

Credit Scoring and Default Risk Prediction: A Comparative Study between Discriminant Analysis & Logistic Regression

Zaghdoudi Khemais¹, Djebali Nesrine² & Mezni Mohamed³

¹ Faculty of Law, Economics and Management of Jendouba, University of Jendouba, Tunisia

Correspondence: Djebali Nesrine, Faculty of Law, Economics and Management of Jendouba, University of Jendouba, Tunisia. Tel: 216-22-961-769. E-mail: djbeli.nesrine@gmail.com

Received: January 16, 2016

Accepted: February 6, 2016

Online Published: March 25, 2016

doi:10.5539/ijef.v8n4p39

URL: <http://dx.doi.org/10.5539/ijef.v8n4p39>

Abstract

This paper aims to develop models for foreseeing default risk of small and medium enterprises (SMEs) for one Tunisian commercial bank using two different methodologies (logistic regression and discriminant analysis). We used a database that consists of 195 credit files granted to Tunisian SMEs which are divided into five sectors “industry, agriculture, tourism, trade and services” for a period from 2012 to 2014. The empirical results that we found support the idea that these two scoring techniques have a statistically significant power in predicting default risk of enterprises. Logistic discrimination classifies enterprises correctly in their original groups with a rate of 76.7% against 76.4% in case of linear discrimination giving so a slight superiority to the first method.

Keywords: credit scoring, probability of default, discriminant analysis, logistic regression, SMEs

1. Introduction

In most developing countries, the banking sector continues to play a leading role in promoting economic growth, and this in spite of efforts to boost their financial markets, which remained underdeveloped. Credit risk is among the most significant risks to which banks are exposed. This allowed the proliferation of risk management methods.

For banks, the big problem has become not only to decide whether to grant a loan or not but also to predict the probability of default of a borrower in case of a credit agreement. It is to anticipate, for a determinate period, the quality of the borrower (good or bad borrower) and its ability to repay its debt. When credit is granted, and regardless of the expected gain, it can be exposed to uncertainty of default. The credit institution is not always safe to recover its funds and is exposed to counter party risk.

The objective of this paper is to present models for predicting the probability of default of the counterparty based on the score method, which pits discriminant analysis against logistic regression. Credit scoring is a method that helps the bank to rationalize its process for credit granting decision. Its principle is to synthesize a set of financial ratios as one indicator able to distinguish between good and bad customers.

To reach this goal, we collected a data from 195 credit files granted to Tunisian SMEs which represent the principal sectors of the economy for a period from 2012 to 2014. In Tunisia, SMEs are the most dominant form of companies. They represent almost more than 80% of the total of companies. Moreover, SMEs are the engine of growth of the economy and the main source of job creation.

The remainder of the paper is organized as follows. In section 2 we present a literature review on credit scoring and default risk foresight. In section 3, we describe our research methodology and credit scoring model. Empirical investigation and results are given respectively in sections 4 and 5. Finally, section 6 is devoted to discussion.

2. Literature Review

Credit scoring is a statistical method for estimating the probability of default of the borrower using historical data and statistical data to reach a single indicator that can distinguish good borrowers from bad borrowers.

The score function is based on a method of financial analysis on financial ratios presented as a single indicator can distinguish between healthy and failing companies.

In the theoretical part of the paper we expand in more detail the two types of credit scoring models, namely

linear discriminant analysis and Logit regression.

Mester (1997) defines credit score following, “credit scoring is a statistical method for estimating the probability of default of the borrower using historical data and statistical data to assess the credit risk of a loan applicant that process quantitative information of an individual or a business that the bank can use to classify borrowers”.

In the same context, (Wallis, 2000) defines credit scoring as a method for estimating the interest, not the risk of a credit. When to (Thomas et al., 2002), Credit Scoring is a set of decision models and underlying techniques that help in the decision granting consumer credits.

Credit scoring systems can be found in many types of credit analysis from consumer loans to commercial loans where it is first to identify factors that determine the probability of default and weigh their weight in a quantitative score second (Saunders & Allen, 2002).

Berger et al. (2002), Frame et al. (2001), have focused on a new method of analyzing the hard information is the “Score” which facilitates the task of decision-making and even the accelerate hence when to them, the use of hard information is not new in concept but the novelty is reflected in the use of a new way of corporate risk calculation while based on the score.

The Credit Scoring uses quantitative measures of performance and characteristics of past loans to predict future performance of loans with similar characteristics. The Credit Scoring does not approve, or reject a loan application; it can predict the rather poor performance of probability of occurrence (default) as defined by the lender (Cairo Kossmann, 2003).

Young, Hunter, and Udell (2004), consider that the bank uses the scores that represent the risk of borrowers for its decision to credit decision.

In the same context, (Berger & Frame, 2005), interpreted that the use of a model “small business credit scoring” is essentially based on two categories of variables, variables relating to the contractor as “its income, default, past bankruptcies ... etc.” and variables related to the company said additional variables that are in the form of financial ratios.

When to (Anderson, 2007), credit scoring is defined as the use of statistical models to transform the “qualitative and quantitative” data in the risk management of bank credit, credit scoring is then defined as the use statistical models to transform data (qualitative, quantitative) into measurable digital indicators for help to the decision to grant or credit rejection.

Tuffery (2007) develops new credit scoring prediction methodologies such as “neural networks, genetic algorithms by credit scoring ... etc.” From which he concluded that there are five types of score in literature namely” the appetite score, the risk score, the score granted, recovery of the attrition score and score. “

Empirically, this new technique was born for the first time in the United States, and that developed subsequently in the rest of the country, allows synthesizing the credit risk also said the risk of some or against default.

Continuing the pioneering work of (Fisher, 1936) on the discriminant analysis, (Durand, 1941) was the first to recognize the possibility of using statistical techniques to discriminate between good and bad borrowers, then, Altman (1968) significantly contributed to the development, promotion and better understanding of the Credit Scoring and his techniques.

Altman (1968) was the first to use more than a ratio that lead to a score for a sample of 22 financial ratios of a period from 1946 to 1965, only 5 financial ratios have tried summers with the most discriminating good filing rate 95% a year before failure, 72% two years before failure and 48%, 29% and 36% three years, four years and five years before failure.

Deakin (1972) used a set of 14 financial ratios for a period of five years; he concludes that the ratios of “profitability, liquidity and solvency” are the most discriminating ratios with significant power.

In the same context, (Canon & Holder, 1979) have used a set of 31 financial ratios for a sample of 190 industrial SMEs which 95 companies are failing firms and 95 firms are sound. Only 5 financial ratios were registered as those with the highest discriminatory power with a percentage of correct classification of companies in their original group of 67% one year prior to default and 69%, 75% from two and three years before failure.

Using logistic regression, (Ohlson 1980), used a set of 9 financial ratios for 363 companies during the period of 1970 up 1976. The study revealed the positive impact of 4 variables that predict bankruptcy, namely, “company’s size, liquidity, performance and financial structure of the company.”

3. Research Methodology

3.1 Credit Scoring Model

The scoring model aims to determine the impact of characteristics of borrowers on their luck to default. It produces scores which are probabilities measuring credit risk of potential borrowers.

From a set of “N” companies divided into two samples, one of failing companies and the other healthy companies are measured “K” ratios that lead to a variable “Z”, and catch values by this variable must be as various possible a subset to another.

In order to calculate the probability of default, we will calculate the score in the first place based on our ratios that have been selected as the most discriminating.

The score is equal to:

$$Z = \sum_{i=0}^n a_i x_i \quad (1)$$

– The coefficients a_i ... represents the weights associated with accounting and financial ratios (X_i).

When the score is high, it presents a satisfactory situation, and otherwise (low score), there is a risk of failure.

The scoring function has a resolving power between healthy and failing company, but that power is not absolute as there is an overlap between the two subsets.

The variable “Z” will therefore include ratios that we ensure the statistical independence.

The probability of default noted DP is presented as follows:

$$DP = 1 / (1 + e^{-Z}) \quad (2)$$

Table 1. Example of borrower rating scale

Default Probability	Risk level	Bank's Decision
0-0.05	Very low	Accepted
0.05-0.15	Low	Accepted
0.15-0.20	Average	Accepted
0.20-0.25	High	Accepted
0.25-0.35	Very high	Accepted / Rejected
0.35-0.50	Too high	Rejected
0.50-0.60	Too high	Rejected
0.60-1	Default	Rejected

3.2 Data Description

The main goal of our model is to allocate credit borrowers to two groups: “good credit” group that is likely to repay the financial obligation or “bad credit” group with a high risk of default on its financial obligations. Our sample consists of 195 SMEs using credit files distributed over five sectors of activity namely “agriculture, industry, services, tourism and trade.”

For the first time, in this study employ a sample consisting of over 195 small and medium companies failed during the period 2012-2014.

Given the number of 195 SMEs, our sample is divided into two sub-samples which are distributed as follows:

- Companies failing: have dropped the bankrupt companies are those having registered a payment delay in repayment of their loans (3 months or more);
- Healthy companies: healthy companies who repaid their entire loan by the deadline and did not have late payment problems.

Table 2. Distribution of healthy and failing companies by sector

Activity		F_S		Total
		Failing	Healthy	
Activity	TRADE	15	25	40
	AGRICULTURE	16	11	27
	INDUSTRY	13	19	32
	SERVICE	3	13	16
	Tourism	51	29	80
Total		98	97	195

To estimate the bank debt, we will be able to formulate the hypotheses to be tested on Tunisian SMEs while providing a set of variables assumed as those having an impact on bank debt.

The credit risk analysis is based on various information about the borrower, which may be summarized as qualitative and quantitative data such as financial ratios, since access on qualitative data seems be difficult, our study will be only based on quantitative variables; hence it becomes necessary to make a wise choice of our financial ratios

The table below allows us to define the variables in our model, where each variable introduced in our model provides us with information.

Table 3. Description of the data

	N	Minimum	Maximum	Average	Standard deviation
RATIO 1	195	-81,6680	318,2960	24,101179	36,0290364
RATIO 2	194	-134,6350	1340,1660	40,558335	104,0103809
RATIO 3	195	-479,0810	978,8270	34,727503	98,1647708
RATIO 4	195	-528,8560	198,3670	15,743287	73,9955638
RATIO 5	195	-6,9850	847,2240	8,231113	60,7867314
RATIO 6	195	,0000	13458,5230	338,564169	1,2850487E3
RATIO 7	195	-21,1560	1306,8780	75,837067	139,8657080
RATIO 8	195	-13,0730	79446,9630	1926,087856	9,9830647E3
RATIO 9	195	-133,9840	167,3790	13,174441	26,6560542
RATIO 10	195	-172,8990	91,8890	6,726282	25,7896096
RATIO 11	195	-24,5660	143,3560	10,218651	18,0394706
RATIO 12	195	-79,2450	81,1890	10,376646	12,8670865
RATIO 13	195	-80,4610	75,1390	7,158672	12,6003230
RATIO 14	195	-1530,8140	1533,1740	51,554390	213,9287882
RATIO 15	195	-2191,2460	1556,6590	17,631949	240,1925506
RATIO 16	195	,0000	145,8330	1,869000	12,7558570
RATIO 17	195	-17760,2770	6018,6560	-91,841933	1,8537492E3
RATIO 18	195	-327,8080	436,5820	27,000579	71,9813496
RATIO 19	195	-342,0140	100,3740	-10,126077	63,1765403
RATIO 20	195	-617,5070	239,1270	12,175082	88,4597086
RATIO 21	195	-419,6280	1122,5560	58,136667	144,3645367
RATIO 22	195	,0000	1175,7220	152,755056	165,9894275
RATIO 23	195	,0000	938,1170	85,462318	128,5738379
RATIO 24	195	-214,7250	323,7510	44,890426	47,1501110
RATIO 25	195	,0000	94,8380	6,115333	12,2088273
RATIO 26	195	-16027,8750	9255,6850	-30,669574	1,3288207E3
RATIO 27	195	-23,0590	157,9280	4,532723	18,8176358
RATIO 28	194	-4716,8680	850593,2980	6071,297732	6,5631145E4

The descriptive statistics table presents the average of the different variables for different classes of the dependent variable; it allows us to show that the values of our twenty eight selected ratios are dispersed.

It is then observed that the values differ widely from one company to another, hence for example R1 (Ratio of value added), the values are spread out over an interval having a minimum of -81, 668 and maximum .296 318.

The wider dispersion was recorded at R28 (Net investment in fixed assets) of a minimum value of -4716.863 and

a maximum value at approximately 850,593.298. The gap between the values allows us to explain the failure phenomenon; hence such a difference in the ratios allows companies to classify healthy and failing firms.

4. Empirical Investigation

4.1 Discriminant Analysis

Discriminant Analysis (FDA) is a classical method developed by (Fisher, 1936) which can be explanatory as a predictive. This method allows to study the membership of a borrower group, such as «failing or healthy companies» and identify the characteristics of groups of borrowers based on the explanatory variables.

The discriminant function of our model is presented as follows:

$$Z = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_N X_N \quad (3)$$

Where:

- a_1, a_2, \dots, a_N : The discriminant coefficients.
- X_1, X_2, \dots, X_N : The discriminating variables.

The score “Z” allows us to synthesize the characteristics of the company with regard to the variable to explain this value allows us to predict the probability of default of each company.

Since Z is a failure score we have:

If:

- Z score is high => The default risk is very low;
- Z-score is very low => The default risk is very high.

However,

- A healthy companies is rated by “1”;
- A failing companies is rated by “0”.

To obtain an estimate of our model by a discriminant analysis, we mainly referring to an estimate through the software “SPSS.18”.

From Analysis menu, we choose the procedure classification and discriminant analysis. Then we will define the qualitative criteria of healthy and failing firms by defining intervals with a minimum value equal to “0” for healthy businesses and a maximum value equal to “1” for failing companies. Once our summers have defined intervals, we will select our financial ratios as explanatory variables while choosing the step by step method.

However, we choose the means of Anova tests and Box as features and lambda Wilks' summary of steps “as the method and” Summary + Classification by elimination “as a classification.

4.1.1 Test of Equality of Group Means

To test the hypothesis of equality of mean vectors of the different classes “healthy firms and failing firms” we will proceed to the equality of means test groups presented in the table below.

The purpose of this test is to discriminate the ratios in order to select the variables that are considered the most discriminating between the groups from which the top three test beings are taken into account namely:

4.1.1.1 Fisher Test

The Fisher test can indicate the existence or non-existence of a significant difference. More F statistic is higher; more the discriminating power of the variable is significant.

4.1.1.2 Wilks's Lambda Test

This test can prove the existence or non-existence of relationship between the dependent variable and the explanatory variables. The Wilks's Lambda test value is always between (0 and 1), a value of 1 means that the middle classes are equal and the most discriminating variable is one that has a lambda value and significance level close to 0.

A low value of Lambda is interpreted as low intra-group variation and therefore strong intergroup variations, hence a significant difference in the mean of classes.

Wilks's Lambda is calculated as follows:

$$\lambda = 1 - (\text{Inter-group variance} / \text{Total Variance}) \quad (4)$$

Table 4. Tests of equality of group means

	Wilks's Lambda	F	df1	df2	Signification
RATIO 1	.974	5.027	1	191	.026
RATIO 2	.981	3.601	1	191	.059
RATIO 3	.994	1.060	1	191	.304
RATIO 4	1.000	.005	1	191	.942
RATIO 5	.997	.563	1	191	.454
RATIO 6	.999	.221	1	191	.639
RATIO 7	.993	1.336	1	191	.249
RATIO 8	.977	4.565	1	191	.034
RATIO 9	.973	5.206	1	191	.024
RATIO 10	.983	3.285	1	191	.071
RATIO 11	.990	1.918	1	191	.168
RATIO 12	.972	5.404	1	191	.021
RATIO 13	.961	7.754	1	191	.006
RATIO 14	.998	.318	1	191	.574
RATIO 15	.985	2.965	1	191	.087
RATIO 16	.984	3.182	1	191	.076
RATIO 17	1.000	.058	1	191	.810
RATIO 18	.916	17.534	1	191	.000
RATIO 19	.980	3.946	1	191	.048
RATIO 20	.873	27.673	1	191	.000
RATIO 21	.998	.415	1	191	.520
RATIO 22	.905	20.089	1	191	.000
RATIO 23	.958	8.333	1	191	.004
RATIO 24	.984	3.091	1	191	.080
RATIO 25	.988	2.231	1	191	.137
RATIO 26	.990	1.901	1	191	.170
RATIO 27	.999	.187	1	191	.666
RATIO 28	.991	1.655	1	191	.200

According to Wilks lambda value, only obtained variables are: settlement period of supplier credit, relative value of working capital, general liquidity.

In addition, the Wilks lambda value is equal to 1 for R4 and R17, from which one can conclude that the average classes are equal, with:

- R4 : Bank's Part;
- R17: Dividend Distribution Capacity.

The lower value of Wilks Lambda was recorded at R20 "Working capital" with a 0873 value.

4.1.1.3 Signification

The test of the significance of each variable allows us to choose the variables:

R1: Ratio of the value added, R8: supplier credit settlement period, R9: Gross margin on revenues, R12: Gross Profitability of total assets, R13: Net Profitability of total assets, R18: Excess of insufficient capital, R19: Excess on insufficient current liabilities, R20: Working capital, R22: Liquidity in the broad sense, R23: Liquidity in the strict sense.

4.1.2 Model Tests

4.1.2.1 Box's Test of Equality of Covariance's Matrices

Among the fundamental tests of discriminant analysis, we can cite the Box test of equality of covariance matrices of the groups. This test checks the null hypothesis of equal variance-covariance matrix, i.e. to verify that the variance-covariance matrix is identical from one indicator to another.

We can say that the covariance matrices for the different classes are not equal. To confirm this, we rely on the Box's test. The M must be the highest possible and the significance of the F test must tend to 0 because if it is greater than 0.05, the analysis is invalid.

- H0: meaning greater than 0.05: Equal variance matrices variances intra classes

- H1: meaning less than 0.05: Unequal variance covariance matrices intra classes.

The following table provides the covariance of the different explanatory variables for each class of the dependent variable

Table 5. Test results

Box's M		443.163
F	Approximately	72.602
	df1	6
	df2	264237.542
	Signification	.000

According to the Box's test, we find that the level of significance is less than 0.05, so we reject the null hypothesis H0 of equal variance-covariance matrices intra classes and we accept the hypothesis H1 of inequality variance-covariance matrices and therefore the condition is verified.

The analysis by the box test allows us to conclude that the test is considered good since he admits quite a high value of 443.163 and meaning Fisher ensured equal to 0.000, hence our discriminant analysis model is validated and the condition is verified.

4.1.3 Canonical Discrimination Function

The canonical correlation is root of the variance between groups divided on the total variance of the discriminant function. It measures the proportion of variation of the discriminant function due to the difference between the groups.

Zero correlation "equal to 0" indicates that the variation of the discriminant function is not due to the groups; so that a correlation equal to 1 means that any variation of the discriminant function is due to the groups.

When the overall correlation, it is measured by the following table which shows us the values that have been associated with various factors as well as the cumulative percentage of the discrimination, more canonical correlation, the higher the model is good.

Table 6. Eigen values

		Eigen values		
Function	Eigen values	% of variance	Cumulative %	Canonical correlation
1	.245a	100,0	100,0	.444

Note. The first 1 canonical discriminant functions were used for analysis.

More canonical correlation is to 1, the more the model is better. In our case, the canonical correlation is equal to 44, 4%. This result is acceptable because this value confirms a significant discriminative power of the extracted discriminant function. In discriminant analysis, the number of Eigen values is equal to (k-1) with K: the number of classes which is equal to two classes in our model and therefore the number of Eigen values is equal to 1 or (2-1).

After reviewing the discriminatory power of the model, we will use the Wilks' lambda test to test the quality of the discriminant function.

4.1.3.1 Wilks Lambda Test

The Wilks' Lambda used to test the importance of the discriminant function by measuring differences discriminant function between groups. More lambda value is lower the more the model is good.

Table 7. Wilks Lambda test

Wilks's Lambda				
Test of functions	Wilks's Lambda	Chi-square	Df	Signification
Through 0 1	.803	41,572	3	.000

The lambda value is equal to 0.803 with a chi-square test significant with a risk of 5%. More significance tends to 0, more our model is good

In our case, the model is good because the significance is equal to 0.000. This means that at the global level, the difference in the group means is significant.

4.1.4 The Score Function

The discriminating power is observed through the table “Standardized Canonical Discriminant Function Coefficients”.

Table 8. Standardized canonical discriminant function coefficients

	Function
	1
RATIO 8	-.381-
RATIO 20	.703
RATIO 22	.521

Z is the linear combination of explanatory variables that allow us to better distinguish good borrowers from bad ones.

The higher the Z score, the higher default risks are low. The most discriminating variable is one that has the highest coefficient in absolute value. In our case it is the variable R20 “relative value of working capital.”

In order to calculate the probability of default, we calculate the score in the first step based on our ratios that have been selected as the most discriminating.

Hence the score function is represented as follows:

$$Z(X) = -0.381 * \text{RATIO8} + 0.703 * \text{RATIO20} + 0.521 * \text{RATIO22}$$

4.1.5 Default Probability Estimation Using Discriminant Analysis

$$DP = 1 / (1 + e^{(-0.381 * \text{RATIO8} + 0.703 * \text{RATIO20} + 0.521 * \text{RATIO22})})$$

The variables that best discriminated between failing and healthy companies are represented as follows: Ratio 8, Ratio 20, and Ratio 22.

4.1.6 The Predictive Power of the Model

The predictive power of the model is provided by the “ranking results.” The good overall clearance rate is the rate of borrowers highly ranked in their original group.

In other words, it is the sum of the number of enterprises of good borrowers classified by the model as good and bad borrowers classified by the model divided by the total number of the sample.

However, discriminant analysis was used to bring out the table 11 which is as follows:

Table 9. Classification results

F-H			Predicted Group Membership		Total
			Failing	Healthy	
Original	Effectif	Failing	75	23	98
		Healthy	23	74	97
	%	Failing	76.5	23.5	100.0
		Healthy	23.7	76.3	100.0
Count %	Effectif	Failing	75	23	98
		Healthy	23	74	97
	%	Failing	76.5	23.5	100.0
		Healthy	23.7	76.3	100.0

Note. Cross-validation is performed only for observations from the analysis. In cross-validation, each observation is classified by the functions derived from any other observations.

76.4% of the original observations classified correctly.

76.4%. of cross-validated observations classified correctly.

A total of 149 well classified observations (75 + 74) among 195 submissions in total.

- 74 companies are classified as “Good” (expected and observed).
- 75 companies are classified as “Bad” (expected and observed).

Generally, the rate of correct classification is retained as the main criterion for assessing the effectiveness of the scoring function to rank borrowers. This table shows that the function score above extracted classifies good, 76.4% ((75 + 74) / 195) of borrowers in the base sample.

This rate can dissect and (a failure is predicted successfully 76.5% while the failure is not predicted successfully 76.3 %.)

- The percentage of good borrowers is equal to $(74/97) = 76.3\%$
- The percentage of bad borrowers is equal to $(75/98) = 76.5\%$

As we know, the sample used to estimate the rate of correct classification directly affects the results. This rate is generally optimistic theory overestimates the real rate of correct classification

Table 10. Functions for group's barycentres

F_H	Function
	1
Failing	-,490
Healthy	,495

Note. Canonical discriminant functions no standardized valued at group averages.

From Table 10 we find that each borrower can be classified according to the following decision rule for good and bad borrowers, where:

- If the score $Z_i > C$, The company is classified as “healthy”.
- If the score $Z_i < C$, The company is classified as “failed”.
- If the score $Z_i = C$, The company is classified by random drawing of the group.

With C: midpoint of the segment joining the centroids of the two groups.

Each individual discriminating individual score is then compared the two average scores assigned to the group and it is the closest. To assign the company failed and healthy, one must determine a score that plays the role of border between the groups. If the groups are of equal size, the critical score is equal to the mean of the scores of the groups. If $Z_i \text{ score} > 0$, then the company is classified as “healthy”.

Each company is classified according to the score of our scoring function; it is assigned to the group in which it obtains the highest score.

The allocation to groups will be based on comparison with a discriminant score “average” for each group. This score is calculated from the discriminant function, where it replaces the individual values by means of the independent variables for the group of which it is responsible.

4.2 Logistic Regression

Logistic regression or Logit model is used to study the relationship between a dependent variable and the explanatory variables. The goal of logistic regression is to test whether the global variable in the model has a significant explanatory power.

The logistic regression equation is shown as follows:

$$Z = B_0 + B_1X_{i1} + B_2X_{i2} + \dots + B_nX_{in} + \xi \quad (5)$$

Where:

- $B_0 \dots B_i$: The coefficients representing the linear combination.
- X_i : measure the different variables that will allow for the probability of default.
- ξ_i : follows a logistics.

For a specification and estimation of the model using logistic regression through SPSS, we select from the “Analysis” menu, the procedure “Binary Logistic” of the “Regression” menu, selecting the options

“classification chart” “Hosmer-Lemeshow adjustment Quality”, “estimates Correlations” and “iteration history.” Then we choose, use the “likelihood of downward relative” method so that SPSS selects only the truly significant variables in the final model.

This summary established lets us know that no comments were removed from the analysis due to a probable illogical or missing value.

4.2.1 Model Generation

The generation of models is represented through the following code table:

Table 11. Dependent variables coding

Original value	Internal value
0	0
1	1

4.2.2 Model Specification and Estimation of Predictive Parameters

However, we can present the results of estimating the parameters of the logistic model in the table below:

Table 12. Estimated parameters of the logistic model

Variables in the equation		A	E.S.	Wald	Df	Sig.	Exp(B)
Step 20a	RATIO1	,031	,011	7,690	1	,006	1,031
	RATIO2	-,009	,005	2,888	1	,089	,992
	RATIO8	,000	,000	6,869	1	,009	1,000
	RATIO15	,005	,002	6,244	1	,012	1,005
	RATIO20	,019	,004	18,365	1	,000	1,019
	RATIO22	,007	,003	6,619	1	,010	1,008
	RATIO23	-,007	,003	4,121	1	,042	,993
	RATIO26	,004	,007	,265	1	,606	1,004
	Constante	-1,014	,437	5,389	1	,020	,363

Note. Variables entered in the step 1: RATIO1, RATIO2, RATIO3, RATIO4, RATIO5, RATIO6, RATIO7, RATIO8, RATIO9, RATIO10, RATIO11, RATIO12, RATIO13, RATIO14, RATIO15, RATIO16, RATIO17, RATIO18, RATIO19, RATIO20, RATIO21, RATIO22, RATIO23, RATIO24, RATIO25, RATIO26, RATIO27, and RATIO28.

The Wald test is used to select the different most significant variables to explain the dependent variable.

The software has selected nine variables that are significant and which calculate the probability of default on the repayment of a company “Ratios Value added, share of households, time settlement of supplier credits, net Return on equity, Value on the working Capital and general liquidity.”

The variables that best discriminated between failing and healthy companies are represented as follows:

Ratio1, Ratio2, Ratio8, Ratio15, Ratio20, Ratio22, Ratio23, Ratio25, Ratio26.

In order to calculate the probability of default, we calculate the score in the first step based on our ratios that have been selected as the most discriminating.

Hence the score function is represented as follows:

$$Z = (-1,014) + 0.01 * ratio1 - 0.09 * ratio2 + 0.0001 * ratio8 + 0.005 * ratio15 + +0.019 * ratio20 + 0.007 * ratio22 - 0.007 * ratio23 - 0.077 * ratio25 + 0.004 * ratio26$$

4.2.3 Default Probability Estimation Using Logistic Regression

$$DP = 1 / (1 + e^{(-1,014) + 0.01 * ratio1 - 0.09 * ratio2 + 0.0001 * ratio8 + 0.005 * ratio15 + +0.019 * ratio20 + 0.007 * ratio22 - 0.007 * ratio23 - 0.077 * ratio25 + 0.004 * ratio26})$$

4.2.4 Model Summary

The quality of the model is evaluated by the Log Likelihood. The objective is to minimize the error term. In this step, we can use the estimator of maximum likelihood through the following table:

Table 13. Model summary

Step		-2log-likelihood	Cox & Snell R square	Nagelkerke R square
dimension0	1	163,842a	,416	,554
	20	181,581b	,359	,479

Note. The estimate was stopped at number 10 iteration because parameter estimates changed by less than .001. The estimate was interrupted iteration number 9 because parameter estimates changed by less than .001.

The quality of the model is evaluated by the -2log likelihood -2log. However, the log-likelihood value measures the explanatory power of the model. Both Pseudo R2 allow us to explain the percentage of the binary dependent variable that is explained by the variables.

The test represents the statistic “Cox & Snell” and “Nagelkerke”. Note that the statistics “Nagelkerke” is an adjusted version of the “Cox & Snell.” We note from this table that of these two indicators increase as the software proceeds with the introduction of a significant variable. In our model, these statistics rise from 35.9% and 47.9% respectively. 47.9% of company failure might be explained by the 9 variables in our model.

The selected explanatory variables are significant individually at the 5% threshold; they are useful and important in explaining and predicting the probability of default for new credit.

However, the overall significance of the model is also provided by “the Hosmer-Lemeshow test” (test adjustment).

This shows whether the specified model is good or bad. It is based on the following hypothesis:

- H0: Bad adjustment
- H1: Good adjustment.

Table 14. Hosmer-Lemeshow test

Step		Chi-chi-square	Df	Sig.
dimension0	1	19,924	8	,011
	20	12,199	8	,143

Note. We find that, at the 5% threshold, the model adjustment is good because the probability (or significance) of the chi square to 8 degrees of freedom (ddl) is 0.011.

4.2.5 Predictive Power of the Model (Classification Performance)

The logistic regression was used to bring out the table is as follows:

Table 15. Global classification table

Observations			Predicted		Correct Percentage
			F_H		
			Failing	Healthy	
Step 1	F_H	Failing	75	22	77,3
		Healthy	15	81	84,4
	Global Percentage				80,8
Step 20	F_H	Failing	72	25	74,2
		Healthy	20	76	79,2
	Global Percentage				76,7

Note. The cut value is ,500.

Table 15 shows the number of companies that are properly classified by the scoring function. For the “failing companies”, 72 of them are properly classified (a rate of 74.2% correct classification), while 76 “healthy companies” are properly classified (a rate of correct classification of 79, 2%).

5. Results

Generally, the rate of correct classification is retained as the main criterion for assessing the effectiveness of the

scoring function to rank borrowers. The model correctly predicts 76.7% of cases with a threshold of 0.5. This shows that the model has a statistically significant power.

The following table presents a comparison of good clearance rates:

Table 16. Comparing companies ranking between the two methods: Performance of Best Two Models

Classification Method	Healthy	Failing	Total
linear discriminant	76,3%	76 ,5%	76,4%
Logistics discrimination	77,4%	74,2%	76,7%

In general, there was slight performance logistics discrimination in relation to the linear discrimination. Indeed, the class discrimination logistics companies correctly in their original groups with a ratio of 76.7% against 76.4% in the case of linear discriminant.

Table 17. Percentage classification for healthy and failing SME companies by sector of activity using the two methods

Sector /H_F	Discriminant Analysis			Logistic Regression		
	Failing	Healthy	Total	Failing	Healthy	Total
Trade	45%	55%	100%	47,5%	52,5%	100%
Agriculture	66,66%	33,33%	100%	66,66%	33,33%	100%
Industry	34,38%	65,62%	100%	34 ,38%	65 ,62%	100%
Services	18,75%	81,25%	100%	18,75	81,25%	100%
Tourisme	60%	40%	100%	61 ,25%	38,75%	100%

Using the two methods, we conclude that the best performing sector with the highest percentage of healthy companies and the lower of failing companies is services with an equal percentage was 81.25% by both methods allowing to show the ability of our two methods “discriminant analysis and logistic regression” to have similar results.

Our second area having the healthiest companies is that of the industry with 65.62% of healthy companies ranked among 32 companies to sector. When the agricultural sector which includes 27 companies, it is characterized as the most failed with 18 failing companies and only 7 healthy companies and thus with equal failure rates was 66.66%.

6. Discussion

In Tunisia, small and medium-sized companies continue to rely on bank lending given the inadequacy of their self-financing rate and the inaccessibility to new instruments of direct finance. The Tunisian banking sector continues also to build on credit as main activity and this in spite of the exhaustion of the principal factors that have long supported the banking intermediation. This activity carries risks especially credit risk. That is why, we presented in this paper models for assessing risk of default to help bankers in their decisions on the granting of credits.

To do this, we used the credit scoring method to properly predict the probability of default of the counterparty by comparing discriminant analysis and logistic regression. The empirical results found show that these two scoring techniques have a statistically significant power in foreseeing default risk of the firms, and credit scoring is important in the decision-making process for granting credit. Logistic discrimination classifies enterprises correctly in their original groups with a rate of 76.7% against 76.4% in case of linear discrimination giving so a slight superiority to the first method.

References

- Allen, N., Berger, W., Scott, F., & Nathan, H. M. (2002). *Credit Scoring and the Availability, Price, and Risk of Small Business Credit*. FRB of Atlanta Working Paper No. 2002-6, FEDS Working Paper No. 2002-26. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=315044
- Altman E. I., Hademan, & Narayanan, R. G. (1977). Zeta analysis, A new model to identify bankruptcy risk of corporation. *Journal of Banking and Finance*, 1, 29-51. [http://dx.doi.org/10.1016/0378-4266\(77\)90017-6](http://dx.doi.org/10.1016/0378-4266(77)90017-6)

- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589-609. <http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Anderson, R. (2007). *The credit scoring Toolkit*. Oxford University Press.
- Bardos, M. (2005). Les scores de la Banque de France: Leur développement, leurs applications, leur maintenance. *Bulletin de la Banque de France*, N° 144. Décembre 2005. Retrieved From https://www.banquefrance.fr/fileadmin/user_upload/banque_de_france/archipel/publications/bdf_bm/etudes_bdf_bm/bdf_bm_144_etu_6.pdf
- Bardos, M. (2008). *Scoring sur données d'entreprises: Instrument de diagnostic individuel et outil d'analyse de portefeuille d'une clientèle*. Mireille Bardos. La revue MODULAD, N°38, 2008. Retrieved From <http://www.modulad.fr/archives/numero-38/Bardos-38/Bardos-38.pdf>
- Bardos, M., & Zhu, W. (1997). Comparaison de l'analyse discriminante linéaire et réseaux de neurones, applications à la détection de défaillance des entreprises. *Revue de statistique appliquée*, XLV4, 65-92. Retrieved from http://archive.numdam.org/ARCHIVE/RSA/RSA_1997__45_4/RSA_1997__45_4_65_0/RSA_1997__45_4_65_0.pdf
- Beaver, W. H. (1968). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111. <http://dx.doi.org/10.2307/2490171>
- Beaver, W. H. (1968). Market prices, financial ratios and the prediction of failure. *Journal of Accounting Research*, Autumn, 179-192. <http://dx.doi.org/10.2307/2490233>
- Berger, Frame, & Miller. (2005). Credit Scoring and the Availability, Price, and Risk of Small Business Credit. *Journal of Money, Credit, and Banking*, 37, 191-222. <http://dx.doi.org/10.1353/mcb.2005.0019>
- Caire, Dean, & Robert, S. K. (2002). *Credit Scoring: Is It Right for Your Bank?* Bannock Consulting, Working paper. Retrieved from http://www.microfinance.com/English/Papers/Scoring_For_Your_Bank.pdf
- Conan, J., & Holder, M. (1979). *Variables explicatives de performances et contrôle de gestion dans les PMI*. Thèse de Doctorat en sciences de gestion, Université de Paris IX.
- Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, spring 1972, 167-179. <http://dx.doi.org/10.2307/2490225>
- DeYoung, Hunter, & Udell. (2004). The Past, Present, and Probable Future for Community Banks. *Journal of Financial Services Research*, 25, 85-133. <http://dx.doi.org/10.1023/B:FINA.0000020656.65653.79>
- Dietsch, M., & Petey, J. (2008). *Mesure et gestion du risque de crédit dans les institutions financières* (2nd ed.). Edition Revue Banque.
- Dominique, L. L. (2005). *Les nouvelles exigences en fonds propres réglementaires des banques et l'évaluation en juste valeur des instruments financiers*. Bâle II et IAS 39 2005. Retrieved From https://www.researchgate.net/publication/268187405_Bale_II_et_IAS_39_Les_nouvelles_exigences_en_fonds_propres_reglementaires_des_banques_et_l_evaluation_en_juste_valeur_des_instruments_financiers
- Fisher. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 179-188. <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Frachot, A., & Georges, P. (2001). Aide à la décision: Avantage au scoring face au système expert. *Banque Magazine*, n°627, Juillet-août 2001, 45-47. Retrieved From http://www.revue-banque.fr/article/aide-decision-avantage-au-scoring-face-au-systeme#restricted_content
- Frame, W. S., Aruna, S., & Lynn, W. (2001). The Effect of Credit Scoring on Small Business Lending. *Journal of Money, Credit, and Banking*, 33(3), 813-825. <http://dx.doi.org/10.2307/2673896>
- Gourieroux, C., & Tiomo, A. (2007). *Risque de crédit: Une approche avancée*. Les cahiers du CREF. Retrieved from <http://neumann.hec.ca/cref/pdf/c-07-05f.pdf>
- Kesseven, P., Carole, H., & Narasimhan, M. S. (2012). Working capital financing preferences: The case of manufacturing small and medium sized entreprises "SME". *Asian Academy of Management Journal of Accounting and Finance AAMJAF*, 8(1), 125-157. Retrieved From <http://web.usm.my/journal/aamjaf/vol%208-1-2012/8-1-6.pdf>
- Loretta, J. M. (1997). What's the Point of Credit Scoring? *Business Review*, SEPTEMBER/OCTOBER 1997. Retrieved from

- <https://philadelphiafed.org/research-and-data/publications/business-review/1997/september-october/brso971m.pdf>
- Matoussi, H., & Krichène, A. (2010). *La prévention du risque de défaut dans les banques Tunisiennes*. Crises ET nouvelles problématiques de la valeur, Nice: France (2010). Retrieved from <https://halshs.archives-ouvertes.fr/hal-00481087/document>.
- Matoussi, H., Mouelhi, R., & Sayah, S. (2011). *La prediction de faillite des entreprises tunisiennes par la regression logistique*. 20ÈME CONGRES DE L'AFC, May 1999, France. pp. CD-Rom, 1999. Retrieved from <https://halshs.archives-ouvertes.fr/halshs-00587769/document>
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, Spring, 109-131. <http://dx.doi.org/10.2307/2490395>
- Philippe, A., & Fredj, F. (2012). *Acces au credit et promotion des PME en Tunisie*. ERUDITE, Université Paris 12. Retrieved from <http://fseg.univ-tlemcen.dz/larevue09/philip%20adair%20+fredj%20fhima.pdf>
- Rangau, U., & Ouertani, N. (2010). Corporate Default Analysis in Tunisia using Credit Scoring Techniques. *International Journal of Business*, 15(2). Retrieved from <http://www.craig.csufresno.edu/ijb/Volumes/Volume%2015/V152-5.pdf>
- Saunders, A., & Allen, L. (2002). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*. New York: John Wiley and Sons. Retrieved From http://www.untag-smd.ac.id/files/Perpustakaan_Digital_1/CREDIT%20RISK%20Credit%20risk%20measurement,%20New%20approaches%20to%20value%20at%20risk%20and%20other%20paradigms.pdf
- Stéphane, T. (2007). *Data Mining et Statistique décisionnelle, l'intelligence des données*. Editions Technip.
- Stewart, C. M., & Nicholas, S. M. (1984). Corporate Financing and investment decisions when firms have information that investor do not have. *Journal of Financial Economics*, 13, 187-221. North-Holland. Retrieved from <http://www.sciencedirect.com/science/article/pii/0304405X84900230>
- Thomas, L., Edelman, C., & Crook, L. N. (2002). *Credit Scoring and Its Applications*. Philadelphia: Society for Industrial and Applied Mathematics. <http://dx.doi.org/10.1137/1.9780898718317>
- Vivien, B. (2009). *Gestion des risques et risques de crédit*. Université d'Evry, pp. 3-38. Retrieved from <http://leguiedurentier.free.fr/documents/risquescredit.pdf>.
- Wallis, L. P. (2000). Credit-Scoring: The Future of Decisioning in the A/R Process. *Business Credit*, 103(3). Retrieved from <https://www.questia.com/magazine/1P3-69560157/credit-scoring-the-future-of-decisioning-in-the-a-r>

Appendix A

Table A. Financial ratio

Ratio	Financial Ratio
Ratio1	Added Value Ratio
Ratio2	Households's Part
Ratio3	Company Part
Ratio4	Bank's Part
Ratio5	Part of the State
Ratio6	Stock's rotation
Ratio7	Recovery Period of Receivables
Ratio8	Settlement periods of supplier credits
Ratio9	Gross Margin
Ratio10	Net Margin
Ratio11	Cash Flow
Ratio12	Gross profitability of total assets
Ratio13	Net profitability of total assets
Ratio14	Gross return on equity
Ratio15	Net return on equity
Ratio16	Dividend Distributions
Ratio17	Dividend distribution capacity

Ratio18	Surplus /Insufficient equity
Ratio19	Surplus/Insufficient of non-current liabilities (financial liabilities)
Ratio20	Working capital (capital items)
Ratio21	Working capital needs
Ratio22	Liquidity in the broad sense
Ratio23	Liquidity in the strict sense
Ratio24	Financial autonomy
Ratio25	financial charges
Ratio26	Repayment capacity
Ratio27	Leverage
Ratio28	Net investment in fixed assets

Appendix B

Table B. Sample distribution

		Effectif	Percentage	Valid Percentage	Cumulative Percentage
Valid	Trade	40	20.5	20.5	20.5
	Agriculture	27	13.8	13.8	34.4
	INDUSTRY	32	16.4	16.4	50.8
	SERVICE	16	8.2	8.2	59.0
	Tourism	80	41.0	41.0	100.0
	Total	195	100.0	100.0	

Appendix C

Table C. Healthy and failing companies distribution

		Effectif	Percent	Valid Percent	Cumulative Percent
Valid	Failing	98	50.3	50.3	50.3
	Healthy	97	49.7	49.7	100.0
	Total	195	100.0	100.0	

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).