ECONOMIC RESEARCH FORUM

منتدى البحوث الاقتصادية

Working Paper Series

ERF 25TH ANNIVERSARY

PREDICTING FINANCIAL DISTRESS OF COMPANIES: COMPARISON BETWEEN MULTIVARIATE DISCRIMINANT ANALYSIS AND MULTILAYER PERCEPTRON FOR TUNISIAN CASE

Fayçal Mraihi and Inane Kanzari

Working Paper No. 1328

PREDICTING FINANCIAL DISTRESS OF COMPANIES: COMPARISON BETWEEN MULTIVARIATE DISCRIMINANT ANALYSIS AND MULTILAYER PERCEPTRON FOR TUNISIAN CASE

Fayçal Mraihi and Inane Kanzari[1]

Working Paper No. 1328

July 2019

Send correspondence to:
Fayçal Mraihi
Higher School of Economics and Business Sciences of Tunis
mraihifaycal@gmail.com

[1] Higher School of Economics and Business Sciences of Tunis, Tunisia. Email: inene.kanzari@gmail.com.

First published in 2019 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract

In this study, we try to develop a model for predicting corporate default based on a multivariate discriminant analysis (ADM) and a multilayer perceptron (MLP). The two models are applied to the Tunisian cases. Our sample consists of 212 companies in the various industries (106 'healthy' companies and 106 "distressed" companies) over the period 2005-2010. The results of the use of a battery of 87 ratios showed that 16 ratios can build the model and that liquidity and solvency have more weight than profitability and management in predicting the distress. Despite the slight superiority of the results provided by the MLP model, on the control sample, the results provided by the two models are good either in terms of correct percentage of classification or in terms of stability of discriminating power over time and space.

Keywords: distressed firms, forecasting model, multivariate discriminant analysis, multilayer perceptron.

JEL Classifications: 17 - 33 – 34.

1. Introduction

The diagnosis of default risk has experienced significant development using both classical statistical methods as methods from artificial intelligence that analyze the financial situation from a given set of ratios. In the present work, we will estimate and compare the discriminating power of the Multivariate Discriminant Analysis (MDA) and the multilayer perceptron (MLP) models. The first is a classic statistical method, while the second belongs to the methods from artificial intelligence.

The principle is relatively simple. With the financial characteristics described using ratios and a sample of companies that cover both "healthy" companies and firms "failing", the objective is to determine the best combination of ratios to differentiate the two business groups. Based on this combination, we will estimate the percentage of correct classification of each method. To achieve this goal, this article will address in the first section, the methodology through the constitution samples, presentation and justification of the two selected models. The estimate of the discriminating power of the two models in the time and space will be of the second section. The third section will compare the results given by the two methods.

2. The methodology

In this work, we will use the Multivariate Discriminant Analysis and Multilayer perceptron (MLP) for the purpose of forecasting corporate failures, and then test their validity in time and in space. However, it is above all, to address the composition of samples, the selection of variables, presenting the models and demonstrate their usefulness.

2.1. The constitution of samples

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

- 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, Metallurgical and Electric	11	13
Services (hotel)	8	9
Chemical industries	4	5
Total	**106**	**106**

2.2. The choice of default indicators

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. In order that our work be as exhaustive as possible, we chose 87 ratios contained in the works of Ramser and Foster (1931), Fitzpatrick (1932), Winakor and Smith (1935), Merwin (1942), Beaver (1966), Altman (1968), Deakin (1972) Edmister (1972), Blum (1974); Altman et al (1977), Taffler (1983) and Zmijewski (1984).

2.3. Overview and principle of the Multivariate Discriminant Analysis model

2.3.1. Literature review

The objective of the multivariate discriminant analysis is to compare the predictive power of the different ratios. This predictive power is measured by the capacity of the selected model to separating healthy firms from failing firms.

Unlike univariate analysis, the assignment of a company to one of the two classes is not based on the value of a single ratio but on the basis of a combination of several ratios or indicators. In effect, Altman (1968) asserts that a one-dimensional analysis is not able to account for the complexity of the failure process.

The objective is to determine a function called Z-score, which is none other than the linear combination of explanatory variables retained. This combination must be able to distinguish at best the two groups through the identification of the level of risk of each company. The linear discriminant analysis requires the observance of two assumptions that of the multi-normality and that of the homoscedasticity. The first assumes that the accounting variables used follow a normal law; the second requires the equality of matrices variance-covariance for the two categories of failing and healthy firms. To circumvent the problem of homoscedasticity, some authors have made use of quadratic discriminant analyzes, which require only the hypothesis of multi-normality of ratios (Lachenbruch and al, 1973; Marks and Dunn, 1974 ; Rose and Giroux, 1984). Only we found that they are always less efficient than the linear analysis and this mainly for two reasons. First, the absence of multi-normality ratios is much more harmful to the effectiveness of the quadratic analysis than to those of the linear analysis (Lachenbruch, 1975); secondly, even in the case of non-respect of the hypothesis of multi-normality, quadratic discriminant analysis is efficient only if it is applied to a sample of large size.

2.3.2. MDA model principle

Developed by Altman (1968), multivariate discriminant analysis (MDA) assumes the existence of two groups of firms each with its own indicators of its financial situation. For these two groups then we can determine a discriminant function that is sharing in the better the set of firms in two separate groups. This discriminant function is a linear combination of the most relevant indicators, to differentiate the two groups we associate a score Zj has each company j.

$$Z_j = \alpha_0 + \alpha_1 x_{1j} + \alpha_2 x_{2j} + \cdots + \alpha_n x_{nj} + c$$

Avec:

x_{nj} : The value taken by the indicator x_n of the enterprise j

α_i : The numerical adjustment coefficients.

c : A constant

The classification in one or the other of the groups is done by comparing the value of the score Zj with a critical value Z*. We must however, during the drafting of the discriminant function maximize the intergroup variance and minimize the intra-group variance.

During this discrimination, there may be two types of errors:
- The error of first species: classify a failing firm with sound.
- The error of second species: classify a healthy firm with failing.

The cost associated with the error of first species is very different from the cost associated with the error of second species. In effect, the first cost is the one that will bear a creditor in the event of failure of its debtor. While the second cost corresponds to the opportunity cost, that is to say, the gap between the pay that a creditor might have been able to collect on the loan refused and the rate of return offered by the use of these funds.

The proportion of correct classification allows you to judge the quality of the discriminant function.

2.4. Overview and principle of the MLP model

2.4.1. Literature review

Warren McCulloch and Walter Pitts (1943) were the pioneers in the field of neural networks by presenting the "formal neuron" as the first attempt to imitate the functioning of the human brain. In 1949, Hebb presents the first rule of learning neural networks, a move which allowed, later, Rosenblatt (1958) to propose the first algorithm of learning making the adjustment of the parameters of a neuron possible.

After publishing their book "perceptrons" in which Papert (1968) shows the limits of monolayer neural networks, connectionism resumed in the 1980s after a long period of hibernation. Indeed, the work of Hopfield (1982), who proposed associative neural networks, induced an interesting renaissance of neural networks.

Rumelhart, Hinton, and Williams (1986) published their work on the error-retroagitation algorithm that optimizes the parameters of a multi-layered neural network. From this date, research on neural networks has expanded greatly and has been integrated in all areas.

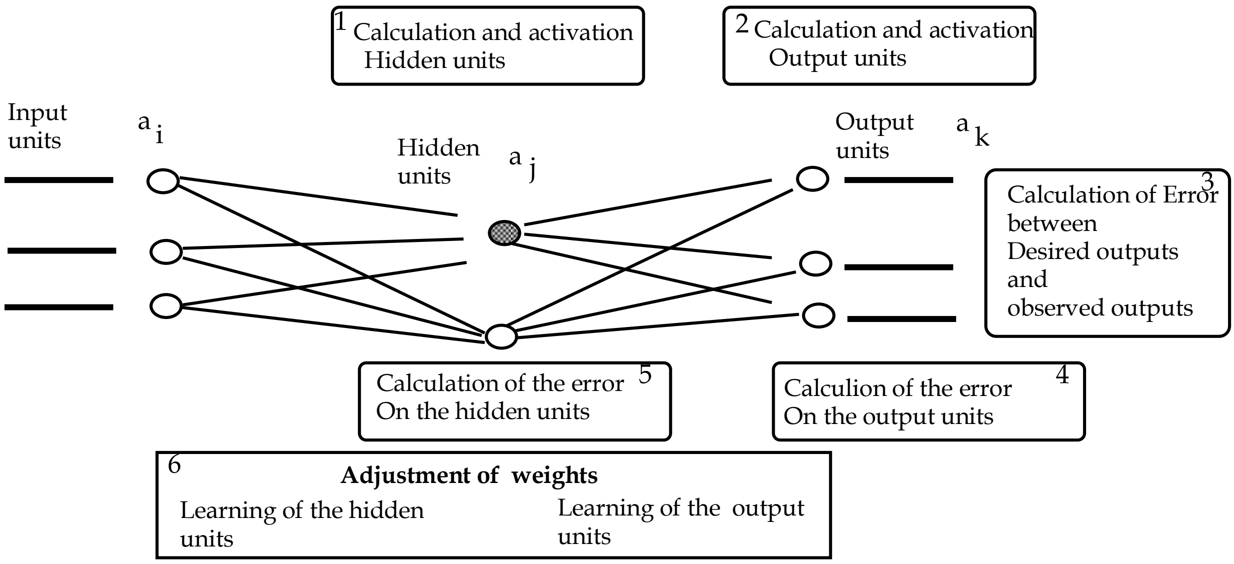
The use of artificial neural networks (ANNs) in failure prediction dates back to 1990. Indeed Odom et al. (1990) were the pioneers in the field.

According to Odom et al. (1990), Raghupathi (1991), Salchenberger (1992), Tam (1997) and Altman (1994), the multilayer perceptron with gradient retro-extension algorithm (RPG) learning remains the reference in failure anticipation.

The use of a learning algorithm other than the RPG technique in the context of the implementation of a multilayer perceptron stems from one of the limits of this type of network, namely its blocking on the local minima.

In the area of failure prediction the multilayer perceptron (PMC) represents the reference network [Poddig (1995)]. However, there are other networks of artificial neurons other than the PMC such as the radial base function networks (FBR) and self-organizing maps of Kohonen. The operating principle of the multilayer perceptron is as follows:

Figure 1. The multilayer perceptron: learning by backpropagation of the error



MLP model principle

The neural networks allowing for estimating a function f such that $f: x \rightarrow y$

with $x^T = [x_1, x_2 \ldots\ldots x_E] \in IR^E$ if $y \in IR^s$ are regressions.

If $y \in [c_1, c_2 \ldots\ldots c_s]$ then it is classification. In this case, we should have as many output neurons as classes.

The desired outputs are of the form:

$$y^{d^T} = [0,0,1\ldots\ldots 0,1\ldots\ldots,0]$$

When estimating the function f we must identify the connection weights between neurons. Now let us recall before all the principle of the formal neuron [Me Cultoch and Pitts, 1943].

Let E inputs x_i et y outputs. The sum of inputs x_i weighted by w_i is equal to α

With : $$\alpha = \sum_{i=1}^{E} w_i x_i + b = \sum_{i=0}^{E} w_i x_i$$ avec $x_0 = 1$

Let φ an activation function that can be linear where we have:

$$y = \varphi(x) = \varphi\left(\sum_{i=0}^{E} w_i x_i\right)$$

If φ is linear the separator is a hyperplane.

If φ is not linear the separator is a hyperbola of dimension E.

We distinguish different activation functions that determine the activation threshold of a neuron.

Identity function: $\varphi(x) = x$

Heaviside function: $\varphi(x) = 0$ si $x < 0$ et $\varphi(x) = 1$ si $x \geq 0$

Sigmoid function: $$\varphi(x) = \frac{1}{1+e^x}$$

Hyperbolic tangent function: $$\varphi(x) = \frac{e^x - e^x}{e^x + e^{-x}} = \frac{e^{2x} - 1}{2^{2x} + 1}$$

Average function (Gaausian = normal)

The available data are as follows:

We have a base of N couples $\{x(n); y^d(n)\}$.

With :

$x(n)$: observations on the independent variables

$y^d(n)$: the desired outputs in value for example n

$X \in IR^{ExN}$

$$X = \{X(n)\} = \left\{ \begin{bmatrix} x_1(n) \\ x_2(n) \\ \\ x_E(n) \end{bmatrix} \right\} = \begin{bmatrix} x_1(1)\ x_1(2) \ldots\ldots x_1(N) \\ x_1(2) \qquad\qquad x_2(N) \\ \\ x_E(1) \qquad\qquad x_E(N) \end{bmatrix}$$

$$y^d \in \{y^d(n)\} = \left\{ \begin{bmatrix} y_1^d(n) \\ \\ \\ y_s^d(n) \end{bmatrix} \right\} = \begin{bmatrix} y_1^d(1)\ y_1^d(2) \ldots\ldots y_1^d(N) \\ y_2^d(1) \qquad\qquad \\ \\ y_s^d(1) \qquad\qquad y_s^d(N) \end{bmatrix}$$

With E : the number of input variables
S : the number of neurons in the output layer

We will assume a multilayer network with inputs (E inputs), a hidden layer with j neurons and an output layer of S neurons.

Are: $W^1 \in IR^{J\,x\,E}$ the matrix of connection weights between inputs $\left(X(n) \in IR^E\right)$ and the J neurons of the hidden layer.

$W^2 \in IR^{S\,x\,J}$ the matrix of connection weights between the J neurons of the hidden layer and the S neurons of the output layer.
So:

$$W^1 = \{W_{ji}\} = \begin{bmatrix} W_{11}\ W_{12} \qquad W_{1E} \\ W_{21} \\ \\ W_{J1} \qquad\qquad W_{JE} \end{bmatrix} \tag{1}$$

$$W^2 = \{W_{sj}\} = \begin{bmatrix} W_{11}\ W_{12} \qquad W_{1J} \\ W_{21} \\ \\ W_{S1} \qquad\qquad W_{SJ} \end{bmatrix} \tag{2}$$

x_1		1		1		
x_2		2		2		
	➔		➔		➔	$y(n) \in IR^S$
x_E		J		S		
$X(N) \in IR^E$		$W^1 = \{W_{ji}\}$		$W^2 = \{W_{sj}\}$		
Inputs		hidden layer composed of J neurons		output layer composed of S neurons		

Let φ^1 and φ^2 two nonlinear activation functions.
φ^1 of the sigmoid type relating to the connections of the hidden layer and φ^2 of the soft max type relative to the connections of the output layer.
Let $Z(n) \in IR^J$ an intermediate variable.

α_j^1 : the weighted sum of the connections between all the E inputs and the jth neurons in the hidden layer.

α_s^2 : the weighted sum of the connections between the J hidden neurons and the sth output neuron.

We then:

$$\alpha_j^1 = \sum_e W_{je} X_e \quad \text{et} \quad Z_j = \varphi^1 \left[\left(\alpha_j^1 \right) \right] \qquad (3)$$

$$Z_j = \varphi^1 \left[\sum_e W_{je} X_e \right] \qquad (4)$$

Once we have finished with the modeling of the passage from the input neurons to the hidden neurons, we will approach the second half of the process, which relates to the passage of the hidden neurons to the output neurons. Indeed:

$$\alpha_s^2 = \sum_j W_{sj} Z_j \quad \text{et} \quad y_s = \varphi^2 \left[\left(\alpha_s^2 \right) \right] \qquad (5)$$

$$= \varphi^2 \left[\sum_j W_{sj} Z_j \right] \qquad (6)$$

then

$$y_s = \varphi^2 \left[\sum_j W_{sj} \varphi^1 \left(\sum_e W_{je} X_e \right) \right] \qquad (7)$$

3. Estimation of the Multivariate Discriminant Analysis model parameters

This part will be devoted to the estimation of the discriminating power of A.D.M. Both in time through its application on the initial sample two and three years before the failure and in space by applying it on the three control sub-samples.

First, we will use 87 explanatory variables (see Appendix 1). To determine the weighting coefficient of each exogenous variable in our discriminant function, we used a software frequently applied in the analysis of the data, the software S.P.S.S.

Applying this software to our sample, we obtained the following results: (see appendix 2)
If we take into account the significance (see Appendix 2) and the redundancy (variance-covariance matrix) of the explanatory variables of the model for a degree of significance of 1%, we must retain only the 16 ratios that will constitute the explanatory variables of the model to be estimated. The estimate by A.D.M. gives us the following results:

Table 2. Eigen values

Fonction	Eigen values	% of variance	% cumulated	Canonical correlation
1	8,669[a]	100,0	100,0	,947

R16	-,023
R19	-3,389
R26	1,855
R33	-,927
R40	8,230
R58	-2,510
R61	-,027
R73	-,631
R78	-,210
R79	8,369
R83	-,493
R84	-4,234
R85	,024
(Constante)	,225

a. The first 1 canonical discriminant functions were used for the analysis.
Non-standardized coefficients

Table 3. Coefficients of canonical discriminant functions

	Fonction
	1
R6	2,891
R7	-9,988
R15	5,942

The last 16 ratios will represent the explanatory variables of our final model:
Z = 2,8907 R6 - 9,9883 R7 + 5,9415 R15 - 0,0225 R16 - 3,3888 R19 + 1,8554 R26 - 0,9273 R33 + 8,23 R40 - 2,5098 R58 - 0,0274 R61 - 0,6312 R73 - 0,2096 R78 + 8,3685 R79 - 0,4930 R83 - 4,2335 R84 + 0,0242 R85 + 0,2247
Avec:

Table 4. The Ratios Retained by the M.D.A Method

Ratios	Formulas
R_6	Permanent Capital / Total Balance Sheet
R_7	Current assets / Total assets
R_{15}	Equity / Total assets
R_{16}	Working capital / Cash flow from operations
R_{19}	Short-Term Debt / Total Liabilities
R_{26}	Amortization of Capital Assets / Gross Fixed Assets
R_{33}	current assets (excluding stocks) / current liabilities
R_{40}	current assets (excluding stock) / Total assets
R_{58}	receivables / Total assets
R_{61}	Medium and long-term debt / Cash flow
R_{73}	Net income / Turnover
R_{78}	Size Ln (Total assets)
R_{79}	Total Liabilities / Total Assets
R_{83}	Value Added / Total Liabilities
R_{84}	Total Fixed Asset / Total assets
R_{85}	Working capital / Cash-flow

In the prediction equation retained by the discriminant analysis, we note the presence of several ratios that have been selected as explanatory variables in previous studies.

Table 5. The Presence of Several Explanatory Ratios in Previous Studies

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

The presence of these ratios in the models makes it possible to cover all aspects of the company, its solvency, its liquidity level, its financial autonomy, its financial structure, the degree of maturity of these debts and the degree of aging of these equipment.

The global significance test used in the MDA regression is the chi-square with k degrees of freedom (K is the number of explanatory variables in our case k = 16). If the critical probability is lower than the level of significance we have set, we can consider that the model is globally significant. In our model, the likelihood ratio statistic (chi-square) is equal to 322,187, the associated critical probability is zero. The model is thus globally very significant, there is indeed a relationship between the explanatory variables and the variable to be explained.

Table 6. Lambda of Wilks

Test of the function (s)	Wilks' Lambda	Chi-square	ddl	Signification
1	,103	322,187	16	,000

Once the overall significance of the chosen model is demonstrated, our work now consists in verifying the discriminatory capacity and the stability of the results presented by the A.D.M. And S.V.M. Both in time using the initial samples one year, two years and three years before the failure than in the space using the control samples.

4. Estimation and validation of the discriminatory power of the MDA model in time and space

4.1. Estimation of the model discriminatory power one year before distress

The estimation of the MDA 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".

Table 7. Estimates of initial sample one years before distress

Classification table[b]

			Predicted		
			Selected observations[a]		
			Y		
	Observations		0	1	Percentage correct
Stape 1	Y	0	76	0	100,0
		1	0	76	100,0
	global Percentage				100,0

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 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).

4.2. Validation of the model discriminatory power over time

4.2.1. For the same companies two years before distress

The validation of model on exercises that come two years before distress gives the results in in the following table.

Table 8. Estimates of initial sample two years before distress

Classification table[c]

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

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

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 3.3% from 100% to 96.7%, the error type I increases from 0 to 6.58%, while the error type II remains zero.

4.2.2 For the same companies three years before distress

By distancing yet the period between the date of the estimates and the date of the failure of a period of time for an additional year, the application of the multivariate discriminant analysis provides the results presented in the table 9.

Table 9. Estimates of initial sample three years before distress

Classification table[c]								
			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	correct percentage	0	1	
Stape 1	Y	0	76	0	100,0	75	1	98,7
		1	0	76	100,0	6	70	92,1
	global Percentage				100,0			95,4

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

In passing from one year to three years before the failure, the method loses more of its accuracy. In fact, the percentage of correct classification increased from 100 per cent to 95.4%. The error of first species (error type I) jumped from 0 per cent to 7.89 %. In effect, the method class 6 companies as "sound", then they really are "faulty".

The error of second species increased from 0 % to 1.32 %. Actually, the discriminant analysis multivariate range a single company in the group of "failed" when it is really "healthy".

Table 10. Results of estimation in the time

	1 year before distress	2 years before distress	3 years before distress
% of correct classification	100 %	96. 71 %	95.4 %
% of classification error	0 %	3. 29 %	4.6 %
% of error type I	0 %	6. 58 %	7.89 %
% of error type II	0 %	0 %	1.32 %

Indeed, we notice that for the model used, the percentage of the error Type I varied only by 7.89% between the first and third years before distress. Furthermore, we find that the correct percentage of classification decreased only by 4.6% (it goes from 100% to 95.4%).

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 ≤ 1.32%.

4.3. 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 list 30 firms. The model will be tested on companies other than those that were originated. The application of our MDA model on these samples gives us the estimates presented in the table 11.

Table 11. 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	Correct percentage	0	1	Correct percentage
Stape 1	Y	0	76	0	100,0	29	1	96,7
		1	0	76	100,0	3	27	90,0
	global Percentage				100,0			93,3

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

In the « healthy » companies group, we find that the model classifies only one firm in the « distressed » group when she is « healthy ». In the « distressed » group, there are also three misclassified firms so they are considered by the model « healthy » when they are actually distressed.

This model has a remarkable accuracy by classifying 93.34% of the control sample correctly. The error Type I is around 10% while the error type II is 3.33%.

Studying companies' exercises of control sample in case of two years before distress, we get the results announced at the table 12.

Table 12. Estimates of control sample two years before distress

Classification table[c]

			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	Percentage correct	0	1	Percentage correct
Stape 1	Y	0	76	0	100,0	29	1	96,7
		1	0	76	100,0	2	28	93,3
	global Percentage				100,0			95

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

In the « healthy » companies group, we find that the model classifies 29 firms correctly so we conclude an error type II equal to 3.33%. While in the group of distressed companies, there is two firm misclassified, giving us an error Type I of about 6.67%.

The increase of the efficiency of the MDA function, in this validation test (it passed from 93.3% to 95%), 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 the following table.

Table 13. Estimates of control sample three years before distress

			Classification table[c]					
			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	Percentage correct	0	1	Percentage correct
Stape 1	Y	0	76	0	100,0	27	3	90,0
		1	0	76	100,0	2	28	93,3
	global Percentage				100,0			91,7

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

There are five misclassified companies. Two are considered as "healthy" when they are actually distressed and three are considered as distressed when they are really "healthy".

If we summarize, we get the following table:

Table 14. 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 %	96. 71 %	95.4 %	93,34%	95%	91,67%
% of classification error	0 %	3. 29 %	4.6 %	6,66%	5%	8,33%
Error type I	0 %	6. 58 %	7.89 %	10%	6,67%	6,67%
Error type II	0 %	0 %	1.32 %	3,33%	3,33%	10%

In effect, from the summary table above, using the initial sample for a maturity of one year prior to the failure, our model presents a rate of correct classification of 100 %. Such a result is consistent with that found by Frydman, Altman & Kao (1985) and Izán (1984) but remains well above those achieved by Yu et al (2014), Serrano-canca and al (2013), Myoung-Jong Kim, Dae-Ki Kang (2012) and Rafiei and al (2011). The same for the coming years two to three years before the failure, the method presents rates of correct classification, respectively, of the order of 96.71 per cent and 95.4 per cent largely superior to those made by Blum (1974), Altman (1968), Moyer (1977), Altman et al (1977), Frydman et al (1985), Dimitras and al (1987), Altman et al (1994), Back and al (1996), Charitou and al (2004) and Wu et al (2007) (see table 15).

By applying our model on a sample test, its percentage of correct classification remains beyond 90 %, outperformance as well the results obtained by Deakin (1972), Taffler (1982), Rose and Giroux (1984), Flagg and al (1991) and Brabazon and Keenan (2004) (see table 15 and 16).

Table 15. The results of literature review

Authors	Year	Method	Percentage of correct classification		
			One year	**Two years**	**Three years**
Altman	1968	MDA	95%	72%	48%
Altman and al	1977	MDA	92,8%	89%	83,5%
Altman and al	1994	MDA	93,2%	88,2%	91,1%
Altman and al	1985	MDA	100%	75%	50%
BACK and al	1996	MDA	85,14%	78,38%	72,97%
Blum	1974	MDA	87%	79%	72%
Boyacioglu and al	2009	MDA	68,18%		
Brabazon and KEENAN	2004	MDA	80,67%	72%	
Brabazon and Keenan	2004	MDA	76% c	69,33% c	64,67% c
CALIA and GANUCI	1997	MDA	60,9%		
Charitou and al	2004	MDA	82,5%	62,5%	68%
Dambolena and Khoury	1980	MDA	91,2%	84,8%	82,6%
DEAKIN	1972	MDA	87% (c)	82% (c)	
DEAKIN	1972	MDA	91,2%	84,8%	
Dimitras and al	1999	MDA	90%	81,3%	77,5%
Gombola and al	1987	MDA	89%	70%	78%
Izan	1984	MDA	100%	70%	40%
Jae H. Min, Young-Chan Lee	2005	MDA	78,81%		
KIRA and al	1997	MDA	93,3%		
Levitan and al	1985	MDA	95%	91%	83%
Moyer	1977	MDA	84,1%	76,6%	68,2%
Myoung-Jong Kim, Dae-Ki Kang	2012	MDA	71,02%		
Rafiei and al	2011	MDA	80,6%		
Serrano-canca and al	2013	MDA	91,79%		
Sharma and Mahajan	1980	MDA	91,7%	78,3%	73,9%
Weinrich	1978	MDA	89%	84,3%	78,1%
WILSON and SHARDA	1994	MDA	88,65%		
Wu and al	2007	MDA	87,5%	85,22%	75%
Yi-Chung Hu and Fang-MeiTseng	2005	MDA	77,94%		
Yu and al	2014	MDA	86,5%		

Table 16. The results of literature review

Authors	Year	Method	Percentage of correct classification					
			Distressed			healthy		
			1year	2 years	3 years	1year	2 years	3 years
ALTMAN	1968	MDA	93,39	71,2%	48,3%	97%	93,9%	
ALTMAN	1983	MDA	94,2%			92,4%		
ALTMAN and al	1994	MDA	92,8%	90,3%		96,5%	86,4%	
BACK and al	1996	MDA	86,49%	75,68%	83,78%	83,78%	81,08%	62,16%
Brabazon and Keenan	2004	MDA	82,7%	74,7%	65,3%	78,7%	69,3%	66,7%
Cadden	1991	MDA	80%	60%	60%	90%	80%	70%
Dambolena and Khoury	1980	MDA	83%	83%	78%	100%	87,%	87%
Deakin	1972	MDA	77%	96%	94%	82%	92%	82%
Diamond J.R	1976	MDA	97,3%	87,8%	80%	90,7%	85,3%	80%
Dimitras and al	1999	MDA	87,5%	75%	67,5%	92,5%	87,5%	87,5%
Dwyer	1992	MDA	76%	70%	43%	57%	54%	57%
Gloubos and Grammatikos	1988	MDA	66,7%	60,9%	64,3%	66,7%	82,6%	85,7%
Laitinen	1991	MDA	90%	72,5%	57,5%	87,5%	65%	52,5%
Moyer	1977	MDA	95%	80%	70%	82%	86%	73%
ROSE and GIROUX	1984	MDA	84,6% (1c)	87,5% (2c)		97,1% (1c)	96,2% (c)	
TAFFLER	1982	MDA	87,9% (1c)	48% (c)		100% (1c)		

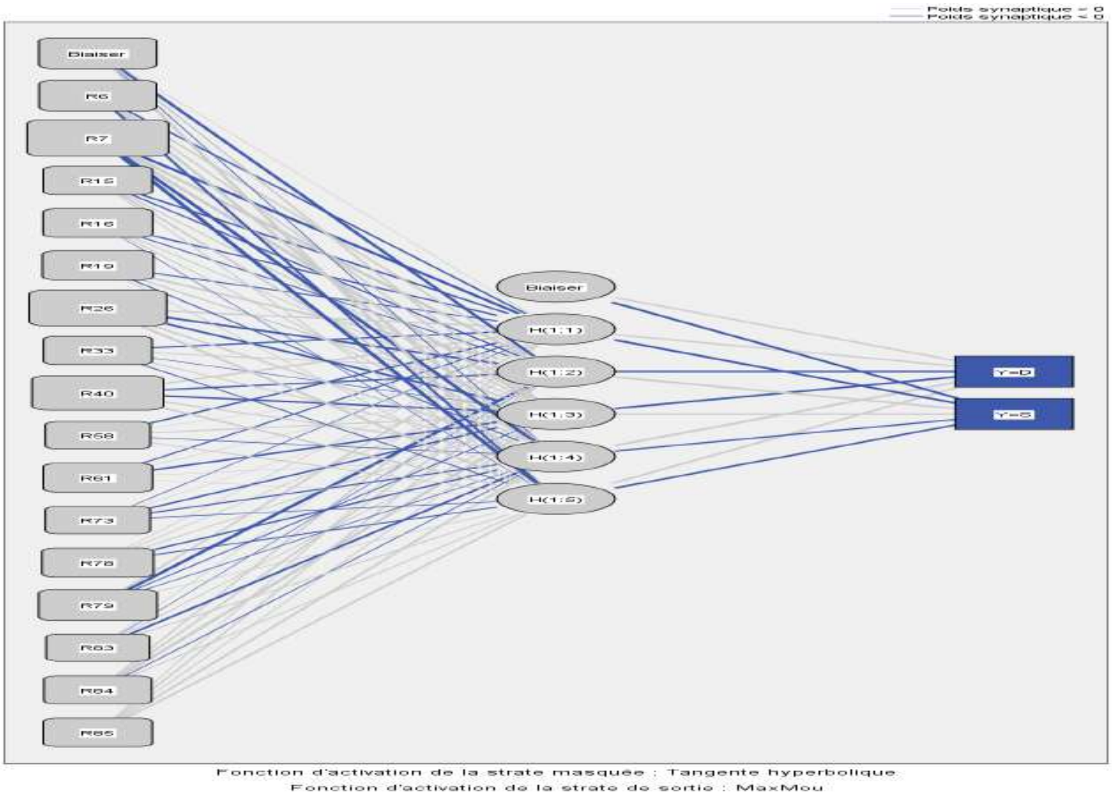
5. Estimation of the MLP model discriminatory power

To estimate and compare the discriminatory ability of the MLP model with that of the multivariate discriminant analysis method, we will use the sixteen independent variables used earlier. Then, the purpose of this section is twofold: estimate the discriminatory ability of the MLP method and check if it is able to maintain its discriminatory power in time and in space. Before presenting the results of the estimation, we must pay particular attention to two levels.

The first level is the array of information on the network and its architecture that allow us to check, first, that the specifications are correct and then to extract the specificities of the network summarized in the following points:

- the number of units in the input stratum corresponds to the number of independent variables (see appendix 3).
- likewise, a unit of a specific result is created for each class of healthy and failing companies for a total of 2 units in the income or output stratum.
- the automatic selection of the architecture chose a single hidden layer consisting of 5 units in addition to a biased one. Indeed, the architecture of the multilayer perceptron retained confirms

Figure 2. Multilayer perceptron architecture



- the activation function used for the hidden layer is of the hyperbolic Tangent type whereas it is of the MaxMou type for the output layer (see appendix 3).

The second level is the model summary (see appendix 4) which displays information on the results of the learning of the final network and its application to the processed sample. Indeed,

- A cross-entropy error occurs because the output layer uses the MaxMou activation function. This is the error function that the network tries to minimize during learning.
- The percentage of incorrect forecasts comes from the league table and will be discussed later in this section.
- The stopping criterion is the indicator that must be imposed on the algorithm, i.e. the criterion which, once satisfied the algorithm, stops and puts an end to all calculations. The stopping criterion can be either a number of variables or iterations, or the absence of a significant variation of an expected result after adding or removing a variable or still obtaining a satisfactory predictive capacity threshold. In our case, learning stopped when the error converged.

5.1. Estimation of the MLP model discriminatory power one year before distress

The MLP method, applied to an original sample "a year before the distress", allows for correctly classifying 100% (152/152) of companies.

Table 17. Estimates of initial sample one years before distress

Classification table[b]

			Predicted		
			Selected observations[a]		
			Y		
	Observations		0	1	Percentage correct
Stape 1	Y	0	76	0	100,0
		1	0	76	100,0
	global Percentage				100,0

a. Selected observations Partition EQ 1
b. Excluded observations Partition NE 1

5.2. Validation of the discriminatory power of the method in time

5.2.1. For the same business two years before failure

In this section, we will keep the same companies, but we will use the data two years before distress.

The results show a slight reduction in accuracy of the method. Indeed, the correct classification percentage moved from 100% to 98.68% due to misclassification of two distressed firms by type I error of about 2.63% (2/76). Type II error remained always zero.

Table 18. Estimates of initial sample two years before distress

Classification table[c]

			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	correct Percentage	0	1	Correct percentage
Stape 1	Y	0	76	0	100,0	76	0	100,0
		1	0	76	100,0	2	74	97,37
	global Percentage				100,0			98,68

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

5.2.2. for the same business three years before failure

Three years from the date of the coming failure, the MLP method yielded the following results:

Table 19. Estimates of initial sample three years before distress

Classification table[c]

			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	correct percentage	0	1	
Stape 1	Y	0	76	0	100,0	73	3	96,05
		1	0	76	100,0	3	73	96,05
	global Percentage				100,0			96,05

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

Table 20. The results provided by the model over time

	1 year before distress	2 years before distress	3 years before distress
% of correct classification	100 %	98.68 %	96.05 %
% of classification error	0 %	1. 32 %	3.95 %
% of error type I	0 %	2. 63 %	3.95 %
% of error type II	0 %	0 %	3.95 %

When evaluating the predictive ability of the model, we found a correct classification percentage varying between 100% (152/152) and 96.05% (146/152), respectively for year one and three years before distress. Similarly, type I and II errors have increased from 0% (0/76) to 3.95% (3/76) during the same period. Despite its application to data located three years before the advent of the distress, the MLP method keeps a decent percentage of correct classification (96.05%), allowing it to keep almost all of its predictive capacity in time.

5.3. Validation of the discriminatory power of the method in space

Since the MLP method was able to keep its predictive ability in time, we will now examine if it is able to keep its capacity in space. To find out, we will apply the method on data one, two and three years before distress for a new firm population, called the control sample. This test sample consists of 60 new firms 30 "healthy" and 30 "distressed”. The obtained results are as follows:

Table 21. 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	Correct ercentage	0	1	Correct percentage
Stape 1	Y	0	76	0	100,0	30	0	100,0
		1	0	76	100,0	1	29	96,67
	global Percentage				100,0			98,33

Table 22. Estimates of control sample two years before distress

		Classification table[c]						
			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	Percentage correct	0	1	Percentage correct
Stape 1	Y	0	76	0	100,0	30	30	100,0
		1	0	76	100,0	0	0	100,0
	global Percentage				100,0			100,0

Table 23. Estimates of control sample three years before distress

		Classification table[c]						
			Predicted					
			Selected observations[a]			Excluded observations[b]		
			Y			Y		
	Observations		0	1	Percentage correct	0	1	Percentage correct
Stape 1	Y	0	76	0	100,0	29	1	96,67
		1	0	76	100,0	0	30	100,0
	global Percentage				100,0			98,33

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

Table 24. The results provided by the model over time and space

Sample	Initial			Control		
	1year	2 years	3 years	1year	2 years	3 years
% of correct classification	100 %	98.68 %	96.05 %	98,33 %	100 %	98,33 %
% of error classification	0 %	1. 32 %	3.95 %	1,67 %	0 %	1,67 %
Error type I	0 %	2. 63 %	3.95 %	3,33%	0 %	0 %
Error type II	0 %	0 %	3.95 %	0 %	0 %	3,33 %

The results show a percentage of correct classification varying between 100 % (60/60) and 98,33% (59/60) for the coming three years before the failure. Type I error has reached 3.33 % (1/30) during the first year since the model has ranked one failed business as healthy. Type II error has reached 3.33 % (1/30) during the third year since the model has ranked one healthy business as failed. (Table 21, 22 and 23).

The MLP method has retained its discriminatory ability both in time with a correct classification rate that remains above 96.05%, and in space with a good ranking ratio of about 98,33%. Referring to the work done in the area, we find that our MLP model has better results than those of Min and Lee (2005), Olson and al (2012), Serrano-canca (2013) and wang and al (2014). However, our results remain below those reported by Wu et al (2007 (Table 25).

Table 25. The results provided by the literature review

Authors	Year	Method	Percentage of correct classification		
			One year	Two year	Three year
Ahn and al	2011	RNA	100%		
Boyacioglu and al	2009	RNA (PMC)	95,5%		
Min and Lee	2005	RNA(PMC)	85,25%		
Min and al	2006	RNA	79,57%		
Olson and al	2012	RNA	79,8%		
Serrano-canca and al	2013	RNA(PMC)	93,93%		
Serrano-canca	1997	RNA	93,94%		
Wang and al	2014	RNA	75,69%		
Wu and al	2007	RNA(PMC)	100%	100%	100%
Hu and Tseng	2005	RNA(PMC)	81,64%		
Hu and Tseng	2005	RNA	81,69%		
Lee and To	2010	RNA	96%		

6. Comparison of methods

6.1. Results from the models applied to initial samples

Table 26. Comparison of the two models applied to initial samples

	MDA	MLP
a) one year before distress		
- % of correct classification	100 %	100 %
- % of error classification	0 %	0 %
- Error du type I	0 %	0 %
- Error du type II	0 %	0 %
b) two years before distress		
- % of correct classification	96,71 %	98,68 %
- % of error classification	3,29 %	1,32 %
- Error du type I	6,58 %	2,63 %
- Error du type II	0 %	0 %
c) three years before distress		
- % of correct classification	**95,4 %**	**96,05** %
- % of error classification	4,6 %	3,95 %
- Error du type I	7,89 %	3,95 %
- Error du type II	1,32 %	3,95 %

The results obtained using the initial samples (validation in time) show a superiority of the MLP compared to the MDA method. Indeed, the MLP has a correct classification percentage that remains beyond 96.05 % against 95.4 % for MDA. Similarly, to the extent that the cost of a Type I error is much higher than that of a Type II error, we find that the maximum rate of error for MLP is largely lower than that of MDA (3.95% against 7.89 %).

6.2. Results from the models applied to control samples

Table 27. Comparison of the two models applied to control samples

	MDA	**MLP**
a) one year before distress		
- % of correct classification	93,34 %	98,33 %
- % of error classification	6,66 %	1,67 %
- Error du type I	10 %	3,33 %
- Error du type II	3,33 %	0 %
b) two years before distress		
- % of correct classification	95 %	100 %
- % of error classification	5 %	0 %
- Error du type I	6,67 %	0 %
- Error du type II	3,33 %	0 %
c) three years before distress		
- % of correct classification	**91,67 %**	**98,33 %**
- % of error classification	8,33 %	1,67 %
- Error du type I	6,67 %	0 %
- Error du type II	10 %	3,33 %

The above comparative table shows a clear superiority of the multilayer perceptron method, both in time and in space, compared to the multivariate discriminant analysis method. Indeed, even over three years of the initial sample and control, the correct classification rate has always remained greater than or equal to 96,05%, well above 91.67% for the multivariate discriminant analysis.

On the revised plan of literature, the superiority of the multilayer perceptron (MLP) is confirmed in the work of Udo (1993), Kumar et al. (1997), Wu (1999), Brabnazon and Keenan (2004), Yi-Chung Hu et al. (2005), Sangjae Lee et al. (2013) and Serrano-Cinca et al. (2013). However, for Coats and Fant (1993) and Stephen et al. (1994), the superiority of the MLP over the ADM is manifested when the data are not linearly separable otherwise their capabilities are identical. For Bardos and Zhu (1997), the fewer input variables that are correctly selected, the more the MLP dominates the ADM. Moreover, Tam and Kiang (1992) indicate that in the presence of a hidden layer the MLP is better than the ADM, otherwise both are equal. For Tam (1991), Odom and Sharda (1993), neural methods perform better than ADM for firms in difficulty, but conversely for healthy firms (Table 28).

Table 28. Comparison between MDA and MLP

Authors	years	Conclusion
Erxeleben	1991	The proof of the superiority of the ANNs on the ADM is not made but the following year it shows that the two tools reach the same results
Bardos and Zhu	1997	The fewer the number of input variables, the more the MLP> ADM
Wilson and Sharda	1994	MLP > ADM
Coats and Fant	1993	MLP > ADM when the data are not linearly separable otherwise their abilities are identical
Coats and Fant	1992	ANN > ADM
Boritz	1995	MLP = ADM, but there are differences between first and second type errors
Tam and Kiang	1992	In the absence of hidden layer we have ANN (MLP) = ADM, but in the presence of hidden layer we have MLP > ADM
Wu	1999	MLP > ADM
Kumar	1997	MLP > ADM
Odom and Sharda	1990	MLP > ADM
Udo	1993	MLP > ADM
Kerling	1994	MLP = ADM
Tsukuda	1994	MLP = ADM for listed companies MLP > ADM for unlisted companies
Altman, Marco and Varetto	1994	MLP ≥ ADM
Chung and Tam	1993	ANN > ADM
Philipe du Jardin	2007	ANN > ADM
Kira, Doreen and Nguyen	1997	ANN < ADM
Tam	1991	ANN > ADM For distressed companies ANN < ADM For healthy businesses
Guan	1993	ANN > ADM
Odom and sharda	1993	ANN > ADM For distressed companies ANN < ADM For healthy businesses
Alici	1996	ANN > ADM
Sung and al	1999	ANN > ADM
Anandarajan and al	2004	ANN > ADM
Cadden	1991	ANN > ADM
Back and al	1996	ANN > ADM For first and third year before failure, but ADM > ANN for second year.
Brabnazon and Keenan	2004	ANN (MLP) > ADM
Charitou and al	2004	ANN > ADM
Dimitras and al	1999	ANN > ADM
Serrano-cinca and al	2013	ANN (MLP) > ADM
Jae H. Min and Young-Chan Lee	2005	ANN > ADM
Wu et al	2007	ANN > ADM
Sangjae Lee and Wu Sung Choi	2013	ANN (MLP)> ADM
Boyacioglu and al	2009	ANN > ADM
Serrano-cinca	1997	ANN > ADM

Stephen P. Curram; John Mingers	1994	ANN > ADM when the data are not linearly separable otherwise their abilities are identical
Juliana Yim, Heather Mitchell	2005	ANN hybride > ADM
Mario Hernandez Tinoco, Nick Wilson	2013	ANN > ADM
Zhou and al	2012	ANN > ADM
Rafiei and al	2011	ANN > ADM
Yi-Chung Hu and Fang-Mei Tseng	2005	ANN (MLP) > ADM
Yi-Chung Hu and Fang-Mei Tseng	2005	ANN (FBR) > ADM
Yi-Chung Hu and Fang-Mei Tseng	2005	ANN (FBR) > ANN (MLP)

7. Conclusion

Both on the initial sample and on the control sample, the results provided by the methods chosen perform well either in terms of percentage of correct classification or the stability of their discriminatory power in time and space.

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

Despite the superiority of the results of the multilayer perceptron compared to those obtained by the multivariate discriminant analysis, the presence of several forecasting methods allows the financial analyst a wider choice and therefore more satisfaction and confidence. Indeed, when applying the models for the same business, we obtained the same results, then the creditor or the financial analyst will take their decision with more confidence. If on the contrary the models gave conflicting results, then the decision-maker is forced to probe more into this company.

Despite the statistical problems and the problems of temporal and sectoral robustness, which are common to all the techniques mentioned, the forecasting methods of firms in difficulty have the advantage of a systematic treatment of the information as well as a saving of time and cost for the decision-maker.

The linear discriminant analysis is the most widely used method from an operational point of view. Indeed, the score function is very useful for practitioners since it will allow them to calculate the posterior probabilities as well as the construction of risk classes for the studied companies.

The recent techniques borrowed from artificial intelligence, mainly neural networks, are very successful academic tools especially after the integration of genetic algorithms in their models, which have avoided local minima. Still in the exploratory phase, they are very promising given the absence of statistical restrictions and the robustness of the used genetic algorithms.

What brings us closer to these methods is the exclusive use of accounting and financial data by omitting qualitative variables such as the quality of human resources management, the degree of customer concentration or the age of the manager that would be probably able of improving the predictive ability of the method.

From the economic and social point of views, the presence of these forecasting models makes it possible to anticipate the failures and the difficulties encountered by companies. These offer financial analysts and economic managers the opportunity to provide corrections and the appropriate remedies allowing thus for preserving the economic fabric of the country and the jobs attached therein.

References

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

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

Altman, E. I., and al. (1994). "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)." journal of banking & finance**18**(3): 505-529.

Altman, E. I. (1984). "A further empirical investigation of the bankruptcy cost question." The journal of finance 39(4): 1067-1089.

Altman, E., (1984). The Success of Business Failure Prediction Models (An International Survey), Journal of Banking and Finance, 8, pp. 171-184.

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

Bardos, M. and 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.

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

Brabazon, A. and P. B. Keenan (2004). "A hybrid genetic model for the prediction of corporate failure." Computational Management Science **1**(3-4): 293-310.

Charitou, A., and al. (2004). "Predicting corporate failure: empirical evidence for the UK." European Accounting Review**13**(3): 465-497.

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

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

Edmister, R. O. (1972). "An empirical test of financial ratio analysis for small business failure prediction." Journal of Financial and Quantitative Analysis**7**(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

Frydman, H., and al. (1985). "Introducing recursive partitioning for financial classification: the case of financial distress." The journal of finance **40**(1): 269-291.

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

Kumar, N., and al. (1997). "Financial decision support with hybrid genetic and neural based modeling tools." European Journal of Operational Research 103(2): 339-349.

Lachenbruch, P. A., and al. (1973). "Robustness of the linear and quadratic discriminant function to certain types of non-normality." Communications in Statistics **1**(1): 39-56.

Lee, S. and W. S. Choi (2013). "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis." Expert Systems with Applications **40**(8): 2941-2946.

Lee, M.-C. and C. To (2010). "Comparison of multilayer perceptronand back propagation neural network in evaluating the enterprise financial distress." arXiv preprint arXiv:1007.5133.

Marks, S. and O. J. Dunn (1974). "Discriminant functions when covariance matrices are unequal." Journal of the American Statistical Association **69**(346): 555-559.

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.

Moyer, R. C. (1977). Forecasting Financial Failure: A Re-examination. Financial Management, 6(1), 11-17.

Olson, D. L., and al. (2012). "Comparative analysis of data mining methods for bankruptcy prediction." Decision Support Systems**52**(2): 464-473.

Platt, H. D. and M. B. Platt (1991). "A note on the use of industry-relative ratios in bankruptcy prediction." journal of banking & finance**15**(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.andGiroux 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. and B. Gutiérrez-Nieto (2013). "Partial least square discriminant analysis for bankruptcy prediction." Decision Support Systems**54**(3): 1245-1255.

Salchenberger, L. M., and al. (1992). "Neural Networks: A New Tool for Predicting Thrift Failures*." Decision Sciences **23**(4): 899-916.

Taffler, R. J. (1983). "Forecasting company failure in the UK using discriminant analysis and financial ratio data." Journall of the Royal Statistical Society. Series A (General): 342-358.

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

Tamari, D. (1964). "Sur quelques problèmes d'associativité." Annales de la Faculté des sciences de l'Université de Clermont. Série Mathématiques **24**(3): 91-107.

Tan, C. N. (1997). "An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System." Bond University.

Tserng, H. P., and al. (2011). "An enforced multilayer perceptronmodel for construction contractor default prediction." Automation in Construction**20**(8): 1242-1249.

Udo, G. (1993). "Neural network performance on the bankruptcy classification problem." Computers & industrial engineering **25**(1): 377-380

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

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, 44 p.

Wu, C.-H., and al. (2007). "A real-valued genetic algorithm to optimize the parameters of multilayer perceptronfor predicting bankruptcy." Expert Systems with Applications**32**(2): 397-408.

Zhou, L., and al. (2012). "Empirical models based on features ranking techniques for corporate financial distress prediction." Computers & mathematics with applications **64**(8): 2484-2496.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, 22, 59-82.

Appendix 1

The battery of 87 ratios initially used

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 if total liabilities exceed total assets, zero otherwise

Appendix 2

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
R1	,991	1,348	1	150	,247
R2	,850	26,417	1	150	,000
R3	1,000	,000	1	150	,989
R4	,926	12,027	1	150	,001
R5	,928	11,667	1	150	,001
R6	,864	23,515	1	150	,000
R7	,883	19,885	1	150	,000
R8	,887	19,065	1	150	,000
R9	,990	1,540	1	150	,216
R10	,998	,234	1	150	,629
R11	,993	1,093	1	150	,298
R12	,849	26,615	1	150	,000
R13	,998	,358	1	150	,551
R14	,976	3,721	1	150	,056
R15	,828	31,080	1	150	,000
R16	,995	,780	1	150	,379
R17	,994	,878	1	150	,350
R18	,943	9,010	1	150	,003
R19	,759	47,732	1	150	,000
R20	1,000	,028	1	150	,868
R21	,981	2,836	1	150	,094
R22	,978	3,432	1	150	,066
R23	,982	2,808	1	150	,096
R24	,979	3,140	1	150	,078
R25	,848	26,807	1	150	,000
R26	,652	79,976	1	150	,000
R27	,998	,352	1	150	,554
R28	,859	24,701	1	150	,000
R29	,987	1,919	1	150	,168
R30	,997	,427	1	150	,514
R31	,890	18,517	1	150	,000
R32	,999	,110	1	150	,740
R33	,883	19,909	1	150	,000
R34	,968	4,950	1	150	,028
R35	,993	1,073	1	150	,302
R36	,995	,730	1	150	,394
R37	,986	2,159	1	150	,144
R38	,959	6,447	1	150	,012
R39	,993	1,030	1	150	,312
R40	,981	2,921	1	150	,090
R41	,977	3,575	1	150	,061
R42	,970	4,677	1	150	,032
R43	,865	23,501	1	150	,000
R44	,857	24,960	1	150	,000
R45	,979	3,290	1	150	,072
R46	,978	3,435	1	150	,066
R47	,813	34,409	1	150	,000
R48	,834	29,925	1	150	,000
R49	,999	,193	1	150	,661
R50	,832	30,369	1	150	,000
R51	,858	24,904	1	150	,000
R52	,957	6,773	1	150	,010
R53	,916	13,843	1	150	,000
R54	,999	,106	1	150	,746
R55	,984	2,372	1	150	,126
R56	,977	3,552	1	150	,061

R57	,999	,225	1	150	,636
R58	,990	1,559	1	150	,214
R59	,991	1,396	1	150	,239
R60	,999	,200	1	150	,655
R61	,970	4,629	1	150	,033
R62	,923	12,465	1	150	,001
R63	,985	2,244	1	150	,136
R64	,933	10,785	1	150	,001
R65	,918	13,351	1	150	,000
R66	,990	1,540	1	150	,216
R67	,992	1,232	1	150	,269
R68	,980	3,008	1	150	,085
R69	,996	,541	1	150	,463
R70	,944	8,910	1	150	,003
R71	,833	29,967	1	150	,000
R72	,985	2,292	1	150	,132
R73	,817	33,588	1	150	,000
R74	,829	30,994	1	150	,000
R75	,998	,268	1	150	,605
R76	,995	,738	1	150	,392
R77	,943	9,010	1	150	,003
R78	,958	6,633	1	150	,011
R79	,785	41,038	1	150	,000
R80	,963	5,803	1	150	,017
R81	,992	1,141	1	150	,287
R82	1,000	,042	1	150	,838
R83	,891	18,401	1	150	,000
R84	1,000	,045	1	150	,832
R85	,988	1,898	1	150	,170
R86	,799	37,684	1	150	,000
R87	,765	45,996	1	150	,000

Appendix 3

Network information

Entrance stratum	Covariables	1	R6
		2	R7
		3	R15
		4	R16
		5	R19
		6	R26
		7	R33
		8	R40
		9	R58
		10	R61
		11	R73
		12	R78
		13	R79
		14	R83
		15	R84
		16	R85
	Number of units[a]		16
	Rescaling method for covariates		standardized
Hidden stratum (s)	Number of hidden layers		1
	Number of units in the hidden stratum 1[a]		5
	Activation function		Hyperbolic tangent
Output stratum	Dependent variables 1		Y
	Number of units		2
	Activation function		MaxMou
	Error function		Cross entropy

a. Exclusion of the biased unit

Appendix 4

Summary of models

Learning	Cross entropy error	,477
	Incorrect percentage forecasts	,0%
	Stopping the rule used	1 consecutive step (s) without decrease in error
	Duration of training	00:00:00,031
Test	Cross entropy error	3,750
	Incorrect percentage forecasts	1,7%

Dependent variable: Y

a. Error calculations are based on the test sample.