₃	Liquidity Ratio	Cash / current liabilities
X ₄	Returned on Assets ROA	Net result / total assets
X ₅	Stock Turnover	COGS / inventories
X ₆	Collection Period	365 / account receivables turnover ratio
X ₇	Credit Period	365 /account payables turnover ratio
X ₈	Solvency ratio (Asset based)	(Net Income + Depreciation) / Total Assets)
X ₉	Earnings Before Interest, Taxes, Depreciation and Amortization Margin (EBITDA Margin)	EBITDA / Revenue
X ₁₀	Interest cover	EBIT / interest expenses
X ₁₁	Profit per employee	Net Revenue / Average Number of Employees

X ₁₂	Debt Ratio	(Long-term debt + Current Liabilities) / Total Assets
-----------------	------------	---

Source: Subramanyam K.R. (2014). "Financial Statement Analysis". 11th Edition McGraw-Hill

Table 8: Calculated Ratios' averages for the Active (A) and the Distressed (D) firms

Ratio		Total Average	A	D
X ₁	Current Liquidity (Current Ratio)	1.90	1.93	1.11
X ₂	Acid test	0.27	0.28	0.06
X ₃	Liquidity Ratio	1.34	1.37	0.72
X ₄	Returned on Assets ROA	3.63	4.13	-9.39
X ₅	Stock Turnover	9.94	9.98	9.04
X ₆	Collection Period	89.25	88.41	111.89
X ₇	Credit Period	58.11	56.95	88.85
X ₈	Solvency ratio (Asset based)	37.21	38.07	15.05
X ₉	Earnings Before Interest, Taxes, Depreciation and Amortization Margin (EBITDA)	6.86	7.23	-2.84
X ₁₀	Interest cover	11.07	11.61	-2.80
X ₁₁	Profit per employee	7.78	8.33	-6.03
X ₁₂	Debt Ratio	0.57	0.56	0.85

Source: Author's Calculation

Table 9a: Total ratios' averages of the sample's ratios per country and group.

County/ Ratios	Current Liquidity		Acid test		Liquidity ratio		ROA using P/L before tax (%)		Stock turnover (x)	
	A	D	A	D	A	D	A	D	A	D
Group										
Belgium	2.07	1.02	0.37	0.10	1.48	0.66	4.35	-13.64	11.43	11.28
France	1.87	1.27	0.31	0.09	1.31	0.79	5.35	-9.87	10.66	9.53
Germany	2.89	1.76	0.49	0.09	1.90	1.11	5.33	-4.35	9.50	9.64
Italy	1.66	1.00	0.21	0.05	1.19	0.67	3.10	-9.59	8.42	8.02
Spain	1.89	1.11	0.20	0.04	1.37	0.69	3.16	-10.22	10.92	9.36
United Kingdom	1.91	1.06	0.34	0.10	1.42	0.71	5.94	-7.46	12.45	12.85

Source: Author's Calculation

Table 9b: Total ratios' averages of the sample's ratios per country and group.

County/ Ratio	Collection period (days)		Credit period (days)		Solvency ratio (Asset based) (%)		EBITDA margin (%)		Interest cover (x)	
	A	D	A	D	A	D	A	D	A	D
Belgium	69.84	77.77	48.07	71.73	41.42	10.26	7.52	-5.02	12.20	-6.53
France	70.24	78.13	49.99	74.39	42.48	19.27	6.32	-3.93	15.37	-5.70
Germany	36.32	37.70	19.87	39.00	37.19	25.47	7.31	1.97	11.05	-1.71
Italy	114.65	135.04	81.17	127.95	33.89	9.05	7.31	-2.82	10.15	-3.17
Spain	106.37	127.01	48.79	68.95	43.49	16.29	7.33	-4.08	9.71	-3.96
United Kingdom	62.54	60.81	38.33	49.24	39.09	16.11	7.56	-1.43	14.98	2.70

Source: Author's Calculation

Table 9c: Total ratios' averages of the sample's ratios per country and group.

Country / Ratio	Profit per employee (th EUR)		Debt ratio	
	A	D	A	D
Belgium	9.15	-13.77	0.55	0.85
France	9.43	-8.83	0.53	0.77
Germany	8.72	-2.90	0.49	0.70
Italy	7.68	-7.45	0.58	0.90
Spain	7.61	-10.10	0.55	0.87
United Kingdom	9.30	-4.77	0.57	0.85

Source: Author's Calculation

6. Application, Analysis and Comparison

6.1 Selection of Independent Variables

The selection of the independent variables (ratios) to be included in the prediction model is a very difficult procedure. There is a wide range of failure models with good classification results, each consisting of different variables and a different number of variables (Daubie, et, al 2002).

The most common strategy for selecting model predictors used in the majority of research studies is based on statistical procedures. Since there is no financial theory indicating the financial ratios that are the best predictors, researchers select those variables that satisfy some distributional requirements (Berger, et al, 2005).

A number of methods have been proposed attempting to relate the importance of individual ratios (Eisenbeis 1977).

In our initial set of the 12 financial ratios (Table 10) derived from the financial statements collected, we apply the test of Kruskal–Wallis in order to overcome multicollinearity problems, reduce the dimensionality and increase the applicability of the model.

Table 10: Kruskal Wallis Test for training sample

Ratio		Chi- Square	Asymptotic Significance
X ₁	Current Ratio	5693.953	0.000
X ₂	Acid test	3151.173	0.000
X ₃	Liquidity Ratio	5420.523	0.000
X ₄	Returned on Assets ROA	8966.232	0.000
X ₅	Stock Turnover	473.170	0.000
X ₆	Collection Period	941.896	0.000
X ₇	Credit Period	2869.599	0.000
X ₈	Solvency ratio (Asset based)	6874.394	0.000
X ₉	(EBITDA) Margin	6938.946	0.000
X ₁₀	Interest cover	7456.665	0.000
X ₁₁	Profit per employee	6088.308	0.000
X ₁₂	Debt Ratio	7421.275	0.000

Source: Author's Calculation

According to the Kruskal-Wallis test, all the ratios (12 out of the 12) were found statistically significant at a level of 5%.

6.2 Developing the Logistic Regression Model

Following the step of testing the variables using Kruskal Wallis Test which resulted in twelve variables to be chosen as predictors in the analysis, we applied the logistic regression model using IBM SPSS Statistics 23, and the results were the following:

1. The Logistic Regression model at 5% significance level.

The application of the LR model using the 12-selected predictors at 5% significance level resulted in an 8-variables equation as shown the Table 6.2. The Variables in the equation are ROA (X₄), EBITDA Margin (X₉), Interest Cover (X₁₀), Collection Period (X₆), Current Ratio (X₁), Solvency Ratio (X₈), Debt Ratio (X₁₂), and Profit Per Employee (X₁₁).

Table 11: Variable in the equation at 5% significance level

Variables in the Equation Set ^a							
		B	S.E.	Wald	df	Sig.	Exp(B)
	ROA	.033***	.004	61.588	1	.000	1.033
	EBITDA Margin	.048***	.004	150.483	1	.000	1.049
	Interest Cover	.023***	.002	87.649	1	.000	1.023
	Collection Period	-.005***	.000	209.503	1	.000	.995
	Current Ratio	.086**	.029	8.771	1	.013	1.090
	Solvency Ratio	.023***	.001	298.265	1	.000	1.023
	Debt Ratio	- 3.696***	.095	1510.87	1	.000	.025
	Profit Per Employee	.014***	.003	31.579	1	.000	1.015
	Constant	2.353***	.094	626.247	1	.000	10.518

Source: Author's Calculation

Note: ** and *** represent 5% and 1% significance level respectively

The classification of the 8-variables equation is shown in Table 12 at 5% significance level with a proper sign.

Table 12: Logistic Regression - Classification Table set^a (5%, 8 variables equation)

Logistic Regression - Classification Table set ^a						
	Predicted					
	Training Sample			Validation Sample		
	Status		Correct %	Status		Correct %
	Distressed	Active		Distressed	Active	
	Distressed	7246	1504	82.8	348	75
Active	1644	7093	81.2	1587	5863	78.7
Overall Percentage	82%			78.9 %		

Source: Author's Calculation

For Set^a LR model, the overall percent of correct classification is 82% for the training sample and 78.9% for the validating sample. The model reaches its highest discrimination accuracy for the active firms of the validating sample with 82.3% of correctly classified.

2. The Logistic Regression model at 1% significance level.

The application of the LR model using the 12-selected predictors at 1% significance level resulted in a 7-variables equation with no constant as shown the Table 6.4. The Variables in the equation are ROA (X4), EBITDA Margin (X9), Interest Cover (X10), Collection Period (X6), Current Ratio (X1), Solvency Ratio (X8), Debt Ratio (X12), Profit Per Employee (X11).

Table 13: Variable in the equation at 1% significance level Set^b

Variables in the Equation Set ^b						
	B	S.E.	Wald	df	Sig.	Exp (B)
ROA	.043***	.004	124.134	1	.000	1.044
EBITDA Margin	.041***	.004	133.944	1	.000	1.041
Interest Cover	.027***	.002	133.525	1	.000	1.027
Collection Period	-.009***	.000	1021.465	1	.000	.991
Current Ratio	.101***	.024	17.472	1	.000	1.106
Solvency Ratio Assets based	.028***	.001	583.677	1	.000	1.028
Profit Per Employee	.012***	.003	21.443	1	.000	1.012

Source: Author's Calculation

Note: *** represent 1% significance level respectively

The classification of the 7-variables LR equation is shown in Table 14 at 1% significance.

Table 14: Logistic Regression - Classification Table set^b (1%, 7 variables equation)

Logistic Regression - Classification Table set ^b						
	Predicted					
	Training Sample			Validation Sample		
	Status		Correct %	Status		Correct %
	Distressed	Active		Distressed	Active	
Distressed	6713	2037	76.7	339	84	80.1
Active	1845	6892	78.9	1479	5971	80.1
Overall Percentage			77.8%			80.1%

Source: Author's Calculation

For Set^b LR model, the overall percent of correct classification is 77.8% for the training sample and 80.1% for the validating sample. The model reaches its highest

discrimination accuracy for the active firms of the validating sample with 80.1% of correctly classified.

As we can see, there are differences in the percentages of correct classification between the two sets (Set^a, Set^b) of equations' variables. The difference occurred because of the appetite of increasing the confidence level.

6.2. Model Results

In this part we would assess and compare the overall usability and predictability of each model regarding our case and circumstances. Tables 15 and 16 depict the overall result and test of each approached model.

The comparison is done using three comparable results:

1. The results of overall correct percentage. The higher the value, the higher the model's predictability.
2. Area Under the Curve (AUC) or it is also known as the operating characteristic curve (ROC Curve) test results. AUC curve tests the models' accuracy of separating the tested groups (Active, Distressed). The accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of 0.5 represents a failed test (Myerson, 2001).
3. Kolmogorov-Smirnov goodness of fit test (One sample K-S Test). K-S Test is a test used to decide if a sample comes from a population with a specific distribution. (Drew, et al, 2008). In our study the K-S Test is applied as a distribution normality test.

Table 15: AUC Results

	LR			
	Training Sample		Validation sample	
	Set ^a	Set ^b	Set ^a	Set ^b
Overall correct %	82%	77.80%	78.90%	80.10%
Average	80%		79.55%	
AUC - ROC	75.3%	71.5%	73.2%	74.3%

Source: Author's Calculation

Table 16. K-S Results

K-S Test of Different Predictors sets (Kolmogorov-Smirnov goodness of fit)		
	LR	
	Set ^a	Set ^b
Test statistic	15.2%	11.7%
Asymp. Sig (2-tailed) Lilliefors Significance Correction	0.000	0.000

Source: Author's Calculation

The above mentioned results show that the model presents accuracy and predictability under both different schemes, although, a considerable imbalanced data set with a small number of defaults have been faced through the modeling process. The averages of the LR model were 80% and 78.90%, which are similar to previous studies.

Assessing the overall significance, effectiveness, efficiency of the two models, two non-parametric tests have been implemented, Area Under the Curve (AUC) or it is also known as the operating characteristic curve (ROC Curve), and Kolmogorov-Smirnov (K-S) goodness of fit test.

The AUC values indicate predictability performance since they have a value that range from 71.5% to 75.3% throughout the samples and the predictors sets.

Kolmogorov-Smirnov goodness of fit test can give an answer if a sample comes from a population with a specific distribution. K-S test can also be helpful in distinguishing two different categories in dual problems (for example defaulted / non-defaulted firm). According to Conover, 1999 K-S test is used to check the normality assumption in Analysis of Variance.

K-S Test of the data sets implies that the distribution of model (is normal. The test statistics predicts goodness of fit for the LR model, with K-S Test values (15.2%, 11.7%).

7 Conclusion – Further Research

In a modern era, there are surrounding threats and factors that affect and shape the business strategies. Corporate risk management is one of them. The discussion of the business environment implies how hard and demanding it is for the modern enterprises to survive the fast-changing business climate and its changes that are driven by many diversified aspects and factors. The changing factors of the business environment cause some severe financial or nonfinancial losses and risks. The evaluation and prediction of credit risk is of utmost importance. Multicriteria decision making approaches and the statistical models are used as helpful tools of the corporate credit rating. In our study, we attempted to evaluate credit risk with a popular technique, namely Logistic Regression. Our sample consisted of manufacturing firms of different EU countries. The results depicted, that under two different schemes (difference significance levels and different variables) the model managed to predict credit risk in an accurate way (round 80% accuracy levels). The AUC (ROC) and Kolmogorov-Smirnov goodness of fit (K-S) tests, were applied to the comparison of the models' predictability and their results were quite comparable to the ones found in other similar studies. One advantage of our study, is the ability of generating a model applicable not only for a country but for set of countries with different economic conditions.

According to the limitation of study, this research has examined one methodology and one sector (manufacturing), for a specific time period of three years. In a further study, these issues could be considered for elaboration. Moreover, different

ratios and the involvement of qualitative factors could be considered for more meaningful and robust results.

References

- [1] Abdou, H. A., & Pointon, J. (2011). "Credit scoring, statistical techniques and evaluation criteria: A review of the literature". *Intelligent Systems in Accounting, Finance and Management*, 18(2-3), pp. 59-88.
- [2] Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., & Franklin, M. (2012). "Challenges and Opportunities with Big Data: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association." Divyakant Agrawal, UC Santa Barbara: Computing Research Association
- [3] Altman, E., Giancarlo, M., & Varetto, F. (1994). "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)". *Journal of Banking and Finance*.
- [4] Atiya, A. F. (2001). "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results". *IEEE Transactions On Neural Networks*.
- [5] Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). "Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy". *Review of Accounting studies*, 10(1), pp. 93-122.
- [6] Berger, A. N., Scott, F. W., & Miller, N. H. (2005). "Credit Scoring and the Availability, Price, and Risk of Small Business Credit". *Journal of Money*, pp. 191-222
- [7] Boguslauskas, V., & Mileris, R. (2009). "Estimation of Credit Risk by Artificial Neural Networks Models". *Economics of Engineering Decisions*, pp. 1392 - 2785.
- [8] Bohn, J. R., & Stein, R. M. (2009). "Active Credit Portfolio Management in Practice The Framework: Definitions and Concepts pp 1-43" First published:29 November 2011 Print ISBN:9780470080184 Willey Online Library
- [9] Bokpin, G. A. (2010). "Financial market development and corporate financing: evidence from emerging market economies". *Journal of Economic Studies*, pp. 96 - 116.
- [10] Boreiko, D., Kaniovski, Y., & Pflug, G. (December 2016). "Modeling dependent credit rating transitions". *Central European Journal of Operations Research*, pp. 989–1007.
- [11] Bulak, M. E., & Turkyilmaz, A. (2014). "Performance assessment of manufacturing SMEs: a frontier approach". *Industrial Management & Data Systems*, 114(5), pp 797-816.

- [12] Conover, W. J. (1999). "Statistics of the Kolmogorov-Smirnov type." *Practical nonparametric statistics*, pp. 428-473.
- [13] Daubie, M., Levecq, P., & Meskens, N. (2002). "A comparison of the rough sets and recursive partitioning induction approaches: An application to commercial loans". *International Transactions in Operational Research*, 9(5), pp. 681-694.
- [14] Drew, J H., Diane, E. L., & Leemis, I. M. (2008). "The Distribution of the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling Test Statistics for Exponential Populations with Estimated Parameters". *Computational Probability Applications* pp. 165-190. Part of the *International Series in Operations Research & Management Science* book series (ISOR, volume 247)
- [15] Eisenbeis, R. A. (1977). "Pitfalls in the application of discriminant analysis in business, finance, and economics". *The Journal of Finance*, 32(3), pp 875-900.
- [16] EUROPA. (2017). "The EU in brief, institutions and bodies, countries, symbols, history, facts and figures". The European Union.
- [17] Gepp, A., & Kumar, K. (2012). "Business failure prediction using statistical techniques: A review". Bond Business School Publications.
- [18] Gestel, T. V., & Baesens, B. (2009). "Credit Risk Management - Basic Concepts". New York: Oxford University Press Inc.,
- [19] Giudici, P. (2003). "Applied Data Mining: Statistical Method for Business and Industry". John Wiley & Sons.
- [20] Handzic, M., Tjandrawibawa, F., & Yeo, J. (2003). "How Neural Networks Can Help Officers to Make Better Informed Applications". *Informing Science*, pp. 1-13.
- [21] Hotchkiss, E., & Altman, E. I. (2006). "Corporate Financial Distress and Bankruptcy. Hoboken", New Jersey John Wiley & Sons, Inc.
- [22] Iazzolino, G., & Laise, D. (2012). "Business multicriteria performance analysis: a tutorial. Benchmarking: An International Journal, Vol. 19 Issue: 3, pp. 395-411, <https://doi.org/10.1108/14635771211243012>.
- [23] Memić, D. (2015). "Assessing Credit Default Using Logistic Regression and Multiple Discriminant Analysis: Empirical Evidence From Bosnia And Herzegovina". *Interdisciplinary Description of Complex Systems*, pp. 128-153.
- [24] Min, J., & Lee, Y.-C. (2005). "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters". *Expert Systems with Applications* 28 (2005) pp. 603–614.
- [25] Mouna, A., & Anis, J. (2015). "Stock Market, Interest Rate and Exchange Rate Risk Effects on non-Financial Stock Returns During the Financial Crisis". *Journal of the Knowledge Economy* September 2017, Volume 8, Issue 3, pp 898–915 Springer.
- [26] Myerson, J. (2001). "Area Under the Curve as a measure of Discount". *Journal of the Experimental Analysis of Behavior*, pp. 235-243.

- [27] NACE Rev2, May 2017. Available at <http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>
- [28] Rottig, D. (2006). "Institutions and emerging markets: effects and implications for multinational corporations". *International Journal of Emerging Markets*. Vol. 11 Issue: 1, pp.2-17,
- [29] Salame, E. J. (2007). "Data Mining Application in the Decision Making Process: the case of credit risk assessment". Centre International de Hautes Etudes Agronomiques Mediterraneennes, Chania (Greece). Institut Agronomique Mediterraneeen Chania, Crece.
- [30] Saunders, A., & Allen, L. (2002). "Credit Risk Measurement New Approaches to Value at Risk and other Paradigms". Book published by John Wiley & Sons, Inc
- [31] Scapolo, F., Geyer, A., Boden, M., Dory, T., & Ducatel, K. (2003). "The Future of Manufacturing in Europe 2015-2020 The Challenge for Sustainability". Institute for Prospective Technological Studies. European Commission Joint Research Center.
- [32] Siddiqi, N. (2006). "Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring". Canada: John Wiley & Sons, Inc., Hoboken, New Jersey.
- [33] Subramanyam K.R. (2014). "Financial Statement Analysis". 11th Edition McGraw-Hill
- [34] Sweeney, M. T., Szwajczewski, M., & Cousens, A. (2016). "The strategic management of manufacturing networks". *Journal of Manufacturing Technology Management*, pp. pp. 124-149.
- [35] Thomas, L., Edelman, D., & Crook, J. (2002). "Credit Scoring and Its Applications". University City Science Center, Philadelphia: The Society for Industrial and Applied Mathematics.
- [36] Vargo, S. L. (2011). "Market systems, stakeholders and value propositions: Toward a service-dominant logic-based theory of the market". *European Journal of Marketing*, pp. 217-222.
- [37] Verbeke, A. (2005). "Financial and Non-Financial Theories of International Diversification". In J. Cantwell, *Financial and Non-Financial Theories of International Diversification* (pp. 59 - 73). Book published by Emerald Group Publishing Limited.
- [38] Wehrspohn, U. (2002). "Credit Risk Evaluation Modeling - Analysis - Management". PhD. Discussion. Germany: Heidelberg University.
- [39] Wickens, M. (2016). "The Eurozone financial crisis: debt, credit ratings and monetary and fiscal policy". *Empirica Springer Link*, May 2016, Volume 43, Issue 2, pp 219–233.
- [40] Wiginton, J. C. (1980). "A note on the comparison of logit and discriminant models of consumer credit behavior". *Journal of Financial and Quantitative Analysis*, 15(3), pp. 757-770.

- [41] Wilson, A., Jayawarna, D., & Macpherson, A., (2006). "Managers' perceptions of management development needs in manufacturing SMEs". *Education + Training Journal*, 48(8/9), pp. 666-681.
- [42] Wu, S., Huang, Z., Chen, H., Hsu, C.-J., & Chen, W.-H.(2004). "Credit rating analysis with support vector machines and neural networks: a market comparative study". *Decision Support Systems*, Volume 37, Issue 4, September 2004, Pages 543-558.
- [43] Zopounidis, C., & Dimitras, A. I. (1998). "Multicriteria Decision Aid Methods for the prediction of business Failure". Netherlands: Kluwer Academic Publisher.