Economics Bulletin

Volume 34, Issue 1

Bankruptcy prediction for Tunisian firms : An application of semi-parametric logistic regression and neural networks approach

Manel Hamdi
International Finance Group Tunisia, El Manar University, Tunisia

Sami Mestiri
Applied Economics and Simulation, Monastir University, Tunisia

Abstract

The paper uses two approaches, semi-parametric logistic regression model and artificial neural networks, to predict bankruptcy of Tunisian companies. A sample of 528 Tunisian firms for the period (1999-2006), was used to investigate the performance of these two approaches. The empirical results indicate that the quality of model prediction of the neural networks is better than the semi-parametric logistic regression model in terms of comparing the rates of misclassification and the area under curve (AUC) measures of the two proposed techniques. This research concludes that neural nets are a very powerful tool in bankruptcy prediction.

Citation: Manel Hamdi and Sami Mestiri, (2014) "Bankruptcy prediction for Tunisian firms : An application of semi-parametric logistic regression and neural networks approach", *Economics Bulletin*, Vol. 34 No. 1 pp. 133-143.
Contact: Manel Hamdi - mannelhamdi@yahoo.fr, Sami Mestiri - mestirisami2007@gmail.com.
Submitted: December 02, 2013. **Published:** January 30, 2014.

1. Introduction

Bankruptcy prediction is a very important topic that has attracted a lot of attention to researchers since past decades. Through the literature focused on this area, we find the pioneering work of Beaver (1966) and Altman (1968). And since time, an increasingly number of studies have been concentrated on models forecasting of financial distress. Different statistical techniques such as discriminant analysis (Deakin, 1972), logistic regression model (Ohlson, 1980; Pang, 2006), and recently the logistic regression with random effects (Mestiri & Hamdi, 2012); have been applied for predicting bankruptcy in companies. In a more recent research, Giordani et al. (2013) compared the logistic regression model with the additive spline regression model to predict bankruptcy risk of Swedish firms. The authors have concluded that the spline model outperforms the standard logistic model in term of good classification rate that has been enhanced from 70% to 90%. On the other hand and for the same purpose, artificial intelligence tools have been widely utilized. Among these tools, artificial neural networks are the most popular and the most used to predict financial distress (Perez, 2006; Chih-Fong & Jhen-Wei, 2008; Odom & Sharda, 1990; Wilson & Sharda, 1994; Khashman, 2011; Ghatge & Halkarnikar, 2013). The bankruptcy forecasting results of artificial neural networks model noticeably outperforms the traditional and statistical techniques according to several studies (Zhang et al. 1999; Yim and Mitchell, 2005; Xie et al. 2004; among them).

In this paper we present a comparative study between semi parametric logistic regression model and artificial neural networks technique to predict bankruptcy for Tunisian companies. The semi parametric logistic regression is used to overcome the shortcomings of linear regression model when the studied phenomenon is complicated. In fact, Zhang and Lin (2003) have proposed a flexible modeling of the effects of explanatory variables and the linear predictor in the regression model is replaced by non-parametric functions therefore a semi parametric logistic regression model has been formulated.

The paper is organized as follows: In Section 2, we present the description of the data of our study. Then, the semi-parametric logistic regression is applied, in section 3, to predict financial distress for Tunisian firms. The fourth section is devoted to the presentation and application of artificial neural networks. In Section 5, we evaluate the quality of model predictions. Finally, main conclusions are presented in section 6.

2. Data description

2.1 Data and exogenous variables

A battery of 26 financial ratios has been selected and calculated from balance sheets and income statements of 528 firms from different activities sectors. The data applied in this study have been collected from the Central Bank of Tunisia covering the period from 1999 to 2006. These ratios represent the exogenous variables of the model (see Table 1).

Table 1. The inputs of the model

$x_{j\ (j:1 \longrightarrow 26)}$	*Ratios definition*
R_1	Raw stock / Total assets
R_2	Duration credit to the customer
R_3	Gross margin rate
R_4	Operating margin rate
R_5	Ratio of personnel expenses
R_6	Net margin rate
R_7	Asset turnover
R_8	Equity turnover
R_9	Economic profitability
R_{10}	Rate of return on assets
R_{11}	Operating profitability of total assets
R_{12}	Gross economic profitability
R_{13}	Net economic profitability
R_{14}	Rate of return on equity
R_{15}	Permanent capital turnover
R_{16}	Return on permanent capital
R_{17}	Rate of long-term debt
R_{18}	Ratio of financial

	independence
R_{19}	Total debt ratio
R_{20}	Immobilisation coverage by equity capital
R_{21}	The long and medium term debt capacity
R_{22}	Ratio of financial expenses
R_{23}	Financial expenses/total debt
R_{24}	Working capital ratio
R_{25}	Relative liquidity ratio
R_{26}	Quick ratio

2.2 Endogenous variable

The criterion of a priori classification adopted in this analysis is the legal state of the company. Therefore, the group of firms is decomposed a priori into two categories: healthy (448 companies) and failing (80 companies). The dependent variable can be written as two binary values:

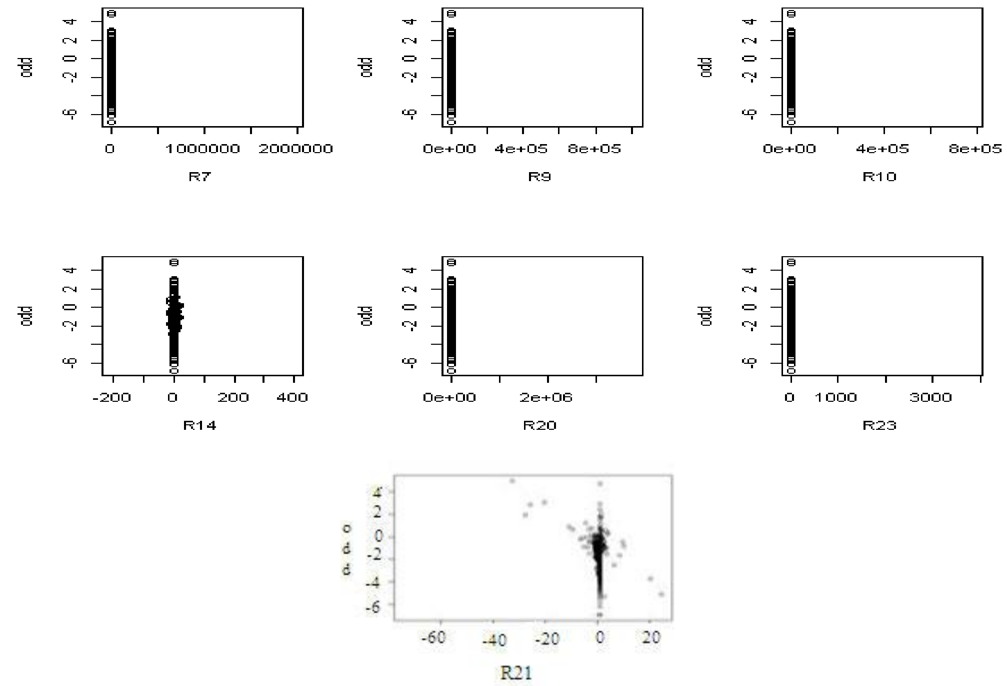
$$y = \begin{cases} 1, \; for \;\; default \; firm \\ 0, \; for \;\; healthy \;\; firm \end{cases} \tag{1}$$

3. Analysis via semi-parametric logistic regression model

3.1 General presentation of the model

Contrary to linear regression, it is not useful to directly plot x_j vs. y. Thus, the semi-parametric logistic regression model assume that the logits of the probability of default should relate in a linear way the explanatory variables X. Therefore, we divide the range of each of the variables x_j into intervals or classes of similar length and estimate their logits in these intervals using the observed frequencies of y = 0 and y = 1. The class centers are then plotted against the estimated logits (see Figure 1).

Figure 1. The scatter plots of the explanatory variables and their logits



Based on the resulting scatter plots presented in Figure 1, it is clear that the scatter plots for $R_7, R_9, R_{10}, R_{14}.R_{20}, R_{23}$ follow a linear trend whereas for the variable R_{21} is not the case. Therefore, the variable R_{21} is the most interesting component for considering a nonparametric modification of the index function. Then, we only consider R_{21} as the variable that to be used within a nonparametric function. The semi parametric modification of the logit model takes the following form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 R_{7i} + \beta_2 R_{9i} + \beta_3 R_{10i} + \beta_4 R_{14i} + \beta_5 R_{20i} + \beta_6 R_{23i} + f(R_{21i}) \qquad (2)$$

Where $p_i = P(y_i = 1 \mid R_i)$, for $(i = 1,\dots,n)$ is the probability of belonging to the group of distress firms, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)'$ is a vector of unknown parameters and f is an unknown smooth function.

3.2 Econometric model presentation

To determine the function $f(R_{21})$, we transform the nonparametric function (2) as a parametric function composed by the quadratic truncated power basis of the second degree:

$$f(R_{21,i}) = \delta_0 + \delta_1 R_{21,i} + \delta_2 R_{21,i}^2 + \sum_{k=1}^{K} b_k (R_{21,i} - \kappa_k)_+^2 \qquad (3)$$

Where $\kappa_1,\dots,\kappa_K$ is a whole of knots distinct drawn from the observations from the variable R_{21} and $X_+ = max(0; X)$. The number of knots K is rather large (of order $K \geq 30$) to ensure the exigibility of the curve (see Figure 2).

Following the approach of Brumback et al. (1999), we suggest formulating the semi parametric logistic regression model by utilizing the truncated power function basis of degree 2. Therefore, we obtained the transformed model (4) with structure of the logistic mixed model by replacing the equation (3) into (2):

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \delta_0 + \delta_1 R_{21,i} + \delta_2 (R_{21,i})^2 + \beta_1 R_{7,i} + \beta_2 R_{9,i}$$
$$+ \beta_3 R_{10} + \beta_4 R_{14,i} + \beta_5 R_{20,i} + \beta_6 R_{23,i} + \sum_{k=1}^{K} b_k (R_{21,i} - \kappa_k)_+^2 \tag{4}$$

To write the model in matrix form, we consider:

$X = \begin{bmatrix} 1 & R_{21,1} & (R_{21,1})^2 & .. & R_{23,1} \\ : & : & : & & : \\ 1 & R_{21,n} & (R_{21,n})^2 & .. & R_{23,n} \end{bmatrix}$ is a matrix composed of the explanatory variables,

$Z = \begin{bmatrix} (R_{21} - \kappa_1)_+^2 & ... & (R_{21} - \kappa_K)_+^2 \\ : & & : \\ (R_{2n} - \kappa_1)_+^2 & ... & (R_{2n} - \kappa_K)_+^2 \end{bmatrix}$ is a matrix (n, K) composed of the bases,

$\beta = (\delta_0 + \beta_0, \delta_1, \delta_2, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)'$ is a vector of unknown parameters, $b = (b_1, ..., b_K)'$ is a vector composed of the coefficient matrix associated to the Z,

$P = (P(y_1 = 1), ..., P(y_n = 1))'$ is the vector of posterior probabilities.

Therefore, equation (4) admits a matrix representation under the following form:

$$\log\left(\frac{P}{1-P}\right) = X\beta + Zb \tag{5}$$

According to Ngo and Wand (2004), the estimation of the equation (2) requires firstly the estimation of the logistic regression model with random effects assuming that the random effects vector (b) is normally distributed.

To estimate the parameters β, we used the method of Penalized Quasi-Likelihood (PQL) developed by Breslow and Clayton (1993). The PQL method has been used to identify the functional vector by $Y^* = X\beta + Zb + \Delta(Y - P)$ with $\Delta = diag\{p_i(1-p_i)\}$ and also, to define the functional matrix of weight by $\Sigma = W^{-1} + ZG_\theta Z'$ with $W = diag\{p_i\}$.

The predictive value of the model (5) is given by the following equation:

$$\log\left(\frac{\hat{P}}{1-\hat{P}}\right) = X\hat{\beta} + Z\hat{b} \tag{6}$$

where $\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y^*$, $\hat{\Sigma}^{-1} = diag\{\hat{p}_i\}^{-1} + ZG_{\hat{\theta}}Z'$ and $\hat{b}' = G_{\hat{\theta}}Z'\hat{\Sigma}^{-1}(Y^* - X\hat{\beta})$.

3.3 The estimation results of the model

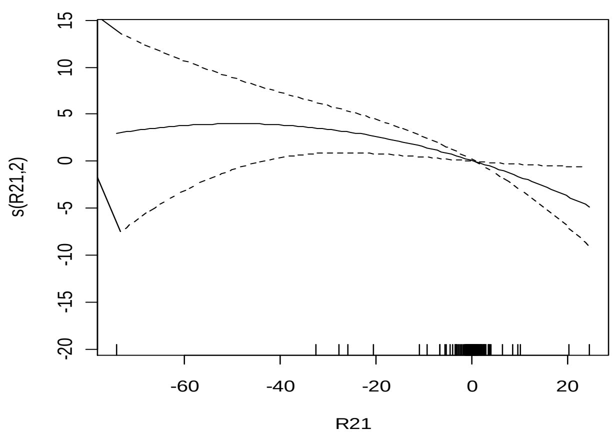
Table (2) summarizes the estimation results of model (6).

Table 2. The coefficients estimation of the semiparametric logistic regression model

	Estimated values	Discriminant power	t-value	p-value
Constant (cst)	-2.0467		-18.687	< 2e-16
R_7 : Asset turnover	0.2795	0.0017	3.846	0.000123
R_9 : Economic profitability	9.8834	0.5277	8.164	5.13e-16
R_{10} : Rate of return on assets	-12.4510	0.4674	-8.210	3.53e-16
R_{14} : Rate of return on equity	0.03277	0.0000	1.742	0.081566
R_{20} : Immobilisation coverage by equity capital	-0.19147	0.0000	-4.691	2.87e-06
R_{23} : Financial expenses/total debt	-0.83057	0.0031	-2.538	0.011203

The discriminant power of R_k is calculated based on this formula: $\frac{\sigma_k^2\beta_k^2}{\sum\sigma_k^2\beta_k^2}$ with σ_k is the standard deviation of the ratio R_k. This indicator reflects the importance of ratio in the score function. Based on Table 2, the ratios R_9 and R_{10} play a fundamental role in the score function . In fact, the two ratios present the highest discriminative power (52.77% and 46.74% respectively for R_9 and R_{10}) therefore a total equal to more than 99%. Moreover, the estimated effect of the variable R_9 has a positive sign (9.8834). As the economic profitability ratio is the ratio between the total of financial expenses and net income over total assets implies that an increase in financial costs reduced profitability and therefore increases the probability of distress. And for the variable R_{10} that equal to the ratio between net income and total assets, has a negative sign (-12.4510) indicating that an increase in net income improves the profitability and subsequently reduces the risk of default.

Figure 2. The curve of the estimated function $f(R_{21})$



The figure (2) shows the curve of the estimated function $f(R_{21})$ with a confidence interval of 95%. The graph of $f(R_{21})$ can be approximated by polynomial function of the 2nd degree.

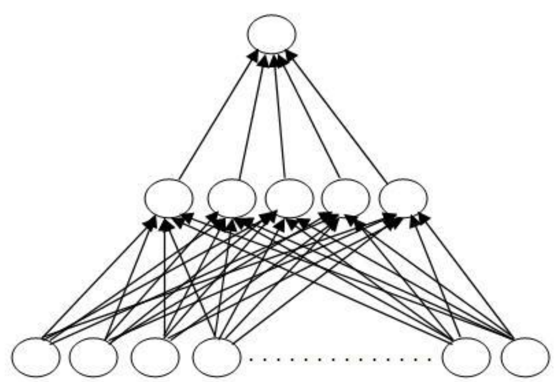
Furthermore, the non-parametric part of model (2) can detect a threshold effect ratios on the probability of being in distress. According to this graph, for a threshold lower than 1, the probability of distress is an increasing function of the long and medium term debt capacity. For a threshold higher than 1, the probability of distress becomes a decreasing function.

4. Analysis via artificial neural networks model

4.1 Overview of the neural network

Artificial neural network is a nonlinear model inspired from the human brain function. This model as shown in Figure 3 consists generally of an input layer, one or more hidden layers and output layer.

Figure 3. Standard architecture of Artificial neural network



The output function calculates the output value of a neuron under the following equation :

$$y_k = g_2\left[\sum_{j=0}^{m} w_{kj}^{(2)} g_1\left(\sum_{i=0}^{N} w_{ij}^{(1)} x_i + b_1\right) + b_2\right] \tag{7}$$

Where; x is the input variable of the network; N is the number of input variables; m is the number of neurons in the hidden layer; k is the number of neurons in the output layer; g is the Transfer/activation function; $w^{(1)}$ is the weights matrix of the hidden layer; $w^{(2)}$ is the weights matrix of the output layer; b_1 = bias of the hidden layer and b_2 = bias of the output layer.

4.2 Neural networks design

In this application, we have used only 80% of the total observations for the training and the remaining (613 observations) to test the predictive ability of the artificial neural network model. The network was trained based on the 7 selected ratios considered as inputs of the model. For the hidden layer, the transfer function used is the sigmoid function as the most used in financial application according to McNelis (2005). And the linear function was used for the output layer.

5. Evaluation of models performance

To assess the discriminative or predictive ability of the established models, two validation metrics are used: the misclassification rate and the Area Under Curve (AUC) as a derived measure from the ROC "Receiver Operating characteristic" curve (see Pepe,2000 for more details).

5.1 The rate of misclassification

The rate of misclassification is the most used measure to evaluate the predictive ability of the model. Based on confusion matrix (see Table 3), we can calculate this indicator by dividing the total number of misclassification (false prediction) on the total of test sample (see Table 4).

Table 3. Confusion matrix

		Predicted class		Total
		$\hat{Y}=1$	$\hat{Y}=0$	
Actual class	$Y=1$	n_{11}	n_{10}	n_1
	$Y=0$	n_{01}	n_{00}	n_0
1 : Bankrupt firm 0 : Healthy firm				

Table 4. Confusion matrix of the estimated models for the test sample

	Artificial neural networks			Semi-parametric logistic regression		
	$\hat{Y}=1$	$\hat{Y}=0$	Total	$\hat{Y}=1$	$\hat{Y}=0$	Total
$Y=1$	29	8	37	24	14	38
$Y=0$	60	516	576	65	510	575
Rate of misclassification: $(n_{10}+n_{01})/(n_0+n_1)$	0.111			0.129		

According to Table 4, neural networks has the lowest rate of misclassification (88.9% of the data are correctly predicted whereas 87.1% for Semi-parametric logistic regression model) .

5.2 The Area Under Curve (AUC)

The area under curve (AUC) reflects the discrimination quality of the model. Numerically, more the AUC is close to unity more the model can predict better the risk of distress. The AUC of neural networks is equal to 0.871 and better than the AUC of semiparametric logistic regression model (0.763).

6. Conclusions

In this study, we applied two techniques in order to predict bankruptcy of Tunisian firms. The first technique is the semiparametric logistic regression and the second is an artificial intelligence tool "the artificial neural network". Based on 3065 credit files, an empirical investigation has been elaborated. The empirical results show that the artificial neural network outperforms the semiparametric logistic regression model in terms of prediction accuracy rate and AUC value. Therefore, artificial neural network is a very promising tool for bankruptcy prediction for firms.

References

Altman, E. I. (1968) ‘‘Financial ratios, discriminant analysis and the prediction of corporate bankruptcy’’ *The Journal of Finance*, 23(4) :589–609.

Beaver, W. H. (1966) ‘‘Financial ratios as predictors of failure’’ *Journal of Accounting Research*, 4 :71–111.

Breslow, N., Clayton, D. G. (1993) ‘‘Approximate Inference in Generalized Linear Mixed Models’’ *Journal of the American Statistical Association* 88 :9 – 25.

Brumback, B. A., Ruppert, D., Wand, M. P. (1999) ‘‘Variable selection and function estimation in additive nonparametric regression using a data-based prior: Comment’’ *Journal of the American Statistical Association*, 94, 794–7.

Chih-Fong, T., Jhen-Wei, W. (2008) ‘‘Using neural network ensembles for bankruptcy prediction and credit scoring’’ *Expert systems with Applications*, n°34, 2639-2649.

Deakin, E. B. (1972) ‘‘A discriminant analysis of predictors of business failure’’ *Journal of Accounting Research*, 10(1), 167–179.

Ghatge, A. R., Halkarnikar, P.P. (2013) ‘‘Ensemble Neural Network Strategy for Predicting Credit Default Evaluation’’ *International Journal of Engineering and Innovative Technology*, Vol 2, Issue 7, pp. 223-225.

Giordani, P., Jacobson, T., Schedvin, E.V., Villani, M. (2013) ‘‘Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios’’ *Sveriges Riksbank Working Paper Series*, No. 256.

Khashman, A. (2011) ‘‘Credit risk evaluation using neural networks: Emotional versus conventional models’’ *Applied Soft Computing*, 11, 5477–5484.

McNelis, P. D. (2005) ‘‘Neural Networks in Finance: Gaining Predictive Edge in the Market’’ (*Academic Press Advanced Finance*).

Mestiri, S., Hamdi, M. (2012) ‘‘Credit Risk Prediction: A comparative study between logistic regression and logistic regression with random effects’’ *International Journal of Management Science and Engineering Management*, 7(3): 200-204.

Ngo, L., Wand, M. (2003) ‘‘Smoothing with mixed model software’’ *Journal of Statistical Software*, 4(1) :1–54.

Odom, M., Sharda, R. (1990) ‘‘A neural network model for bankruptcy prediction’’ *In Proceedings of the international joint conference on neural networks* Vol. 2, IEEE Press, Alamitos, CA (p. 163–168).

Ohlson, J. A. (1980) ‘‘Financial ratios and the probabilistic prediction of bankruptcy’’ *Journal of Accounting Research*, 18(1), 109–131.

Pang, S. (2006) ‘‘Application of Logistic Regression Model in Credit Risk Analysis’’ *Mathematics in Practice and Theory*, (9), 129–137.

Pepe, M. S. (2000) ‘‘Receiver operating characteristic methodology’’ *Journal of the American Statistical Association*, 95(449) :308–311.

Perez, M. (2006) ‘‘Artificial neural networks and bankruptcy forecasting: a state of the art’’ *Neural Computing & Applications*, n°15, 154-163.

Wilson, R. L., Sharda, R. (1994) ‘‘Bankruptcy prediction using neural networks’’ *Decision Support Systems*, 11, 545–557.

Xie, J. G., Wang, J., Qiu, Z. D. (2004) ‘‘Effectiveness of Neural Networks for Prediction of Corporate Financial Distress in China’’ *Advances in Neural Networks, Lecture Notes in Computer Science* (3174), pp 994-999.

Yim, J., Mitchell, H. (2005) ‘‘A comparison of corporate distress prediction models in Brazil: hybrid neural networks, logit models and discriminant analysis’’ *Nova Economia* 15 (1) 73-93.

Zhang, D. , Lin, X. (2003) ‘‘Hypothesis testing in semiparametric additive mixed models’’ *Biostat*, 4(1) :57–74.

Zhang, G., Hu, M.Y., Patuwo, B.E., Indro, D.C. (1999) ‘‘Artificial neural networks in bankruptcy prediction:General framework and cross-validation analysis’’ *European Journal of Operational Research* 116, 16-32.