Quarterly Journal of Business Studies
Vol. 2, No. 1, 2015, 34-54

Distressed Company Prediction Using Logistic Regression: Tunisian's Case

Fayçal Mraihi[1]

Abstract

In this study, we try to develop a model for predicting corporate default based on a logistic regression (logit) and applied to the case of Tunisia. Our sample consists of 212 companies in the various industries (106 companies 'healthy' and 106 companies "distressed") over the period 2005-2010. The results of the use of a battery of 87 ratios showed that 12 ratios can build the model and that liquidity and solvency have more weight than profitability and management in predicting the distress. Both on the original sample and the control one, these results are good either in terms of correct percentage of classification or in terms of stability of discriminating power over time (on, two and three years before the distress) and space.

Keywords: distressed firms, forecasting model, logistic regression model

1. Introduction

Many firms react very late or improperly facing the first signs of distress. Three to five years elapse, usually between the early difficulties encountered by the company and the first operating mechanisms.

This delay generally results from a lack of understanding of the mechanisms and causes the degradation of process and an obvious lack of foresight. Thus, it is useful to examine the sequence that implies that process and to define, in the area of prevention, methods or models to predict the decline of the company in the medium term.

An objective definition of a distressed company or a firm in a difficult situation does not exist, so we can refer to the definitions suggested by Haehl (1981) and The French Superior Council of Economic Professions (FSCEP).According to the first definition « In state of difficulty the company which, because of certain economic, financial or human imbalance, revealed by the conjunction of diverse indications, ratios, and the examination of all elements, cannot envisage in the predictable, short and medium-term future, to continue its activity in a normal way or could only by proceeding in transactions of partial liquidation, economic transformation, inflow of outer permanent capital or redundancy of a part of the staff ».

For the second definition « In the absence of legal definition on the subject, and to define the firm in difficulties we can base on the criteria of liquidity, solvency, profitability and added value and to consider that a company is in a difficult situation from the moment it evolves in such a way, for economic, financial, organizational, social or other reasons, it will meet sooner or later difficulties to generate the sufficient income to fill its legal and contractual commitments and make the necessary investments ».

In such context, to which is added a bubbling socioeconomic environment, the regular appeal to the diagnosis establishes not only a requirement of good management, but also an imperative for the survival of the company.

[1]Faculty of Law, Economics and Management of Jendouba, Tunisia

A successful diagnostic has to detect, in time, the causes of the distressing. These causes show themselves in the company by a battery of indicators that must be identified as soon as possible to a successful recovery plan.

The diagnostics of default risk knew an important development through the use of multivariate statistical methods to analyze the financial situation from a given set of ratios. Among the most commonly used statistical methods, we find logistic regression. The principle of this method is the following: having the characteristics described by financial ratios, and a sample of companies that cover both "healthy" companies and "distressed" companies, logistic regression leads to determine the best combination of ratios to differentiate the two business groups.

To achieve this goal and to develop a model for predicting corporate default based on a logistic regression, this article will address, in a first section, the methodology through the presentation, writing and justification of the model used, the constitution of the samples and the set of distressed determinants, while being interested in the Tunisian case. The estimate of the discriminatory power of the model in time and space will be in the second section. The third section analyzes the sensitivity that will allow us to test the elasticity of the model results due to the variation of the explanatory variables. Thus, we try to classify, in the fourth section, each ratio according to its degree of participation in the discriminatory power of the model.

2. The Methodology

In this work, we use regression for predicting business distress, and then we test its validity in time and space. However, it is primordial to define what a logistic model is, explain its approach and show its usefulness, then present the hypotheses and tests to perform and discuss the constitution of the samples.

- **Overview and Principle of the Logistic Model**
  - **Literature Review**

Logistic regression, viewed as a generalization of linear discriminant analysis, has been introduced by Day & Kerridge (1967), Cox (1970), and developed by Anderson (1972, 1982), Martin (1977), Olshon (1980) who was the pioneer in the use of logistic regression in the domain of prediction of business distresss. Among the major works that have used this method we can cite Mensah (1984), Albert & Lesaffre (1986), Aziz & *al* (1988), Bardos (1989), Burgstahler & *al* (1989), Flagg & *al* (1991), Platt & Platt (1991), Zopounidis (1995), Bardos et Zhu (1997), Mossman & *al* (1998) and more recently Altman & *al* (2005), Jones & Hensher (2004, 2007, 2007a), Zeitun & *al* (2007), Li & *al* (2011), Ahn & al (2011), Tserng & al (2011), kim & Kang (2012), Serrano-cenca & al (2013) et Wang & al (2014), Yu & al (2014).

As in multiple linear regression, it is relates to estimate parameters of model, to measure its adequacy (quality of adjustment) and to deduce the significance and the interpretation of the estimated parameters. Logistic regression is an econometric technique with a dichotomous dependent variable y_i, representing the state of the company that takes:

- The value 1 if the company is "distressed"
- The value 0 if the firm is "healthy".

This type of regression allows to determinate the probability that a firm is classified in the group of « healthy » or the group of « distressed ».

At this discrimination, there can be two types of errors:

- The error of the first kind I: classify a distressed company with the healthy ones.
- The type of the second kind II: classify a healthy company with distressed ones.

We must notice, however, that the cost associated with the error of the first kind is very different from that associated with type II. Indeed, the first cost is that a creditor support in case of default of the debtor. While the second one is an opportunity cost representing the difference between remuneration that a creditor could collect on the, not accepted, and the rate of return offered by the use of these funds.

To the extent that the cost of a Type I error is much higher than that of a Type II error (about 1 to 20 according to Altman et al. "Zeta analysis" in 1977), then it seems more relevant to judge the quality of the model on the basis of correct classification percentages, in general, and the error rate of type I that it induces, in a particular way.

In general, from a sample of base and a set of ratios, we will proceed as follows:

- Check the distribution normality of selected ratios by eliminating those not responding to the corresponding test.
- Examine the individual discriminating power of these ratios by classifying them by categories.
- Evaluate the existing correlations between the ratios by eliminating those that are redundant.
- Observe the discriminating power of different combinations and select by iteration the combination that offers the best correct percentage of classification with the lowest cost of the first kind, that is the one that provides the best value:

intergroup dispersion / intragroup dispersion.

○ **Logistic Model Principle**

we have :

$y_1, y_2, ...y_n$: random variables, called dependent variables, each taking the value 1 or 0, values that correspond to groups G1 and G2 to discriminate.

$x_1, x_2, ...x_J$: the components of a multi-dimensional vector $X = (x_1, x_2, ...x_J)$ and that represent random variables called explanatory or independent variables.

$(\beta) = (\beta_0, \beta_1, ...\beta_J)$: are the unknown coefficients of the model to be estimated.

The idea is to build a model linking $\pi(x) = p[\ Y=1/ X\ ]$ (he probability that Y = 1 given X).

With :

$$\text{probability of default } [\pi(x)] = P(Y = 1 / X = x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+.......+\beta_K x_K)}}$$

and

$$\text{probability of non-default } [1-\pi(x)\ ] = P(Y = 0 / X = x) = 1 - \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+.......+\beta_K x_K)}}$$

- **Hypotheses and Significance Tests of the Coefficients**

Formally, the null hypothesis is as follows:

$$H_0 : a_1 = a_2 = \cdot \cdot \cdot = a_k = 0$$

This is a global evaluation assessment of the regression. Indeed, if the null hypothesis is accepted, it would mean that none of the explanatory variables contribute to the explanation of the dependent variable. The model can be rejected.

H1: at least one of the coefficients is non-zero.

The objective of significance tests is to determine the role of each of several or all, of explanatory variables.

We have two approaches to test the hypotheses:

- Use the principle of the likelihood ratio. The approach is generic and consistent with the process of parameter estimation. It can detect better the alternative hypothesis when it is true. The disadvantage is that it is heavier in terms machine. Indeed, every hypothesis to evaluate gives rise to a new estimation of the parameters, so to a process of optimization. Certainly, software and computers today are very efficient, but when the databases processed are important, the calculations to be made will not be as significant as that.

- Use the asymptotic normality of estimators (maximum likelihood). We talk about Wald test. The main advantage is that the information that we want to use, are all-available when estimating the global mode, including all variables. The obtaining of the results is immediate. A disadvantage is that the Wald test is conservative; it tends to favor the null hypothesis.

- **The Constitution of Samples and Variables Determination**

The choice of the sample posed us serious problems. Indeed, the implementation of logistic regression assumes the existence of two business groups « healthy » and « distressed ». The selection of the reference population leads to a choice between two alternatives:

- Constitute a sample the widest possible, which includes companies from different industries, size, geographical location and economic environments.

- Choose a reference population so as to guarantee the homogeneity of the sample, leave to limit its size.

In practice, and according to most studies [Beaver (1966), Altman (1968), Edmister (1972)], we adopted the option of a larger sample affecting several sectors. Our sample consists of 212Tunisian companies in the various sectors (which will be discussed below), (106 "healthy" companies and 106 "distressed" companies) over the period 2005-2010.

The "healthy" companies were selected from the Tunisian stock exchange and among statutory accountants. While "distressed" companies come from the office of assistance to companies in difficulty, which sits at the Ministry of Industry. The selection of firms in difficulty was based on the following criteria:

- Be suspension of payments for at least six months
- Have very serious social problems,
- Must be identified by statutory auditors, National Social Security Fund or fiscal institutions

From this basic sample, and referring to the approach of Platt and Platt, (1991); Altman et al, (1994); Bardos (1998a) and Varetto (1998), it was possible to set up two sub-samples:

- A first, called "Initial" sample consisting of 152 companies, 76 "healthy" and 76 "distressed". We'll take the last three years of the same companies to form three sub-samples we call "Initial one year prior to distress," "Initial two years before distress" and "Initial three years prior to distress." these sub-samples used to develop the model and to test its validity in time.

- A second sample, called "Control" sample, composed of 60 other companies, 30 "healthy" and 30 "distressed". From the last three years of these companies, we will establish three sub-samples that we call "control one year prior to distress," "Control two years prior to distress" and "Control three years prior to distress." These sub-samples are designed to test the validity of the model in space.

Companies belonging to both sample of "healthy" and the "distressed" companies are distributed between the different sectors as follows:

Table 1: The Distribution of the Companies between the Different Sectors

Companies / Sectors	Healthy	Distressed
Textile , Clothing and Leather Industries	28	23
Food-processing industry	23	19
Various industries	19	19
Industries of Building materials, Ceramic and Glass	13	18
Mechanical engineering industries, Metallic,	11	13
Metallurgical and Electric	8	9
Services (hotel)	4	5
Chemical industries		
Total	**106**	**106**

In the absence of a theory of business distress, the choice of indicators is completely subjective. Indeed, it is based on experience and intuition of the one who develops the model. Generally, this choice often results from previous choices, this is to say the choice of all first authors of reference(Ramser and Foster, 1931 ; Fitzpatrick, 1932 ; Winakor and Smith, 1935 ; Merwin, 1942 ; Beaver, 1966 ; Altman, 1968 ; Deakin, 1972 ; Edmister, 1972 ; Blum, 1974 ; Altman and al, 1977 ; Taffler, 1983).

The number of ratios that can be included in a financial analysis is extremely high. To avoid making an excessively statistical treatment, we limited ourselves to ratios calculated on the basis of different valuesrelative to the same year and concerningthe Fundamental and classic aspects of the financial analysis: liquidity, funding, debt, profitability, balance sheet structure and financing costs.

Moreover, for each category, we selected three or four ratios, in order to avoid a high number of ratios for the study to be carried out and thus avoid the redundancy phenomenon. But on the other hand the number of ratios should not be too small for all aspects of business situation are covered.. Despite these limitations, we were finally brought to retain only 87 ratios shown in Appendix 1.

The assignment of a ratio to one or to the other categories can be discussed. Indeed, among selected ratios some are composite in nature and thus reflect, at the same time, several aspects of corporate behavior to be taken into account in the interpretation. This classification has only for objective the convenience of the presentation and the analysis of the results.

3. Estimation of the Model Parameters

From the three subsamples which we called "Initial one year prior to distress," "Initial two years before distress" and "Initial three years before distress," each consist of the same 152 firms (76 "distressed" and 76 "healthy") but for different years (each sample is interested in the same year for all companies), and a set of 87 ratios (Appendix 1), we will try to formulate a logistic model, estimate its coefficients, calculate the probability of default in posteriori and develop a decision rule.

Table 2 : Classification Table[a,b]

			Predicted		
			Y		
		Observed	0	1	Percentage Correct
Step 0	Y	0	0	76	,0
		1	0	76	100,0
		Overall Percentage			50,0

Constant is included in the model.
The cut value is ,500

To perform the estimation, we used the "SPSS" software.

In a first step, it was assumed a model with 87 explanatory variables. The estimated model has provided us with results rather critical because the error rate is 50%:

Such an error rate is explained by the importance of correlations between the explanatory variables: collinearity problem, correlation matrix and variance-covariance. Thing that leads us to take great care in selecting all ratios. Indeed, the number of ratios should not be too high for the study to be performed (Rose and Giroux (1984) identified more than 130 different ratios). Also, the phenomenon of redundancy between ratios must be avoided: from the analysis of the correlation matrix, we observed a strong correlation between some explanatory variables; there is a great redundancy (the same information is provided by several ratios).

To solve this problem of collinearity, we opted for the "Feedward" method. It consists in introducing into the model, each time, the most correlated explanatory variable with the dependent variable until the matrix becomes not inversible. During this operation, we must be careful and retain only the independent variables that are significant at the 5% and can improve the $\overline{R}^2$ and we will ensure that all aspects of the situation of the company are covered.

Once this is done, based on 87ratios initially taken, we are left with only 12ratios, which will constitute the explanatory variables of the model to be estimated.

The estimate by the logit model gives the following results:

Table 3 : Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	R_5	14,088	15960,342	,000***	1	,999	1312882,320	,000	.
	R_6	-131,311	43256,749	,000***	1	,998	,000	,000	.
	R_7	-272,144	40875,140	,000***	1	,995	,000	,000	.
	R_{15}	10,482	20133,088	,000***	1	1,000	35663,913	,000	.
	R_{19}	-23,350	13228,722	,000***	1	,999	,000	,000	.
	R_{26}	66,129	15652,150	,000***	1	,997	5,243E28	,000	.
	R_{28}	178,682	40767,715	,000***	1	,997	3,988E77	,000	.
	R_{33}	-13,401	6809,594	,000***	1	,998	,000	,000	.
	R_{40}	87,654	29863,406	,000***	1	,998	1,169E38	,000	.
	R_{61}	-,502	319,246	,000***	1	,999	,606	,000	3,348E271
	R_{74}	-15,515	25788,736	,000***	1	1,000	,000	,000	.
	R_{79}	52,925	14977,442	,000***	1	,997	9,663E22	,000	.
	Constant	126,426	38236,323	,000***	1	,997	8,052E54		

a. Variable(s) entered on step 1: R5, R6, R7, R15, R19, R26, R28, R33, R40, R61, R74, R79.

A careful analysis of the Wald test shows that all the variables used by the model are significant at a rate of 5 %.

The last twelve ratios represent the explanatory variables in our final model:

$\boldsymbol{Z} =$ 14,057 R_5 - 131,311 R_6-272,144 R_7 + 10,482 R_{15} - 23,350 R_{19} + 66,129 R_{26} + *178,682* R_{28} - *13,401* R_{33} + 87,654 R_{40} - 0,501 R_{61} - 15,515 R_{74} + 52,925 R_{79} + *126,426*

R_5 = Cash and cash equivalents / current liabilities

This is the quick ratio (ratio of immediate liquidity), which determines the proportion of current liabilities covered by cash and equivalents of liquidity.

R_6 = Permanent Capital / Total Balance Sheet

This is a ratio that measures the creditworthiness (solvency) of the company reporting the means of stable funding to total assets.

R_7 = Current assets / Total assets

This ratio represents the degree of liquidity; it defines the importance of current assets relative to total real assets.

R_{15} = Equity / Total Assets

This ratio, called the ratio of financial autonomy is particularly studied by bankers because their equity represents a guarantee. Indeed, in case of liquidation of the company, share holders will be last served in case of the sale of assets. If the assets are insufficient to cover liabilities, the loss will thus be imputed on stockholders' equity before being on other debts.

R_{19} = Short-Term Debt / Total Liabilities. It measures the share of short-term debt of the company in all of its liabilities. It is an indicator of the debt structure.

R_{26} = Amortization of Capital Assets / Gross Fixed Assets. This ratio is often used as an indicator of the degree of aging equipment

R_{28} = Working Capital / Total Assets. This ratio expresses the degree of liquidity of the firm. Indeed, he reports the excess of current assets after providing for short-term debt relative to total assets.

R_{33} = current assets (excluding stocks) / current liabilities. The ratio of reduced liquidity is a more restrictive measure of the liquidity of a company than the current ratio. It indicates the portion of current liabilities covered by current assets excluding stocks.

R_{40} = current assets (excluding stock) / Total assets. This ratio is an indicator of the liquidity of the company; it expresses the proportion represented by trade receivables, investments and other current assets, liquidity and cash equivalents to total assets.

R_{61} = Medium and long-term debt / Cash flow

It is a debt ratio, it gives us information on the proportion that debt in the medium and long terms represents over resources generated by the activity of the company in terms of cash. This cash allows the firm to invest and continue its development.

R_{74} = Net Income / Total liabilities

It is a profitability ratio that expresses the proportion of net income for each currency of liabilities invested in the company.

R_{79} = Total Liabilities / Total Assets

This overall solvency ratio must be significantly less than one. Indeed, if its value is equal to ½, this means that the company has a significant debt capacity because in case of liquidation, for example, the value of its assets can be used to repay twice all its commitments.

In the equation used by logistic regression forecasting, we notice the presence of several ratios that have been selected as explanatory variables in previous studies.

Table 4: the presence of several explanatory ratios in previous studies

Ratio	Authors
R_6	Conan & Holder (1979) ; Holder & al (1984)
R_7	Deakin (1972) ; Taffler (1982) ; Holder & al (1984)
R_{15}	Le crédit commercial de France (1995)]
R_{19}	Beaver (1966) ; Plat & Plat (1991)
R_{26}	Altman & al (1974) ; le modèle du C.E.S.A. (1974)
R_{33}	Deakin (1972) ; Edmister (1972) ; Houghton (1984) ; Burgstahler & al (1989) ; Michalopoulas & al (1993)
R_{40}	Conan & Holder (1979)]
R_{61}	Conan & Holder (1979) ; Bardos (1984)
R_{79}	Deakin (1972) ; Rose & Giroux (1984) ; Burgstahler & al (1989) ; Michalopoulas & al (1993) ; Altman & al (1994)

The overall significance test used in the logistic regression is the chi-square with k degrees of freedom (k is the number of explanatory variables in our case k = 12). If the critical probability is less than the significance level that one is fixed, we can consider that the model is globally significant. In our model the statistical likelihood ratio (chi-square) is equal to 210.717; the critical probability associated is zero. The model is generally very significant, there is indeed a relationship between the explanatory variables and the variable to be explained.

Table 5: Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	210,717	12	,000
	Block	210,717	12	,000
	Model	210,717	12	,000

Similarly decrease in value - 2 logliklihood from one stage to another also indicates the same result, that the introduction of new variables improves the model. In our case, this value down from 210.717 to zero.

Table 6 : Itération History[a,b,c]

Itération		-2 Log likelihood	Coefficients
			Constant
Step 0	1	210,717	,000

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 210,717

c. Estimation terminated at iteration number 1 because parameter estimates changed by less than ,001.

Table 7 : Model Summary

Step	-2 Log likelihood	Cox &Snell R Square	Nagelkerke R Square
	,000[a]	,750	1,000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached.

Cox & Snell R Square and Nagelkerke R Square tests help to determinate the percentage of the binary dependent variable that is explained by the explanatory variables retained confirmed the significativity of our model. Indeed, the Nagelkerke R Square test is an adjusted version of the Cox &

Snell R Square one and therefore closer to reality. So, for our model, we notice that 100% of the variation in the dichotomous variable could be explained by the explanatory variables used and retained.

Once the overall significance of the model used is demonstrated, it remains to be seen whether the explanatory variables are significant. The Wald test in the logistic regression (see table above) demonstrates that, the twelve explanatory variables, retained in our model, are significant at 5 %.

The Hosmer and Lemeshow test divided into deciles based on predicted probabilities, then computes a chi-square from observed and expected frequencies. The value p = 100% here is calculated from the chi-square distribution with 6 degrees of freedom, it indicates that the logistic model used is excellent.

Table 8 : Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	,000	6	1,000

After checking the overall significance of the model and the significance of the explanatory variables, our job is now to verify the performance and stability of the logit model retained both in time, by applying it to the initial samples a year, two and three years prior to distress and in space using control samples a year, two years and three years before distress (Appendix 3-1, 3-2, 3-3, 3-4 and 3-5).

Estimation and Validation of the Discriminatory Power of the Model in Time and Space

- **Estimation of the Model Discriminatory Power One Year before Distress**

The estimation of the logit model on the original sample, one year prior distress, shows that in the "healthy" firms group, the model classifies all "healthy" firms in their original group correctly.

In the distressed companies group, that interests us the most, we find no firm misclassified, so the model classifies successfully both companies "healthy" as "distressed" (Appendix 1 and Appendix 3-1).

As far as the error Type I cost is much higher than that of an error type II [about 1 to 20 in Altman and al (1977)], then it seems more appropriate to judge the quality of the model on the base of the correct percentages of classification, in general, and of the error type I rate that it induces, in a particular way. These results "appear" as a whole interesting because they have the advantage of providing a combination of ratios based on which one can make a diagnostic of the company.

We say "appear interesting" because we should not judge the model before testing the performance over time (testing the model on the same companies but for different periods of time, two years and three before distress) and in space (testing the model on a control sample consisting of companies other than those in the sample of origin).

- **Validation of The Model Discriminatory Power Over Time**
  - **For The Same Companies Two Years Before Distress**

The validation of model on exercises that come two years before distress gives the results in Appendix 1 and Appendix 3-2.

In the « healthy » companies group, we find that the model correctly classifies all « healthy » firms in their original group. In the « distressed » firms group, there are five firms misclassified, so the firms are considered as "healthy" when they are actually distressed. The model retains thus its discriminatory power, since the percentage of correct classification varies by only 0.66% from 100% to 99.34%, the error type I increases from 0 to 1.32%, while the error type II remains zero.

  - **For the Same Companies Three Years Before Distress**

We will proceed in the same way as before, the same firms but for three years before distress, we get the results presented in Appendix 1 and Appendix 3-3.

In the group of « failed » firms, we find that the model classifies four firms in the group of « healthy » one, while they are « distressed » which produces an error type I of about 5.26%. In the group of « healthy » companies, all companies are correctly classified and we have a percentage of error Type II equal to zero.

The forecasting ability of selected ratios, showed a satisfactory stability over time, since the overall error rate only increased from 0% to 3.29% % over the last three years preceding the distress, particularly some stability is noted for the classification of « healthy » companies .The following table will present a summary of changes in correct percentages of classifications and in errors of type I and II in time.

Table 9: Results of Estimation in the Time

	1 year before distress	**2 years before distress**	**3 years before distress**
% of correct classification	100 %	99. 34 %	96.71 %
% of classement error	0 %	0. 66 %	3.29 %
% of error type I	0 %	1. 32 %	6.58 %
% of error type II	0 %	0 %	0 %

Indeed, we notice that for the model used, the percentage of the error Type I varied only by 6.58% between the first and third years before distress. Furthermore, we find that the correct percentage of classification decreased only by 3.29% (it goes from 100% to 96.71%).

For our model, the most interesting element, in addition to its high correct percentage of classification, it is the weakness of the error Type I whose cost is higher. Concerning the error type II, we see that it remains zero.

- **Validation of the Model Discriminatory Power in Space**

To test the discriminatory power of the model in space, we use a control sample consisting of two new groups. The first contains the distressed firms while the second contains "healthy" companies, each lists 30 firms. The model will be tested on companies other than those that were originated. The application of our Logit model on these samples gives us the estimates presented in Appendix 2 and Appendix 3-1.

In the « healthy » companies group, we find that the model classifies two firms in the « distressed » group when they are « healthy ». In the « distressed » group, there are also misclassified firms so they are considered by the model « healthy » when they are actually distressed.

This model has a remarkable accuracy by classifying 95% of the control sample correctly. The error Type I is around 10% while the error type II is zero.

Studying companies' exercises of control sample in case of two years before distress, we get the results announced at Appendix 2 and Appendix 3-4.

In the « healthy » companies group, we find that the model classifies all firms correctly so we conclude an error type II equal to zero. While in the group of distressed companies, there is a single firm misclassified, giving us an error Type I of about 3.33%.

The increase of the efficiency of the Logit function, in this validation test (it passed from 5% to 98.33%), is due to the fact that the two samples of distressed firms (the initial sample and the control one) are randomly selected from a pool of 106failed firms. Moreover, as the samples are both small, the distributions of firms by size and industry differ considerably and this affects the efficiency of the function.

If we further increase the time period between the prediction date and the advent of distress, using the same control sample but for three years before distress, we obtain the results reported in Appendix 2 and Appendix 3-5

In the « healthy » companies group, all firms are correctly classified. But, in the « distressed » firms group, there are two misclassified companies so they are considered as "healthy" when they are actually distressed.

If we summarize, we get the following table:

Table 10: Results of Estimation in the Time and Space

	Initial sample			Control sample		
	1year	**2 years**	**3 years**	**1year**	**2 years**	**3 years**
% of correct classification	100%	99,34 %	96,71 %	95 %	98,33 %	96,67 %
% of classement error	0%	0,66 %	3,29 %	5 %	1,67 %	3,33 %
Error type I	0%	1,32 %	6,58 %	10 %	3,33%	6,67 %
Error type II	0%	0 %	0 %	0 %	0 %	0 %

We notice that the percentage of correct classification, in the initial sample, varies from 100% to 96.71% (a change of 3.29%). It is a result that remains well above those achieved by Ohlson (1980) and Olson et al (2012). Note that Ohlson was the pioneer in the use of logistic regression in the prediction of business distresss. For the control sample that percentage increased from 95% to 96.67%, a negative variation of 1.67%. Overall, the results provided by our model outperforms those presented by Wilcox (1973), Zavgren (1985), Flagg and al (1991), Barniv and Mcdonald (1992), Back and al (1996), Charalambous and al (2000), Charitou and al (2004), Wu and al (2007), Ahn and al (2011), Tserng and al (2011), Serrano-cenca and al (2013) and Wang and al (2014) (Appendix 4 et 5).

The Determinant Power of Variables

The basic equation of the model is:

$$Z = 14{,}057\, R_5 - 131{,}311\, R_6 - 272{,}144\, R_7 + 10{,}482\, R_{15} - 23{,}350\, R_{19} + 66{,}129\, R_{26} + 178{,}682\, R_{28} - 13{,}401\, R_{33} + 87{,}654\, R_{40} - 0{,}501\, R_{61} - 15{,}515\, R_{74} + 52{,}925\, R_{79} + 126{,}426$$

Table 11: The variance of selected ratios

Ratio	Variance
R_5	0,08
R_6	0,63
R_7	0,07
R_{15}	0,647
R_{19}	0,071
R_{26}	0,059
R_{28}	0,645
R_{33}	0,608
R_{40}	0,059
R_{61}	137,076
R_{74}	0,135
R_{79}	0,692

The contribution of the j variable j = $|b_j \sigma_j|$ with

bj : Ratio weighting coefficient of R_j in the function LOGIT

σj :standard deviation of ratio R_j for all companies of initial sample.

The observation of the coefficients of the previous equation does not allow us to evaluate the contribution of each ratio. To do this, we made an adjustment by multiplying the coefficients of these variables by their standard deviation, in order to transform them into a scalar vector. Indeed, since the variance matrix is as follows:

Our objective now is to classify each ratio according to its degree of participation in the discriminatory power of the model to deduce the most determinant ones.

Table 12: The Contribution of the Selected Variables

	Coefficients b_j	standard deviation σ_j	Scalar vector $\lvert b_j\sigma_j \rvert$	classification
R_5	14,088	0,282842712	3,984688133	12
R_6	-131,311	0,793725393	-104,2248751	2
R_7	-272,144	0,264575131	-72,00253448	3
R_{15}	10,482	0,804363102	8,431334036	8
R_{19}	-23,35	0,266458252	-6,221800182	9
R_{26}	66,129	0,242899156	16,06267829	6
R_{28}	178,682	0,80311892	143,5028949	1
R_{33}	-13,401	0,779743548	-10,44934328	7
R_{40}	87,654	0,242899156	21,29108262	5
R_{61}	-0,502	11,70794602	-5,877388902	10
R_{74}	-15,515	0,367423461	-5,700575004	11
R_{79}	52,925	0,831865374	44,0264749	4

From this table, we can conclude that the three most significant variables of distress risk in the model are: R_{28}, $R_{06}$and $R_{07.}$

Thus, we see that the liquidity and solvency have more weight in predicting the distress than profitability and management. This is logical and consistent with reality since the filing of corporate balance sheets is never caused by the deficits, but rather a cash flow problem that is manifested by the inability of the company to meet its obligations or an insolvency problem.

4. Conclusion

Both on the original sample as the control sample, the results provided by the method used are very efficient either in terms of correct percentage of classification or in terms of discriminative power stability over time and space.

The ratios selected and used in the model can cover all aspects of the company: its solvency, its degree of liquidity, financial independence sees its financial structure, the level of payment of its debts, and the degree of ageing its equipment.

Despite the relevance of the results obtained by logistic regression, the presence of several predicting methods allows us a wider choice and therefore more satisfaction and confidence.

Indeed, if the application of models for the same company, gives us the same result (different models apply the same classification) then the creditor or financial analyst make its decision with more confidence. If instead the models give contradictory results, then the decision maker is forced to push more research on this company.

References

Ahn, J., & al. (2011). "Trade finance and the great trade collapse." The American Economic Review 101(3): 298-302.

Albert, A. & E. Lesaffre (1986). "Multiple group logistic discrimination." Computers and mathematics with applications 12(2): 209-224.

Altman, E. I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." The journal of finance 23(4): 589-609.

Altman, E. I., & al. (1977). " Zeta Analyses : A new model to identify bankruptcy risk of corporations." journal of banking and finance 1(1): 29-54.

Altman, E. I., & al. (1994). "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)." journal of banking and finance18(3): 505-529.

Altman, E. I., & al. (2005). "The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications*." The Journal of Business 78(6): 2203-2228.

Anderson, J. & V. Blair (1982). "Penalized maximum likelihood estimation in logistic regression and discrimination." Biometrika 69(1): 123-136.

Anderson, J. A. (1972). "Separate sample logistic discrimination." Biometrika59(1): 19-35.

Aziz, A., & al. (1988). "Bankruptcy prediction-an investigation of cash flow based models [1]." Journal of Management Studies 25(5): 419-437.

Back, B., et al. (1996). "Neural networks and genetic algorithms for bankruptcy predictions." Expert Systems with Applications 11(4): 407-413.

Bardos M. (1984) « Le risque de défaillance d'entreprise » Cahiers économiques et monétaires de la banque de France, p. 35, n° 19, 190 p

Bardos M. (1998a). « Detecting the risk of company failure at the Banque de France", Journal of Banking and Finance, vol. 22, pp. 1405-1419.

Bardos, M. & W. Zhu (1997). "Comparaison de l'analyse discriminante linéaire et des réseaux de neurones: Application à la détection de défaillance d'entreprises." Revue de statistique appliquée 45(4): 65-92.

Bardos, M. (1989). "Trois méthodes d'analyse discriminante." Cahiers économiques et monétaires 33: 151-190.

Barniv, R. & J. B. McDonald (1992). "Identifying financial distress in the insurance industry: A synthesis of m&hodological and empirical issues." Journal of Risk and Insurance 59: 543-543.

Barniv, R. & R. A. Hershbarger (1990). "Classifying financial distress in the life insurance industry." Journal of Risk and Insurance: 110-136.

Beaver, W. H. (1966). "Financial ratios as predictors of failure." Journal of accounting research: 71-111.

Blum, M. (1974), "Failing Company Discriminant Analysis, Journal of Accounting Research", vol. 12, n° 1, pp. 1-25.

Boyacioglu, M. A., & al. (2009). "Predicting bank financial failures using neural n&works, support vector machines and multivariate statistical m&hods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey." Expert Systems with Applications 36(2): 3355-3366.

Burgstahler, D., & al. (1989). "Changes in the probability of bankruptcy and equity value." Journal of Accounting and Economics11(2): 207-224.

Charalambous, C., & al. (2000). "Comparative analysis of artificial neural n&work models: Application in bankruptcy prediction." Annals of operations research 99(1-4): 403-425.

Charitou, A., & al. (2004). "Predicting corporate failure: empirical evidence for the UK." European Accounting Review13(3): 465-497.

Chen, J., & al. (2006). "Financial distress prediction in China." Review of Pacific Basin Financial Mark&s and Policies 9(02): 317-336.

Conan, J. & M. Holder (1979). "Variables explicatives de performances & contrôle de gestion dans les PMI." These d'&at, CERG, Université Paris Dauphine.

Cox, D. (1970). "Analysis of binary data, 1970." Methuen, London: 103-108.

Day, N. E. & D. F. Kerridge (1967). "A general maximum likelihood discriminant." Biometrics 23(2): 313-323.

Deakin, E. B. (1972). "A discriminant analysis of predictors of business failure." Journal of accounting research: 167-179.

Dimitras, A., & al. (1999). "Business failure prediction using rough s&s." European Journal of Operational Research 114(2): 263-280.

Dwyer w. J. (1992) "Changes in the Helping Behaviors of Adult Children as Caregivers" Research on Aging September 1992 14:351-375,

Edmister, R. O. (1972). "An empirical test of financial ratio analysis for small business failure prediction." Journal of Financial and Quantitative Analysis7(02): 1477-1493.

Fitzpatrick, P. J. (1932), "A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms", The Certified Public Accountant October 1932, 598-605

Flagg, J. C., & al. (1991). "Predicting corporate bankruptcy using failing firms." Review of financial Economics1(1): 67-78.

Gloubos, G., Grammatikos, T.1(988). "Success of bankruptcy prediction models in Greece. Studies in Banking and Finance: International business failure prediction models", vol. 7: 37-46

Haehl, J. (1981). « Les techniques de renflouement des entreprises en difficulté », Librairies techniques.

Holder, M., & al. (1984). "Le score de l'entreprise." Nouvelles Editions Fiduciaires, Paris.

Houghton, K. A. & R. Sengupta (1984). "The effect of prior probability disclosure and information set construction on bankers' ability to predict failure." Journal of accounting research: 768-775.

Jones S. & Hensher, D. A. (2007a). "Forecasting corporate bankruptcy: optimizing the performance of the mixed logit model." Abacus 43(3): 241-264.

Jones, S. & D. A. Hensher (2007). "Modelling corporate failure: A multinomial nested logit analysis for unordered outcomes." The British Accounting Review 39(1): 89-107.

Jones, S. and D. A. Hensher (2004). "Predicting firm financial distress: a mixed logit model." The Accounting Review 79(4): 1011-1038.

Kim, M.-J. & D.-K. Kang (2012). "Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction." Expert Systems with Applications 39(10): 9308-9314.

Kira D.S, Doreen D., Ngnyen D. (1997) « An application of artificial neutral networks and statistical methods in qualitative evaluation of small business loans » ASMD Summer, April.

Laitinen, E. K. & T. Laitinen (1998). "Cash management behavior and failure prediction." Journal of Business Finance and Accounting 25(7-8): 893-919.

Laitinen, E. K. & T. Laitinen (2001). "Bankruptcy prediction: application of the Taylor's expansion in logistic regression." International Review of Financial Analysis 9(4): 327-349.

Lau, A. H.-L. (1987). "A five-state financial distress prediction model." Journal of accounting research: 127-138.

Li, H. & J. Sun (2011). "Empirical research of hybridizing principal component analysis with multivariate discriminant analysis and logistic regression for business failure prediction." Expert Systems with Applications 38(5): 6244-6253.

Mahmood, M. A. and E. C. Lawrence. (1987). "A Performance Analysis of Parametric and Nonparametric Discriminant Approaches to Business Decision Making." Decision Sciences 18: 308-326.

Martin, D. (1977). "Early warning of bank failure: A logit regression approach." journal of banking and finance1(3): 249-276.

Mensah, Y. M. (1984). "An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study." Journal of accounting research: 380-395.

Merwin, C. L. (1942), “Financing Small Corporations in Five Manufacturing Industries”, 1926-1936, National Bureau of Economic Research, Financial Research Program III, Studies in Business Financing, 172 p. December, pp. 598-605, 656-662, 727-731.

Min, J. H. & Y.-C. Lee (2005). "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters." Expert Systems with Applications 28(4): 603-614.

Mossman, C. E., & al. (1998). "An empirical comparison of bankruptcy models." Financial Review 33(2): 35-54.

Nam, J. H. & T. Jinn (2000). "Bankruptcy prediction: Evidence from Korean listed companies during the IMF crisis." Journal of International Financial Management and Accounting 11(3): 178-197.

Ohlson, J. A. (1980). "Financial ratios and the probabilistic prediction of bankruptcy." Journal of accounting research: 109-131.

Olson, D. L., & al. (2012). "Comparative analysis of data mining m&hods for bankruptcy prediction." Decision Support Systems52(2): 464-473.

Peel, M. & D. Peel (1987). "Some further empirical evidence on predicting private company failure." Accounting and Business Research 18(69): 57-66.

Platt, H. D. & M. B. Platt (1991). "A note on the use of industry-relative ratios in bankruptcy prediction." journal of banking & finance15(6): 1183-1194.

Ramser, J., Foster, L. (1931), “A Demonstration of Ratio Analysis”, Bulletin n° 40, University of Illinois, Bureau of Business Research, 52 p.

Rose P. S. & Giroux G.A. (1984). “Predicting Corporate Bankruptcy: An Analytical and Empirical Evaluation”, Review of Business and Economic Research, vol. XIX, n° 2, pp. 1-12.

Serrano-Cinca, C. & B. Gutiérrez-Nieto (2013). "Partial least square discriminant analysis for bankruptcy prediction." Decision Support Systems54(3): 1245-1255.

Taffler, R. J. (1982). "Forecasting company failure in the UK using discriminant analysis and financial ratio data." Journal of the Royal Statistical Soci&y. Series A (General): 342-358.

Taffler, R. J. (1983), “The Assessment of Company Solvency and Performance Using a Statistical Model, Accounting and Business Research”, vol. 13, n° 52, pp. 295-307.

Tam, K. Y. & M. Y. Kiang (1992). "Managerial applications of neural n&works: the case of bank failure predictions." Management Science 38(7): 926-947.

Tserng, H. P., & al. (2011). "An enforced support vector machine model for construction contractor default prediction." Automation in Construction20(8): 1242-1249.

Varetto F. (1998). “Genetic Algorithms Applications in the Analysis of Insolvency Risk”, Journal of Banking and Finance, vol. 22, n° 10-11, October, pp. 1421-39.

Wang, Jian Ma, Shanlin Yang (2014)“An improved boosting based on feature selection for corporate bankruptcy prediction”. Expert Systems with Applications, Volume 41, Issue 5, April, Pages 2353-2361

Wilcox, J. W. (1973). "A prediction of business failure using accounting data." Journal of accounting research: 163-179.

Winakor, A. H., Smith, R. F. (1935), “Changes in the Financial Structure of Unsuccessful Industrial Corporations”, University of Illinois, Bureau of Business Research, Bulletin n° 51, p.44

Wu, C.-H., & al. (2007). "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy." Expert Systems with Applications32(2): 397-408.

Yu, Q., & al. (2014). "Bankruptcy prediction using Extreme Learning Machine and financial expertise." Neurocomputing 128: 296-302.

Zavgren, C. V. (1985). "Assessing the vulnerability to failure of American industrial firms: a logistic analysis." Journal of Business Finance and Accounting12(1): 19-45.

Zeitun, R., & al. (2007). "Default probability for the Jordanian companies: A test of cash flow theory." International Research Journal of Finance and Economics, 8, 147-162

Zopounidis, C. (1995). Evaluation du risque de défaillance de l'entreprise: Méthodes et cas d'application, Economic

Appendix 1: Ratios Used in the Study

R1= Financial expenses / Operating income
R2= Cash-flow / Turnoverexcluding taxes
R3= Cash-flow / Total debt
R4= Cash-flow / Equity
R5 = Cash and cash equivalents/ Current liabilities
R6= Permanent capital/ Total Balance Sheet
R7= Current assets / Total Assets
R8= Financial expenses / Turnover
R9= Personnel costs / Added value
R10= Operating income / Added value
R11= Total debt / Equity
R12= Working Capital /Turnover
R13= Added value / Fixed assets
R14= Financial expenses/ Added value
R15= Equity /Total Assets
R16= Working Capital / Cash-flow
R17= Cash and cash equivalents/ Short-term debt
R18= Stocks / Total Assets
R19= Short-term debt / Total Liabilities

R20= Turnovers / Equity
R21= Total Debts/ Total Liabilities
R22= Equity / Permanent equity
R23= Permanent equity / Net fixed assets
R24= Equity / Net fixed assets
R25= Current assets / Current liabilities
R26= Amortization of Capital Assets / Gross Fixed Assets
R27= Added value / Actifs non courants
R28= Working Capital / Total Assets
R29= Added value / Total Assets
R30= Turnover / Total Assets
R31= Cash-Flow / Short-term debt
R32= Short-term debt / Equity
R33= Current assets (excluding stocks)/ Current liabilities
R34= Added value / Turnovers
R35 = Staff costs / Trade accounts payable
R36 = Current assets t – Current assets t-1 / Current assets t-1
R37 = Non-current assetst – Non-current assetst-1 / Non-current assetst-1
R38 = Current assets (excluding stocks) / Turnover
R39 = Current assets (excluding stocks) / Current bank accounts
R40 = Current assets (excluding stocks) / Total Assets
R41 = Current assets (excluding stocks) / Current assets
R42 = Current assets / Turnover
R43 = EBIT(Earnings Before Interest and Taxes) (/ Total Assets
R44 = EBIT / Turnover
R45 = EBIT / Financial expenses
R46 = Net operating result / Equity
R47 = Net operating result / Turnover
R48 = Net operating result / Total Assets
R49 = Working capital requirements / Working capital
R50 = Cash Flow / Total Liabilities
R51 = Cash-Flow / Turnoverexcluding taxes
R52 = Cash-Flow / Non-current liabilities
R53 = Cash Flow / Total Assets
R54 = Staff costs / Gross operating incomes
R55 = Turnover t – Turnover t-1 / Turnover t-1
R56 = Turnover t-1 / Total Assets t-1
R57 = Purchase cost of materials consumed (or purchase cost of production sold) / Average stock material or production
R58 = Receivables/ Total Assets
R59 = Receivables + Stocks / Suppliers
R60 = Non-current liabilities/ Equity

R61 = Medium and long-term debt / Cash flow
R62 = Customer credits Duration
R63 = Credits suppliersDuration
R64 = Gross operating incomes/ Turnover
R65 = Gross operating incomes/ Total Assets
R66 = Gross operating incomes/ Added value
R67 = Working Capital/ Added value
R68 = Non-current liabilities / Non-current assets
R69 = Reserves / Total Assets
R70 = Pre-tax income/ Current liabilities
R71 = Gross operating incomes / Total Assets
R72 = Net Income / Equity
R73 = Net Income / Turnover
R74 = Net Income / Total Liabilities
R75 = Inventory turnover
R76 = Working capital requirements turnover
R77 = Stocks / Total Assets
R78 = Size[Ln (total assets)]
R79 = Total Liabilities / Total Assets
R80 = Growth rate of real assets = (Total Assets t – Total Assets t-1) / Total Assets t-1
R81 = Growth rate of Equity – Growth rate of assets
R82 = Added value t – Added value t-1 / Added value t-1
R83 = Added value / Total Liabilities
R84 = Net fixed assets / Total Assets
R85 = Working Capital/ Cash-flow
R86 = 1 if net income is negative for the past two years, zero otherwise
R87 = 1 iftotal liabilitiesexceedtotal assets, zerootherwise

Appendix 2 : Estimates of Initial Samples

		Correctly classified	Misclassified	Total
One year before distress	Healthy	76 (100 %)	0 (0 %)	76 (100 %)
	Distressed	76 (100 %)	0 (0 %)	76 (100 %)
	Total	152 (100 %)	0 (0 %)	152 (100 %)
		Correctly classified	Misclassified	Total
Two years before distress	Healthy	76 (100 %)	0 (0 %)	76 (100 %)
	Distressed	75 (98. 68 %)	1 (1. 32 %)	76 (100 %)
	Total	151(99. 34 %)	1 (0.66) %	152 (100%)
		Correctly classified	Misclassified	Total
Three years before distress	Healthy	76 (100 %)	0 (0 %)	76 (100 %)
	Distressed	71(93.42 %)	5 (6.58 %)	76 (100 %)
	Total	147 (96.71 %)	5 (3.29 %)	152 (100 %)

Appendix 3 : Estimates of Control Samples

		Correctly classified	Misclassified	Total
One year before distress	Healthy	30 (100 %)	0 (0 %)	30 (100 %)
	Distressed	27 (90 %)	3 (10 %)	30 (100 %)
	Total	57 (95 %)	3 (5 %)	60 (100 %)
		Correctly classified	Misclassified	Total
Two years before distress	Healthy	30 (100 %)	0 (0 %)	30 (100 %)
	Distressed	29 (96.67 %)	1 (3.33 %)	30 (100 %)
	Total	59 (98.33 %)	1 (1.67 %)	60 (100 %)
		Correctly classified	Misclassified	Total
Three years before distress	Healthy	30 (100 %)	0 (0 %)	30 (100 %)
	Distressed	28 (93.33 %)	2 (6.67 %)	30 (100 %)
	Total	58 (96.67 %)	2 (3.33 %)	60 (100 %)

Appendix 3-1 : Estimates of Initial and Control Samples One Year Before Distress : Classification Table[c]

			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	Percentage correct	0	1	Percentage correct
Etape 1	Y	0	76	0	100,0	30	0	100,0
		1	0	76	100,0	3	27	90,0
	Pourcentage global				100,0			95,0

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 1
c. The cut value is ,500

Appendix 3-2: Estimates of initial sample two years before distress : Classification table[c]

			Predicted					
			Selected observations[a]			**Excluded observations[b]**		
			Y		**Percentage correct**	**Y**		**Percentage correct**
	Observations		**0**	**1**		**0**	**1**	
Etape 1	Y	0	76	0	100,0	76	0	100,0
		1	0	76	100,0	1	75	98,7
	Pourcentage global				100,0			99,3

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 1
c. The cut value is ,500

Appendix 3-3: Estimates of Initial Sample Three Years before Distress: Classification Table[c]

			Predicted					
			Selected observations[a]			**Excluded observations[b]**		
			Y		**Percentage correct**	**Y**		**Percentage correct**
	Observations		**0**	**1**		**0**	**1**	
Etape 1	Y	0	76	0	100,0	76	0	100,0
		1	0	76	100,0	5	71	93,4
	Pourcentage global				100,0			96,7

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 1
c. The cut value is ,500

Appendix 3-4: Estimates of Control Sample Two Years before Distress:

Classification Table[c]

				Predicted					
				Selected observations[a]			Excluded observations[b]		
				Y		Percentage correct	Y		Percentage correct
Observations				0	1		0	1	
Etape 1	Y	0		76	0	100,0	30	0	100,0
		1		0	76	100,0	1	29	96,7
	Pourcentage global					100,0			98,3

a. Selected observations Partition EQ 1

b. Excluded observations Partition NE 1

c. The cut value is ,500

Appendix 3-5: Estimates of Control Sample Three Years before Distress:

Classification table[c]

			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y		Percentage correct	Y		Percentage correct
	Observations		0	1		0	1	
Etape 1	Y	0	76	0	100,0	30	0	100,0
		1	0	76	100,0	2	28	93,3
	Pourcentage global				100,0			96,7

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 1
c. The cut value is ,500

Appendix 4:

Authors	Year	Method	Percentage of Correct Classification		
			One year	Two years	Three years
Ahn & al	2011	LOGIT	89,47%		
Aziz & al	1988	LOGIT	91,8%	84,7%	78,6%
Back & al	1996	LOGIT	96,49%	71,6%	74,3%
Barniv & Hershbarger	1990	LOGIT	91,1%	85,7%	
Barniv& MCdonald	1992	LOGIT	83,7%	80%	71,9%
Boyacioglu & al	2009	LOGIT	81,81%		
Charalambous & al	2000	LOGIT	82,3%	74,5%	69,8%
Charitou & al	2004	LOGIT	80,95%	73,81%	72,92%
Dimitras & al	1999	LOGIT	90%	82,5%	78,75%
Min & Lee	2005	LOGIT	79,31%		
Kira & al	1997	LOGIT	95,5%		
Laitinen & Laitinen	1998	LOGIT	86,6%	68,3%	
Laitinen & Laitinen	2001	LOGIT	74,7%	65,3%	
Lau	1987	LOGIT	80%	79%	85%
Min & al	2006	LOGIT	78,13%		
Nam & Jinn	2000	LOGIT	84,4%	76,1%	76,1%
Ohlson	1980	LOGIT	82,84%	86%	
Olson & al	2012	LOGIT	79,8%		
Serrano-canca & al	2013	LOGIT	95,36%		
Tserng & al	2011	LOGIT	73,61%		

Wang & al	2014	LOGIT	73,9%		
Wilcox	1973	LOGIT	94%	90%	88%
Wu & al	2007	LOGIT	92,05%	89,78%	80,68%
Zavgren	1985	LOGIT	96%	96%	96%
Chen & al	2006	LOGIT	84,68%		

Appendix 5

Authors	**Year**	**Method**	**Percentage of correct classification**					
			Distressed			**Healthy**		
			1 year	**2 years**	**3 years**	**1 year**	**2 years**	**3 years**
Aziz& al	1988	LOGIT	85,7%	85,7%	79,6%	98%	83 ,7%	77,6%
Back& al	1996	LOGIT	86,49%	72,97%	83,78%	86,49%	70,27%	64,86%
Barniv& Hershbarger	1990	LOGIT	89,3%	89,3%		89,3%	85,7%	
Barniv& MCdonald	1992	LOGIT	80%	75,4%	61,1%	87,1%	84,2%	81,2%
Dimitras & al	1999	LOGIT	92,5%	77,5%	77,5%	87,5%	87,5%	80%
Dwyer	1992	LOGIT	90%	97%	80%	62%	57%	43%
Flagg& al	1991	LOGIT	73%			97%		
Globos & Grammatikos	1988	LOGIT	66,7%	60,9%	50%	85,7%	82,6%	78,6%
Jiang	1993	LOGIT	76%	78%	84%	82%	71%	74%
Laitinen & Laitinen	1998	LOGIT	87,8%	65,9%		85,4%	61,7%	
Laitinen & Laitinen	2000	LOGIT	74,1%	61,2%		75,3%	69,4%	
Mahmood & Lawrence	1987	LOGIT	52,4%	45,2%	31%	92,7%	94,7%	91,7%
Martin	1977	LOGIT	91,3%	83,3%	92,3%	91,1%	90,3%	87,4%
Mossman & al	1998	LOGIT	80%			70%		
Ohlson	1980	LOGIT	87,6%			82,6%		
Peel	1987	LOGIT	67%	75%	92%	79%	83%	88%
Philipe Du Jardin	2007	LOGIT	89,56%			90,44%		90%
Platt & Platt	1991	LOGIT	85%			88%		
Suominen	1988	LOGIT	71%	57%	33%	86%	84%	89%
Tam & Kiang	1992	LOGIT	68%	85%		95%	100%	