Quest Journals
Journal of Research in Business and Management
Volume 4 ~ Issue 3 (2016) pp: 10-15
ISSN(Online) : 2347-3002
www.questjournals.org

Research Paper

The Artificial Neural Networks for Classification: Case of Business Failures

Aomar IBOURK[1], Abdelkrim AAZZAB[2]

Received 09 April, 2016; Accepted 25 April, 2016 © The author(s) 2015. Published with open access at **www.questjournals.org**

ABSTRACT:-*The statistical classification techniques can be divided into parametric and non-parametric techniques which we find artificial neural networks (ANN). The ANN, unlike conventional classification techniques require no assumption about variables and they are quite suitable for unstructured complex problems [10]. This robustness has allowed these techniques a special enthusiasm in various areas of research and mainly in corporate failures. Indeed, the first experiments modeling business failure by applying artificial neural networks start in 1990 with Bell & al .. Next, the use of this technique has intensified with the work of Tam (1991) , Keasey and Watson (1991), Dimitras et al. (1996), Altman and Narayanan (1997), Wong et al. (1997), Zhang et al. (1998), Coakley and Brown (2000), Aziz and Dar (2004), and Ooghe Balcaen (2004, 2006), Ravi Kumar and Ravi (2007) and Lin (2009).*
The objective of this paper is to present and implement the methodology of multilayers neural networks as a classification tool for the issue of business failure based on accounting data from a sample of Moroccan companies.

Key words:-*classification techniques, artificial neural networks, corporate failures*

I. INTRODUCTION

The artificial neural networks (ANN), inspired by biological neural networks is an area of research for addressing the problems of perception, memory, learning and reasoning. As statistical classification techniques, unlike parametric techniques, ANN are robust to errors specification and appear to be universal approximators particularly parsimonious [11] [12] [13].

The use of neural networks for the prediction of business failures really began in the 1990s [3] [4] with the work of Odom & Sharda (1990). This method, which is based on the information processing performed by the human brain, is to develop a learning algorithm that processes a set of information to get a result. Multiple studies and research works on business failure have practiced this technique which are found Bell and al. (1990), Keasey and Watson (1991), Dimitras et al. (1996), Altman et al. (1994), Wong et al. (1997), Zhang et al. (1998), Coakley and Brown (2000), Aziz and Dar (2004), Ooghe and Balcaen (2004, 2006), Ravi Kumar and Ravi (2007) and Lin (2009).

In our study, drawing on empirical work, we selected first two classes of firms by non-defaulting and defaulting criterion, then, we have chosen a set of explanatory variables and then say it was sought to establish a statistical relationship between these variables and the dichotomous state of being or not being faulty. The quality of the model developed depends on the rate of correct classification of enterprises in the corresponding class.

II. DATA AND RESEARCH METHODOLOGY

Firstly, in order to keep only the relevant variables, the most discriminating, with the aim of improving the model prediction quality, we started with the selection of explanatory variables among a set of variables candidates selected on the basis of previous empirical work. Thus, the procedure for selecting variables is based on 1000 bootstarps samples and variables used are those with the highest power of discrimination. This power is calculated by ranking the variables in ascending order according to Fisher statistic and in selecting frequency [17]. Secondly, we have developed models predicting failure based on the artificial neural network method.

1.1 Selection of firms

Our approach to data collection consists of three steps: the choice of the database, the selection of the firms and the choice of indicators of failure. In effect, to constitute our sample, based on an official source of information, we have purchased the accounting synthesis documents from OMPIC. It is the departure of 160 firms whose 50% represents the failing firms. As well, to delineate our field of investigation and to ensure the maximum homogeneity of the composite sample, we selected companies operative in the sector of industry and which are of small and medium sizes. The choice of this sector lies first in the significant number of failed companies operating there, and to the ability to calculate the set of financial ratios described by the theory, a thing which is not possible for services companies, for example, who do not have some indicators.
Thus, the criterion size affects the companies which have achieved, during the year that it was retained for the analysis, an annual turnover not exceeding 75MAD or a balance sheet total not exceeding 50MAD. Our final sample consists of 132 companies, half of which has failed. This balance between the two types of companies up to empirical considerations which show that an imbalance between classes has a negative effect on the correct classification rate of each group and the overall correct classification rate [1] [14].
For the companies in good health, we have begun a choice at random without any other hypothesis. Whereas for the failing firms, we have identified them with the commercial courts prior to requesting their states of syntheses. The commercial courts chosen are those of Agadir, Marrakech and Casablanca. Our choice here is motivated by the ease of access to information and by the proximity.

Thus, for each failed company, we have requested the synthesis documents of an accounting period before the date of declaration of default. For the non-defaulting, it is also one exercise pulls randomly. Also, our sample covers a five-year period from 2006 to 2010. The choice of this period is mainly due to the difficulties related to the identification of failing companies on a shorter period. The following table summarizes the description of the businesses that make up our database according to the type and by regions.

Table 1 : Distribution of companies by regions

	Agadir	Marrakech	Casablanca	Total
Non-faulty	18	22	26	66
Faulty	13	17	36	66
Total	31	39	62	132

1.2 Choice of variables

Our database is composed of 18 financial ratios. These ratios are calculated on the basis of the documents collected in order to constitute a battery relevant and credible likely to respond to our question concerning the explanatory factors of business failure. The justification of choice of these ratios is based mainly on the theoretical and empirical literature [1] [6] [8], Pink and Giroux, 1984, redone, 2004).
Thus, the variable to explain is dichotomous, it takes the value 1 if the firm is faulty and the value 0 if the firm is non-faulty. For the explanatory variables, and to the extent that there is no unifying theory defining the failure of businesses, our work is also included in the same way that most of the empirical models that begin with a high number of factors and reduce in order to keep only a few judges as the most explanatory of the risk of failure. Then, we are therefore limited to a basic battery consisting of 18 ratios according to their popularity and their performance in the previous studies. Annex 1 summarizes the ratios of our study which represent the set of financial indicators chosen.

1.3 Variables selection methods

We opted for the automatic variables selection methods by comparing between two methods to finally select the one which presents more precision [5] [9]. In effect, we have proceeded to the selection of variables using the method called "Stepwise discriminant analysis SDA". Then, we compared between the Forward approach and the so-called Backward. For these methods of variables selection, and to assess the significant role of a variable, we use the statistic F of Fisher. Therefore, it would be sufficient to compare the p-value calculated for the variable to assess and compare with the level of significance chosen. As well, the Wilks lambda, which varies between 0 and 1, represents the preferred indicator for the statistical evaluation of the model [18]. It indicates to what extent the centers of classes are separate from each other in the space of representation. As long as it tends to 0 the model will be good because the clouds are quite distinct.

1.4 Construction of the neuronal model

The network of neuron developed is of type "multi-layers Perceptrons" with the simple gradient descent based on the error backpropagation algorithm [19] as optimization technique. Thus, we have retained the hyperbolic tangent as activation function and the error of least squares as a cost function. Moreover, for the modification of the weight of the network, we have opted for a term of time and each layer is begun of a bias and a term of regularization. Finally, we have retained the sum of square errors (SSE) as a performance measurement function.

For network setup, we adopted supervised learning for a layered network, not curly, fully connected, with a hidden layer and a linear output.
For the input layer, it is the vector of variables selected candidates for learning. For the number of neurons to introduce in the hidden layer, it is to test the different configurations which led to a level of learning high. For the output layer, the variable to explain is dichotomous. It is a vector that takes the value 1 if the company is faulty and the value 0 if the company is not-faulty. As well, to ensure a better learning and to stabilize the process of selection of variables, we have employed bootstrap [17] techniques of resampling.
Too, we resorted to the definition of a random generator by creation of a variable partitioning in order to recreate exactly the samples used in the analyzes. It is a randomly Bernoulli variable generated with a probability parameter of 0.7, modified so as to take the value 1 or -1, instead of 1 or 0 (faulty or not-faulty). Then, the observations containing positive values on the variable of partitioning are assigned to the sample of learning, those with negative values are assigned to the validation sample and those with a value equal to 0 are assigned to the test sample. The latter is formed to avoid the problems of over-learning in order to help the network to remain "on the right track". For the other parameters of the network (the learning step, the term of time and the terms of regularization of weights), the values are set on the basis of the empirical work found in the literature. As well, the number of iterations to retain is the one for which the error does varies almost more beyond this number. Finally, in the aim to delete all that is modeled in order to reduce the complexity of the network and to accelerate its convergence, we performed pretreatments on the standardization of data based on the Min-Max method.

III ANALYSIS OF RESULTS

1.5 Performance of classification and selection of variables

1.5.1 Confusion matrix

The confusion matrix of the two variables selection methods (Table 3) indicate a rate of misclassification of 0.0909 for the SDAF

Table 2 and a rate of 0.0985 for the named SDAB. The error rates calculated on the training data are then very optimistic and the estimator of the error bootstrap gives the advantage to the SDAF which has a value of 0.1221 instead of 0.1279 for the SDAB.

Table 2: Classification performance of variable selection methods

SDA (FORWARD)

Error rate	0,0909		
Bootsrap error estimation	0,1221		
Confusion matrix			
	D	ND	Sum
D	61	5	66
ND	7	59	66
Sum	68	64	132

SDA (BACKWARD)

Error rate	0,0985		
Bootsrap error estimation	0,1279		
Confusion matrix			
	D	ND	Sum
D	59	7	66
ND	9	57	66
Sum	68	64	132

The first method indicates that 61 failing companies have been well reclassified and 5 have incorrectly been. Similarly, for the companies not-faulty, 7 of them have been incorrectly reclassified and 59 are well reclassified. In total, it is therefore 120 firms (60 + 57) which have been correctly reclassified with a rate of correct classification of 90.90 %.

1.5.2 The MANOVA Test

The analysis of the multivariate variance (Table 3) shows that's the method of SDAF which shows good results. In effect, it has the more low of Wilks lambda statistics (0.37). This result is confirmed by the transformations of Bartlett or Rao who adjudicate on the significance of deviations, and which lead to the same conclusion on the threshold of error of 5%. We then rejects the hypothesis that the centers of classes are combined (p-value= 0).

Table 3 : The analysis of variance multivariate

SDA (FORWARD)			SDA (BACKWARD)		
Stat	Value	p-value	Stat	Value	p-value
Wilks' Lambda	0,3789	-	Wilks' Lambda	0,3909	-
Bartlett C(18)	123,74	0,00	Bartlett C(18)	120,22	0,00
Rao F(18, 113)	41,31	0,00	Rao F(18, 113)	49,47	0,00

Then, by marrying the result of MANOVA test with that of the confusion matrix, we understand that the proper holding of the model holds especially for the application of the method of forward stepwise discriminant analysis

1.5.3 Selecting variables

The individual assessment of the predictor variables shows that five variables that contribute to the explanation of the failure to the SDAF and four variables for the SDAB. Thus, table 4 shows that these results also indicates that four common variables between the two methods (R3, R5, R7 and R16).

Table 4 : Individual assessment of the predictor variables

SDA (FORWARD)	SDA (BACKWARD)
R7, R16, R5, R3 et R15	R3, R5, R7 et R16

All of these variables are selected on the basis of the statistics F which is significantly different from zero, because the p-value is less than 5%.

1.5.4 Neural network

1.5.4.1 Architecture of neural models

According to the table 5, we note that, by the employment of all the candidate variables, the best architecture is the one using a hidden layer with a single neuron (Net1_1 (18 1 1). In effect, this is the architecture for which the sum of the quadratic error is minimum for the learning sample (7.68). The SSE for the sample test is the 2.89, this is not therefore the minimum value but it corresponds as even at a low value if it is compared with the other. This architecture has been used to record a rate of correct classification of 87.3% for the learning sample and a rate of 85% for the sample test.

However, the optimum architecture corresponding to the employment of selected variables by the method SDAF is composed of a hidden layer with 9 neurons (Net2_6 (6 9 1). With regard to it, this architecture has enabled us to save the lowest value of the sum square error for the learning sample with 7.86% and an error rate of 2.98% for the sample test. For this network, the rate of correct classification is that of 84.8% for the learning sample and of 85% for the sample test.

Table 5 : Summary of tests of network architectures

Employment of 18 candidate variables			Employment of 5 selection variables (SDAF)		
Architecture	SSE [1] of the learning sample	SSE [2] of the test sample	Architecture	SSE of the learning sample	SSE of the test sample
Net1_1 [18 1 1]	**7,68**	**2,89**	Net2_1 [5 1 1]	10,83	2,78
Net1_2 [18 3 1]	10,15	2,39	Net2_2 [5 3 1]	13,79	2,69
Net1_3 [18 5 1]	9,61	2,82	Net2_3 [5 5 1]	9,83	3,08
Net1_4 [18 7 1]	9,96	2,61	Net2_4 [5 7 1]	10,89	3,22
Net1_5 [18 8 1]	11,42	2,83	Net2_5 [5 8 1]	9,75	3,11
Net1_6 [18 9 1]	11,30	2,74	**Net2_6 [5 9 1]**	**7,86**	**2,98**
Net1_7 [18 10 1]	9,58	2,91	Net2_7 [5 10 1]	11,17	2,91

[1] The sum of the quadratic errors (Sum squared error SSE) committed at the time of the classification of firms in the sample of learning.

[2] The sum of the quadratic errors (Sum squared error) committed during the classification of firms from the test sample.

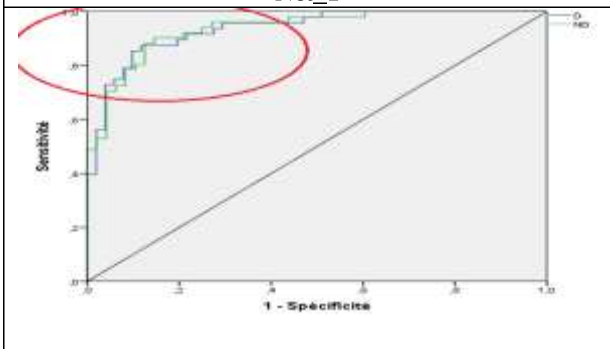
Table 6 : The confusion matrix neural models

Modele_1:Employment of 18 candidate variables (Net_1)					**Modele_2: Employment of 5 selected variables (Net_2)**				
Sample		Forecasts			Sample		Forecasts		
		D	ND	% Correct			D	ND	%
Learning	D	34	6	85.0 %	Learning	D	34	6	85.0 %
	ND	4	35	89.7 %		ND	6	33	84.6 %
	% Global	48.1 %	51.9 %	**87.3 %**		% Global	50.6 %	49.4 %	**84.8 %**
Test	D	7	1	87.5 %	Test	D	7	1	87.5 %
	ND	2	10	83.3 %		ND	2	10	83.3 %
	% Global	45.0 %	55.0 %	**85.0 %**		% Global	45.0 %	55.0 %	**85.0 %**
Validation	D	15	3	83.3 %	Validation	D	16	2	88.9 %
	ND	1	13	92.9 %		ND	1	13	92.9 %
	% Global	50.0 %	50.0 %	**87.5 %**		% Global	53.1 %	46.9 %	**90.6 %**

1.5.4.2 Validation of neural models

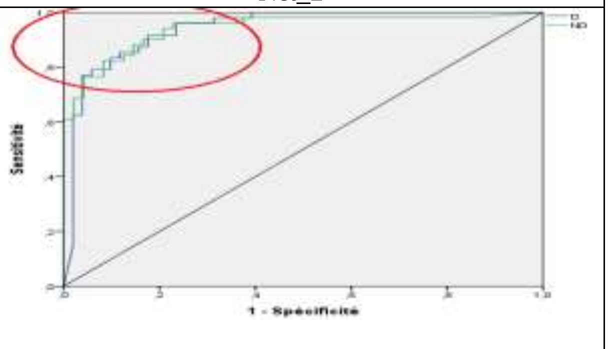
As for the validation sample, that evaluates the final neural network and thus validates the model, the percentage of correct classification shows that 88.9% of the failing companies its well classified by the network Net_2 whereas the network Net_1 arrives only to properly classifying 83.3% of these companies. For the companies non-defaulting, the both networks have correctly classified 92.9% of them. Therefore, the overall rate of correct classification of the Net_2 displays is of 90.6% and that of Net_1 is only 87.5%. The validation of neural models can be strengthened by the analysis of ROC curves[3].

Net_1



Area under the curve	
Faulty	0,936
Not-faulty	0,936

Net_2



Area under the curve	
Faulty	0,941
Not-faulty	0,941

Graph 1 : the ROC curves of the two neural networks

We find for the two networks that the ROC curves are a little close to the top corner-left. Then, the performance of discrimination factors is acceptable. This means that the probability that the Score function, developed by the neural model, place a failing company before a company non-defaulting is almost close to 1 for the two neural models. Thus, for the network of neurons Net_2, for a random choice of a failing firm and a company not-faulty, there is a probability of 94.1% that the pseudo-probability of breach provided by the model is higher for failed company. That is to say that the probability that the network place a failing firm before a company not-faulty is of 94.1%. This rate is 93.6% for the network Net_1.

It is apparent that, from the analysis of the whole of these elements of validation, the two neural models are valid and record of good results. As well, the rates recorded by the two models are very optimistic that this is for the learning sample, the test sample or the validation sample. However, to decide between them, we can say that, without doubt, the neural model based on 6 variables is the most powerful on all levels. In effect, with the

An ROC (Receiver Operating Characteristic) curve displays the modalities for each dependent variable qualitative. It presents a visual display of *sensitivity* and *specificity* for all possible hyphenation should in a unique diagram, which constitutes a tool more clear and more powerful than a series of tables.

exception of the learning sample, this network has recorded the highest rates for the sample test and for the validation sample. More, with a reduced number of variables, the second model has been used to record the results more salient compared to the first model based on 18 variables. Then, our model of neural networks chosen is the one based on 6 explanatory variables with an architecture consisting of a hidden layer with 9 neurons.

IV. CONCLUSION

The results show that the analysis of financial variables (ratios) has allowed us to detect those most indicative of the failure. These are the variables from four of the five dimensions of financial analysis that are the basis of the company's failure explanation to know the financial structure, activity, liquidity and management. This result then confirms the successful outcomes of Conan & Holder (1979) or Combier & Blazy (1997).

The comparison of the two classification methods in terms of predictability shows in our case the performance of conventional models over than the neural networks networks. In fact, the percentage of correct classification measured by the linear discriminant analysis is better than artificial neural networks on the learning samples and test sample, with the exception of the validation sample or the neural networks show a slight superiority.

This result thus cripples those already established empirical studies that have shown the success of these nonparametric methods in predicting business failure [15] [16].
Note finally that this study has some limitations in the frame where the models developed are based on a small number of observations and multicollinearity tests, multi-normality tests and homoscedasticity tests are not checked.

REFERENCES

[1]. ALTMAN E. I., [1968], "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, vol. 23, n° 4, pp. 589-609.

[2]. ALTMAN E. I., MARCO G. et VARETTO F., [1994], "Corporate Distress Diagnosis : Comparisons Using Linear Discriminant Analysis and Neural Network – The Italian Experience", *Journal of Banking and Finance*, vol. 18, n° 3, pp. 505-529.

[3]. ANANDARAJAN M., LEE P. et ANANDARAJAN A., [2001], "Bankruptcy Prediction of Financially Stressed firms: An Examination of the Predictive Accuracy of Artificial Neural Networks", *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 10, n° 2 pp 69-81.

[4]. ATIYA A. F., [2001], "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results", *IEEE Transactions on Neural Networks*, vol. 12, n° 4, pp. 929-935.

[5]. BACK B., LAITINEN T., SERE K. et VAN WEZEL M., [1996 A], "Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms", Turku Centre for Computer Science, Technical Report, n° 40, 18 p.

[6]. BARDOS M. et ZHU W. H., [1997], « Comparaison de l'analyse discriminante et des réseaux de neurones. Application à la détection des défaillances d'entreprises », *Revue de statistique appliquée*, tome 45, n°4, pp. 65-92.

[7]. BLAZY R. et COMBIER J., [1997], « La défaillance d'entreprise : Causes économiques, traitement judiciaire et impact financier », INSEE Méthodes n° 72-73, *Économica*, Paris ;

[8]. CONAN J. et M. HOLDER [1979], « Analyse des causes de faillite des PMI », Thèse d'Etat, Université Paris-Dauphine, CEREG, cités par DAIGNE, 1986.

[9]. DESMET P., [1996], « Comparaison de la prédictivité d'un réseau de neurone à rétropropagation avec celles des méthodes de régression linéaire, logistique et AID pour le calcul des scores en marketing direct», Recherches et Application en Marketing, vol.11, n°2, pp.17-27.

[10]. PARIZEAU M., [2004], « Réseaux de neurones », Université de laval, 117 p.

[11]. CYBENKO G., [1989], "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals and Systems*, 2, pp. 303-314.

[12]. HORNIK et *al.*, [1989], « Multilayer Feedforward Networks are Universal Approximators", *Neural Networks*, vol. 2, n° 5, 359 p.

[13]. DREYFUS G. et *al.*, [2008], « Apprentissage statistique », Editions Eyrolles, 88 p.

[14]. BERG D., [2005], "Bankruptcy Prediction by Generalized Additive Models", University of Oslo, Department of Mathematics, Statistical Report, n° 1, January, 36 p.

[15]. ODOM M. D. et SHARDA R., [1990], "A neural network model for bankruptcy prediction", In : Proceedings of the IEEE International Joint Conference on Neural Networks. San Diego, CA, pp. 163–168.

[16]. BALCEAN S. et OOGHE H., [2004], "Alternative Methodologies in Studies on Business Failure : Do They Produce Better Results than the Classical Statistical Methods ?", Ghent University, Working Paper, n° 2004-249, 40 p.

[17]. SAUERBREI W. et SCHUMACHER M., [1992], "A Bootstrap Resampling Procedure for Model Building: Application to the Cox Regression Model", Statistics in Medicine, vol. 11, n° 16, pp. 2093-2109.

[18]. BARDOS M., [2001], « Analyse Discriminante - Application au risque et scoring financier », Dunod.

[19]. RUMELHART D. E., HINTON, G. E. et WILLIAMS R. J., [1986b], "Learning Representation by Back-Propagation Errors", Nature, Vol. 323, No. 9, pp. 533-536.

*Corresponding Author: Aomar IBOURK