Modelling Risk of Non-Repayment of Bank Credit by the Method of Scoring

Jimbo Henri Claver[1*], Ngongo Isidore Séraphin[2], Dongmo Tsamo Arthur[3], Andjiga Gabriel Nicolas[3] and Etoua Rémy Magloire[4]

[1]Department of Applied Mathematics and Statistics, AUAF & Waseda University, Tokyo, Japan
[2]Department of Mathematics, ENS, University of Yaoundé I , Cameroon
[3]Department of Mathematics, University of Yaoundé 1, Cameroon
[4]Department of Mathematics, Higher National Polytechnic School. Yaoundé 1, Cameroon
Email: jimbo.maths@gmail.com

Abstract The risk of non-repayment of bank credit is a variable that banks are seeking to master in order to save their profitability and to protect themselves against bankruptcy. In this article, we have shown how to model the risk using tools for the decision making purposes using mathematical techniques of the method of Scoring. We construct a score function capable of minimising the probability that a client may not repay the credit at the fixed date. The construction of such function is done through Fischer discriminant analysis and the logistic regression. The methodology used inthis work relies on statistical analysis techniques and the probability of Scoring. Finally we applied our approach to a given company and found that the risk of non-repayment of the bank credit depends mainly on the loans ratios, global cash flow, global indebtedness, capital funds and net result, capital funds.

Keywords: Banks, risks of non-repayment, method of scoring, Fischer discriminant analysis, logistic regression, score function.

1 Introduction

The risk of non-repayment is the most ancient form of risk on the capital market and the most known in the world. It is defined as the risk to see one's client not respecting their financial engagement at the fixed date [1,2]. It nowadays constitutes a capital stake for bank institutions, for a non-repaid debt can influence the survival of a company in the short or long run, as well as it constitutes a loss economically, supported by the creditor. It is then necessary to set an efficient system that can help bank institutions to considerably reduce the risk linked credits. Thus, how can banks minimise the probability that a client may not respect their engagements? To succeed, the banker must find a way to distinguish the potential good clients to bad ones. Which needs to have at his disposal on one hand the information on the clients and on the other hand, the mathematical tools more precisely statistic and probabilistic offering him more details to make a decision. The mathematical tool which offer an easy evaluation regarding the management of risk of non-repayment of credits is based on the method of Scoring [3]. It is defined as a set of techniques which helps in the decision making: such as helping or not credit to a person or to a company [4]. This approach permits to establish a background profile of good and bad clients from which we can easily classify a new client within a group of membership: the group of good clients and the group of bad clients.

The pioneer of the method of Scoring is Beaver [5], it was about to start with the development of other models such as the Z-score of Altman [6], the logit and probit regression models, the neurons networks, the Scoring by genetic analysis [7,8], which were then explore for the prediction.

In this article, we have shown how to model a helping tool to the decision making by using mathematical techniques of the method of Scoring. Specifically, we have constructed an explicative model of credit risk of non-repayment and we have deducted the determiners of the reliability of the bank credit. This study is laid upon a hypothesis according to which the risk of non repayment of bank credit depends on characteristics proper to clients. This article is subdivided in 6 sections: Section 2 is devoted to Fischer discriminant analysis, section 3 is dedicated to the logistic regression, section deals with the computer application of techniques aforementioned as well as the interpretation of obtained results, for section 5 is reserved to the conclusion and in the end section 6 is for the Appendix.

2 Fischer Discriminant Analysis

Given a sample of n individuals collected from a population of two groups: the group of good clients that will be noted 0 and that of bad clients that will be noted 1. Let's measure simultaneously a variable to explain Y and p independent and continuous explicative variables $X = (X_i)_{1 \leqslant i \leqslant p}$.
Let's note by:

- $\mathbb{P}_k = \mathbb{P}(Y = k)$ the prior probability, i.e the probability for an individual to be in the group k, $k = 0, 1$.
- $\mathbb{P}(Y = k \mid X)$ the posteriori probability, i.e the probability for an individual to be in a group k, knowing the variable X which explains the membership.
- $f_k : \mathbb{R}^p \longmapsto [0, 1]$] the conditional density of $X \mid Y = k$ for the measure of Lebesgue over $\mathbb{R}^p$. f_k is the density of the variable in the group k.

The density of the variable X can be expressed as a mixture of conditional density. It is given by:

$$f_X = \sum_{i=0}^{1} \mathbb{P}_i f_i. \tag{1}$$

We formulate the following hypothesis:

$\mathcal{H}_1$: The conditional law of X knowing that Y follows a multi-normal of average and of the matrix of variance-covariance $\sum$. $X \mid Y = k \rightsquigarrow \mathcal{N}_n(\mu_i, \sum)$.

Either $x \in \mathbb{R}^p$ a new observation of the explicative variable X. We wish to predict the variable Y associated to a new observation x of the explicative variable.

2.1 Rule of Bayesian Decision

It permits to attribute a new observation x, the group of membership the most probable for this; i.e, the group for which the posteriori probability is maximal [9]. It is given by:

$$g_k = argmax\ \mathbb{P}(Y = k \mid X = x). \tag{2}$$

Theorem 2.1. *The rule of Bayesian Decision is given by [10]:*

$$g_k = argmax\ \mathbb{P}_k f_k(x). \tag{3}$$

where under the hypothesis $\mathcal{H}_1$,

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}} exp\left[\frac{-1}{2}(x - \mu_k)^\top \sum{}^{-1}(x - \mu_k)\right]. \tag{4}$$

Theorem 2.2. *The rule of Bayesian Decision is written:*

$$g_k = argmax\ d_k(x). \tag{5}$$

with

$$d_k(x) = \ln(\mathbb{P}_k) - \frac{1}{2}(x - \mu_k)^\top \sum^{-1}(x - \mu_k). \tag{6}$$

The Proof of the theorem is found in Appendix

2.2 Estimation of Parameters $\mathbb{P}_k$, μ_k, $\sum$

Practically speaking, we do not know the $\mathbb{P}_k$, the μ_k, and $\sum$. In this section, we will estimate the parameters of the model. Let's consider the parameter θ, defined by $\theta = (\mathbb{P}_k, \mu_k, \sum)$. We can estimate the parameter θ.

Proposition 2.3. *The estimators of parameter θ obtained by the method of maximum of likelihood are:*

$$\begin{cases} \widehat{\mathbb{P}}_k = \frac{n_k}{n}. \quad where \;\; n_k = card(k) \\ \widehat{\mu}_k = \frac{1}{n_k} \sum_{i \in k} x_i \\ \widehat{\sum} = \frac{1}{n} \sum_{k=0}^{1} \sum_{i \in k} (x_i - \widehat{\mu}_k)(x_i - \widehat{\mu}_k)^\top \end{cases} \tag{7}$$

Let's note that the estimator of $\sum$ is biased hence, the estimator without bias of $\sum$ is:

$$\widehat{\sum} = \frac{1}{n-2} \sum_{k=0}^{1} \sum_{i \in k} (x_i - \widehat{\mu}_k)(x_i - \widehat{\mu}_k)^\top. \tag{8}$$

2.3 Estimation of the Posteriori Probability

Proposition 2.4. *The posteriori probability is estimated by:*

$$\widehat{\mathbb{P}}(Y = k \mid X = x) = \frac{exp[\widehat{d}_k(x)]}{\sum_{k=0}^{1} exp[\widehat{d}_k(x)]}. \tag{9}$$

where

$$\widehat{d}_k(x) = \frac{1}{2}(x - \widehat{\mu}_k)^\top \widehat{\sum}^{-1}(x - \widehat{\mu}_k) + g(x). \tag{10}$$

and

$$g(x) = \begin{cases} \ln(\mathbb{P}_k) & \textit{If the prior probabilities are not equal;} \\ 0 & \textit{Or else.} \end{cases}$$

The proof of the proposition is found in Appendix.

In general, the estimation of the posteriori probability is called score function. We then affect a new observation to the group which has a score is maximal. In general, the score function noted $S(X)$, obtained by the linear discriminant analysis is expressed as follows:

$$S(X) = \sum_{i=1}^{p} a_i X_i + b. \tag{11}$$

where

- a_i are the associated coefficients to the variables X_i.
- b is a constant.

We wish to have a quantitative measure between 0 and 1; to succeed we shall use the a posteriori probability. Thus, every new observation will have a mark which from a threshold will be assigned to a group of membership.

2.4 Score and Theoretical Threshold

A new observation x will be classified in the group 1 when $g_0 < g_1$.

$$g_0 < g_1 \Leftrightarrow \mathbb{P}(Y=0 \mid X=x) < \mathbb{P}(Y=1 \mid X=x) \quad (12)$$

$$\Leftrightarrow \mathbb{P}_0 f_0(x) < \mathbb{P}_1 f_1(x) \quad (13)$$

where

$$\begin{cases} f_0(x) = \dfrac{1}{(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}} exp\left[\frac{-1}{2}(x-\mu_0)^\top \sum^{-1}(x-\mu_0)\right] \\ f_1(x) = \dfrac{1}{(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}} exp\left[\frac{-1}{2}(x-\mu_1)^\top \sum^{-1}(x-\mu_1)\right] \end{cases} \quad (14)$$

In the process of the natural logarithm (which does not change the rule of transfer) in every equation of the system **(14)**, we have:

$$\begin{cases} S(x) = x^\top \sum^{-1}(\mu_0-\mu_1) \\ s_0 = \ln`\left(\frac{\mathbb{P}_1}{\mathbb{P}_0}\right) + \frac{1}{2}(\mu_0+\mu_1)^\top \sum^{-1}(\mu_0-\mu_1) \end{cases} \quad (15)$$

An individual will be assigned to the group 1 when $S(x) \geqslant s_0$ and will be assigned to the group 0 when $S(x) < s_0$. The Calculus of the threshold separating the choice of $\mathbb{P}_0$ from the choice of $\mathbb{P}_1$ is delicate since the prior probabilities are not always known. By formulating the hypothesis according to which the prior probabilities are equal to $\frac{1}{2}$, the threshold becomes $s_0 = \frac{1}{2}(\mu_0+\mu_1)^\top \sum^{-1}(\mu_0-\mu_1)$ and the rule of transfer remains unchanged.

2.5 Selection of the Significant Variables

The selection of significant variables can be seen as the research of an optimum model. For that and at the image of the optimizing possibilities, one can either carry out an exhaustive research (since the number of possible models is finished) logit and probit, or do a test, in the circumstances the Lambda test of Wilk's, in order to determine the most significant variables. It is interpreted as follows:

Table 1. Interpretation of tests for the selection of variables

	significant variables	Non significant variables
Fischer test	The statistic F is high	The statistic F is weak
Lambda of Wilk's	p-value lower to the threshold fixed α	p-value higher to the threshold fixed α

2.6 Validation of the Model

Wilk's lambda test The statistics of the Lambda of Wilk's noted Λ is the ratio of determinant of the matrix of variances covariances intra-group (W) to the determinant of the amount of the matrices of variances covariances inter (B) and intra-group. So it is given by:

$$\Lambda = \frac{\mid W \mid}{\mid B+W \mid}. \quad (16)$$

This report is comprised between the numbers 0 and 1; where 0 means that the explicative variables explain perfectly he variable Y , and 1 means that the explicative variables does not explain the model. Thus, the nearer of 0, the lambda of Wilk's is, the better is the model. In order to be able to report to the table of the law of Fischer-Snedecor (Because of the rarity of the table of the Law of Wilk's in the softwares of statistics, we can apply to it the transformation of Roa in order to reformulate the statistics of the test, which will be:

$$F_{\Lambda} = \frac{1-\Lambda}{\Lambda}. \tag{17}$$

This statistic of the test follows the law of Fischer-Snedecor of parameter $(p, n-p-1)$. The hypothesis $\mathbf{H_0}$: **The obtained model is adequate**, is rejected when the p-value of the statistic of the Wilk's Lambda Test is higher to the threshold α fixed, and accepted in the opposite case.

Box M test It is a parametric approach which permits to test if the matrices of variances-covariances associated at $X \mid Y = k$ are equal. The statistics of the M of Box is defined by:

$$M = (n-2)\ln \mid H \mid - \sum_{k=0}^{1}(n_k - 1)\ln \mid V_k \mid . \tag{18}$$

where

$$H = \frac{1}{n-2}\sum_{k=0}^{1}(n_k-1)V_k. \tag{19}$$

In order to be able to report the statistics of M of Box at the table of the Law of $\mathcal{X}^2$, it is appropriate to apply to it the following transformation: $\mathcal{X}^2 = M(1-\delta)$, where

$$\delta = \frac{2p^2+3p+1}{6p+1}\left[\sum_{k=0}^{1}\frac{1}{(n_k-1)} - \frac{1}{n-2}\right]. \tag{20}$$

The statistics of M of Box, follows then a law of Khi two to $\frac{p(p+1)}{2}$ degrees of liberty. The M of Box must be The highest possible and the hypothesis $\mathbf{H_0}$: **The matrices of variances covariances of group k are equal**, is rejected when the p-value of the statistic of the Box M Test is higher to the threshold α fixed, and accepted in the opposite case.

Canonic correlation The canonic Correlation is the root of the variance between the divided groups by the total variance of the discriminant function [11]. It helps to evaluate the relation between several variables. Moreover, it helps to measure the proportion of the variation of the discriminant function due to the difference between groups. A correlation equal to 0 indicates that the variations of the discriminant function are not lined to groups; and a correlation equal to 1 means that all the variations of the discriminant function are linked to groups. In general, the nearer of 1, the lambda of Wilk's is, the better is the model.

3 Logistic Regression

Also called binomial model of the multiple linear regression model, it is as the discriminant analysis, a statistic method that we use to study the existing relation between a qualitative variable to explain with two modalities and several explicative variables. It differs from that by the fact that the variable to explain is conditioned by the explicative variable following a binomial law. We seek the a posteriori probability $\mathbb{P}(Y = k \mid X = x)$ where $k \in \{0, 1\}$ and x is an observation of the explicative variable. Since we model the discrete probabilities, we then have a constraint which is: $\sum_{j=0}^{1}\mathbb{P}(Y = k \mid X = x) = 1.$. So the Knowledge of $\mathbb{P}(Y = 1 \mid X = x)$ induct immediately that of $\mathbb{P}(Y = 0 \mid X = x)$. Since $\mathbb{P}(Y = k \mid X = x)$ is a probability then it is always included between 0 and 1. Thus every attempt of adjusting a cloud of points of probability using a line will be invalid since the line is not bound. In order to solve that, Instead of modelling $\mathbb{P}(Y = k \mid X = x)$, we shall rather model the report $\frac{\mathbb{P}(Y=1 \mid X=x)}{\mathbb{P}(Y=0 \mid X=x)}$. For the following part, we formulate the fundamental hypothesis following:

$$\mathcal{H}_{\mathbf{1}} : \ln\left(\frac{\mathbb{P}(X = x \mid Y = 1)}{\mathbb{P}(X = x \mid Y = 0)}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots b_px_p. \tag{21}$$

The logistic regression model [15] is defined as follows:

$$\ln\left(\frac{\mathbb{P}(Y=1 \mid X=x)}{1-\mathbb{P}(Y=1 \mid X=x)}\right)=\beta_0+\beta_1 x_1+\cdots\cdot\beta_p x_p=x^{\top}\beta \tag{22}$$

or simply

$$Logit(\mathbb{P}(Y=1 \mid X=x))=x^{\top}\beta. \tag{23}$$

where

- Y is the binear explicative variable with two modalities 0 and 1,
- $X=(1,X_1,X_2,\cdots\cdot X_p)^{\top}$ are p explicative variables,
- $\beta=(\beta_0,\beta_1,\cdots\cdot,\beta_p)$ is a united matrix column of dimension $(p+1)\times 1$ of general term $(\beta_i)_{0\leqslant i\leqslant p}$ and with values in $\mathbb{R}$,
- $Logit: p\mapsto \ln\left(\frac{p}{1-p}\right)$ is the bijective derivable of $]0,1[$ in $\mathbb{R}$. It is the function of link which permits To express the $\ln\left(\frac{\mathbb{P}(Y=1 \mid X=x)}{1-\mathbb{P}(Y=1 \mid X=x)}\right)$ Considering the explicative variables.

3.1 Estimation of the Parameters $\boldsymbol{\beta}$

We want to estimate the unknown parameters β. This estimation will be done by using the method of the maximum of likelihood. We are in presence of n observations of variables $\{X_{i1},\cdots,X_{ip},Y_i\}_{i=1,n}$ whose i^{eme} is noted (x_i,y_i), $y_i\in\{0,1\}$. The conditional likelihood of $Y \mid X=x_i$ associated to the observation i is written:

$$L(y_i,\beta)=\mathbb{P}(Y=1 \mid X=x_i)^{y_i}[1-\mathbb{P}(Y=1 \mid X=x_i)]^{1-y_i}. \tag{24}$$

And so, the conditional likelihood of the model is of the form:

$$L(y,\beta)=\prod_{i=1}^{n}\mathbb{P}(Y=1 \mid X=x_i)^{y_i}[1-\mathbb{P}(Y=1 \mid X=x_i)]^{1-y_i} \tag{25}$$

Let's note by $\mathcal{L}(y,\beta)$ the log-likelihood of the model. In the process of the natural logarithm, in the equation 25 we obtain,

$$\mathcal{L}(y,\beta)=\sum_{i=1}^{n}\left\{y_i x_i^{\top}\beta-\ln[1+exp(x_i^{\top}\beta)]\right\}. \tag{26}$$

To have the maximum $\widehat{\beta}$ de β, we must cancel the gradient of $\mathcal{L}(y,\beta)$. practical point of view, this is not easy because of the macro form $\mathcal{L}(y,\beta)$ and its big size n. We then need numerical methods of optimisation to obtain the value of the estimator of β. The algorithm of Newton - Raphson and the Fischer score algorithm permit to calculate the estimator of the maximum of likelihood.

3.2 Estimation of the Probability $\mathbb{P}(\boldsymbol{Y=1 \mid X=x})$

Proposition 3.1. *The estimator of* $\mathbb{P}(Y=1 \mid X=x)$ *est:*

$$\widehat{\mathbb{P}}(Y=1 \mid X=x)=\frac{exp(x^{\top}\widehat{\beta})}{1+exp(x^{\top}\widehat{\beta})}. \tag{27}$$

The demonstration of the proposal is found in Appendix

3.3 Score and Theoretical Threshold

To determine the membership class of the new observation x, we compare the probabilities, $\widehat{\mathbb{P}}(Y = 1 \mid X = x)$ and $\widehat{\mathbb{P}}(Y = 0 \mid X = x)$. The report of these probabilities is given by:

$$\frac{\widehat{\mathbb{P}}(Y = 1 \mid X = x)}{\widehat{\mathbb{P}}(Y = 0 \mid X = x)} = \frac{\widehat{\mathbb{P}}(Y = 1 \mid X = x)}{1 - \widehat{\mathbb{P}}(Y = 1 \mid X = x)}. \tag{28}$$

By using the function of *Logit* link, we have:

$$\ln\left(\frac{\widehat{\mathbb{P}}(Y = 1 \mid X = x)}{\widehat{\mathbb{P}}(Y = 0 \mid X = x)}\right) = Logit[\widehat{\mathbb{P}}(Y = 1 \mid X = x)]. \tag{29}$$

The rule of Decision is then:

- $Logit[\widehat{\mathbb{P}}(Y = 1 \mid X = x)] > 0 \Leftrightarrow x^\top \widehat{\beta} > 0$, then the new observation is assigned to the group of bad clients;
- $Logit[\widehat{\mathbb{P}}(Y = 1 \mid X = x)] < 0 \Leftrightarrow x^\top \widehat{\beta} < 0$, then the new observation is assigned to the group of good clients;
- $x^\top \widehat{\beta} = 0$, we can say nothing.

The obtained score function by the logistic regression is scalar, $S(x) = x^\top \widehat{\beta}$. We notice in this case that the threshold appears clearly and it is $s = 0$. In the practice, it is inappropriate to focus on this theoretical threshold; we must then make vary the threshold and keep the one making the forecast better.

3.4 Selection of the Significant Variables

The preliminary step in the selection significant variables is the choice of the criterion which aims at penalizing by a number of parameter the log-likelihood, in order to have reasonable size model only having significant variables. The criteria answering to this specificity [12] are:

– The criteria of choice Akaike Informative Criterion (AIC) for a model p with variables, defined by:

$$AIC = -2\mathcal{L} + 2p. \tag{30}$$

where $\mathcal{L}$ the log-likelihood of the logistic regression model and $2p$ is the function of the number of parameters;

– The criterion of choice Bayesian Informative Criterion (BIC) for a model p with parameters, defined by:

$$BIC = -2\mathcal{L} + p\ln(n). \tag{31}$$

For each competitive model, the criterion of choice is calculated and the model which has the least AIC or BIC is the one we keep. The last step consist to apply a method of selection step by step which must be either: descending, ascending, or stepwise.

Descending method It consists to examine a model including all the p explicative variables, then we eliminate one after one the variables of the model. At each elimination, we calculate the criterion of choice of the model, in order to spot the model. We stop when all the variables are withdrawn of the model or when the criterion of choice diminishes no more. The final model is then a model only having significant variables.

Ascending method It consists to examine the model with only an explicative variable, then we introduce one after one other explicative variables. At each introduction of those variables, we calculate the criterion of choice of the model, in order to spot the model. We stop when all the variables are integrated in the model or when the criterion of choice diminishes no more. The final model is then a model only having significant variables.

Stepwise procedure (progressive method) It permits to sort out the inconvenience of the ascending method: once the variable is introduced, it can not be withdrawn: for a variable considered as significant to a step can at another become less significant. Its philosophy is identical to that of the ascending method.

3.5 Validation of the Model

Test of Hosmer-Lemeshow The null and alternative hypotheses of the test are formulated in the following way: $\mathbf{H_0}$: The model considered is appropriate; $\mathbf{H_1}$: The model considered is not adequate. The principle is the following [13]:

- For each observation x_i, we calculate the forecast probability by the model $\widehat{\mathbb{P}}(Y=1 \mid X=x_i)$, that we classify in ascending order;
- Then we subdivide those probabilities in G same size groups. In general for a very big sample, we attribute to G the value 10. The last group, that of $\widehat{\mathbb{P}}(Y=1 \mid X=x_i)$ the biggest, have a higher number to the others. Then, we note m_g the number of the group g, o_g the number of cases $Y=1$ in the group g, μ_g the average of $\widehat{\mathbb{P}}(Y=1 \mid X=x_i)$ in the group g.

The statistic of the test of Hosmer-Lemeshow is defined as follows:

$$C^2 = \sum_{g=1}^{G} \frac{(o_g - m_g\mu_g)^2}{m_g\mu_g(1-\mu_g)}. \tag{32}$$

Under $\mathbf{H_0}$, the statistics of the test C^2 follows approximatively a law of $\mathcal{X}^2(G-2)$ degrees of liberty. If we designate by $q_{1-\alpha}(G-2)$ the quantity of level $1-\alpha$ of the law of $\mathcal{X}^2(G-2)$ and by C_{calc} the value of C^2, then at the level α, we reject $\mathbf{H_0}$ (i.e that the model is not appropriate) if $C_{calc} > q_{1-\alpha}(G-2)$, and we reject $\mathbf{H_1}$ i.e the model is appropriate if $C_{calc} \leqslant q_{1-\alpha}(G-2)$

Residuals analysis The residual analysis helps to diagnose the quality of the regression model elaborated. It is used within the framework of individual data (the probability that two individuals have the same description is null) and there are three 03 types:

- The actual residuals which have for expression, $\varepsilon = Y_i - \widehat{P}(Y=1 \mid X=x_i)$. They permit to evaluate the quality of adjustment of the model on each observation. But the analysis of variance of the residuals permits us to conclude that one cannot compare them since they do not have the same variance. In order to solve this problem, we standardise the actual residuals by the theoretical variance of Y_i, which helps us to introduce the Pearson residuals.
- The Pearson residuals are the standardised actual residuals by the theoretical variance of Y_i, and are of the form,
$$RP_i = \frac{Y_i - \widehat{P}(Y=1 \mid X=x_i)}{\sqrt{(\widehat{P}(Y=1 \mid X=x_i)(1-\widehat{P}(Y=1 \mid X=x_i))}}. \tag{33}$$
They help to identify the ill modelled points and their values are all the more higher than $\widehat{P}(Y=1 \mid X=x_i)$ is beside 0 or 1. The distributions follow approximatively the reduced normal centred law. Thus, any point out of the interval $[-2,+2]$(at the level of confidence 95%) is absurd and must be analysed [14]. This enables us to have criteria of validation of the model by graphic reading of the Pearson residuals.
- The residuals of deviance, they are defined as follows:
$$RD_i = \sqrt{2Y_i \ln \widehat{P}(Y=1 \mid X=x_i) + (1-Y_i)\ln(1-\widehat{P}(Y=1 \mid X=x_i))}. \tag{34}$$
In logistic regression they are those which are generally used since they approach better the normal law than the Pearson residuals. They take their values in the interval $[-2,+2]$(at the level of confidence 95%) and the model is valid for the reasons mentioned before.

4 Application

4.1 Description of the Database

The data used in the framework of study are those of the year 2012, issued from the database published on the website (www.ansd.sn). They are constituted of a sample of 32 companies whose 12 bankrupted. The bankruptcy is observed when the company does not repay totally or partially its debt to its creditor. The reliability of a company is obtained by the variable Y, distributed following the hereafter graphic:

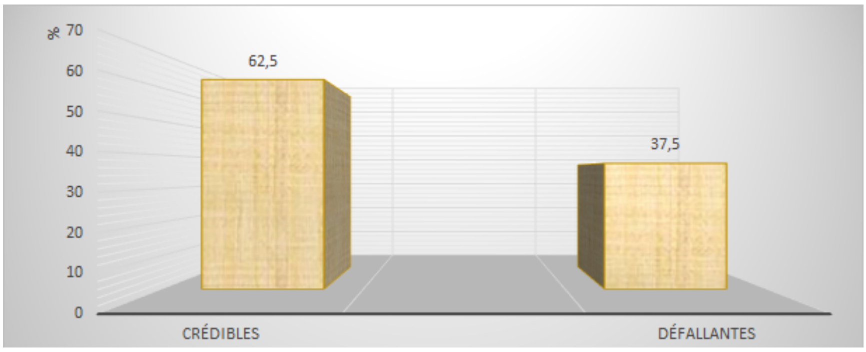

Figure 1. Percentage of credible and bankrupted companies

4.2 Choice of the Ratios

A ratio is an indicator used to conduct the financial analysis of companies. It is the result of a division between two elements issued from the results account, the balance sheet, the stock information, and it is represented by a coefficient or a percentage. The table below summarizes the 13 ratios that we have kept for our study.

Table 2. Table of ratios

Ratio	Formula
R1	Charges of the staff / global added value
R2	Gross operating profit / turnover
R3	Operating results or profit/ gross added value
R4	Loans/ global cash flow
R5	Global indebtedness/ capital funds
R6	Capital funds/ loans
R7	Stable resources/ permanent assets
R8	Financial debt/ total liabilities
R9	Permanent capital/ long run debt
R10	Stockholder's equity/ total liabilities
R11	Cash flow/ financial debts
R12	Clients credits/ turn over AIT
R13	Net profit/ stockholder's equity

4.3 Study of the Collinearity of Ratios

The study of the collinearity of ratios is the very first step in the construction of a score function for the correlated ratios bring the same information; thus will affect the robustness and the performance of the elaborated model. To study this collinearity, we shall have recourse to the cloud of points for it permits to indicate the degree of correlation between two or several linked variables. The reading of this graph

Figure 2. Cloud of points two after two of data

permits to suspect a strong connection between the ratios R6 and R9. To confirm or contradict that, we carry out the Spearman test on the data and we obtain the matrix of the following correlation coefficient: We notice that the data of this matrix (except the data in diagonal) belong to the interval. $[-0,48,0,44]$.

Table 3. Matrix of Spearman correlation

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13
R1	1.00	-0.27	-0.34	0.15	-0.09	-0.16	0.15	0.18	-0.23	-0.35	-0.35	0.27	-0.07
R2	-0.27	1.00	0.34	-0.39	-0.33	-0.01	0.21	0.15	0.29	0.35	-0.00	0.09	0.15
R3	-0.34	0.34	1.00	-0.36	-0.08	0.31	-0.11	0.00	0.25	0.44	0.02	-0.48	0.27
R4	0.15	-0.39	-0.36	1.00	0.10	-0.21	-0.00	-0.35	-0.26	-0.36	-0.10	0.02	-0.36
R5	-0.09	-0.33	-0.08	0.10	1.00	0.11	-0.35	-0.13	-0.02	-0.15	0.27	0.01	-0.14
R6	-0.16	-0.01	0.31	-0.21	0.11	1.00	-0.41	-0.25	0.41	0.18	-0.07	-0.11	0.13
R7	0.15	0.21	-0.11	-0.00	-0.35	-0.41	1.00	0.41	-0.30	-0.07	0.09	0.37	-0.02
R8	0.18	0.15	0.00	-0.35	-0.13	-0.25	0.41	1.00	-0.10	0.11	0.33	0.26	0.00
R9	-0.23	0.29	0.25	-0.26	-0.02	0.41	-0.30	-0.10	1.00	0.14	0.15	0.04	0.20
R10	-0.35	0.35	0.44	-0.36	-0.15	0.18	-0.07	0.11	0.14	1.00	-0.03	-0.13	0.34
R11	-0.35	-0.00	0.02	-0.10	0.27	-0.07	0.09	0.33	0.15	-0.03	1.00	0.07	0.14
R12	0.27	0.09	-0.48	0.02	0.01	-0.11	0.37	0.26	0.04	-0.13	0.07	1.00	0.02
R13	-0.07	0.15	0.27	-0.36	-0.14	0.13	-0.02	0.00	0.20	0.34	0.14	0.02	1.00

We conclude then that the ratios are weakly correlated between them.

4.4 Score Function Obtained by Fischer Discriminant Analysis

To verify the existence of a strong relation between the ratios and the membership to a group, we carry out an equality test of averages of groups; the more discriminating ratios must have higher values for the Fischer test statistic and the p-value of the Wilk's Lambda Test statistics must be lower to a threshold α fixed. The results of the equality test of averages of groups is: In this table, we notice that at the threshold

Table 4. Equality test of averages of groups

Ratios	Lambda.de.Wilks	F	ddl1	ddl2	Signification
R1	0.98	0.44	1	30	0.51
R2	1.00	0.01	1	30	0.94
R3	0.95	1.68	1	30	0.20
R4	0.98	0.47	1	30	0.50
R5	0.97	0.87	1	30	0.36
R6	0.95	1.44	1	30	0.24
R7	0.88	4.20	1	30	0.05
R8	0.94	1.89	1	30	0.18
R9	0.97	0.79	1	30	0.38
R10	0.99	0.20	1	30	0.66
R11	0.97	0.77	1	30	0.39
R12	0.81	6.83	1	30	0.01
R13	0.95	1.56	1	30	0.22

of significance of 10have a p-value lower to and a higher Fischer test statistics, which shows that they differentiate well the sound companies to bankrupted companies. We conclude then that the ratios are more discriminating.
To establish the equation of the score function, we have estimated the coefficients of the discriminating function from the following table: So, the score function is written:

$$S(x) = 0,458\ R7 + 0,75\ R12. \tag{35}$$

The following table shows that there is not equality between the matrices variances covariances of the groups. Indeed at the threshold of significance 5%, the p-value of the statistics of the M Box Test is higher than $0,05$. Therefore, there is no equality of matrices of variances of covariances between the groups. Moreover, with the canonic Correlation coefficient which is of 46,80%, we can conclude that the score function constructed by the Fischer discriminant Analysis is not valid (although the

4.5 Score Function Obtained by Logistic Regression

The constructed model has retained the ratios, R2, R3, R4, R5, R7, R8, R10, R11 and R13 as significant. The estimated coefficients of those ratios are presented in the following table:

Different values of Pr(> |z|), we notice that the coefficients which are statistically null at a level of confidence of 90% are those of the ratios R2, R3, R7, R8, R10. Thus, the Score function obtained by logistic regression is:

$$S(X) = -20,94 + 1,78\ R4 + 2,61\ R5 + 19,57\ R13. \tag{36}$$

The model being constructed, only awaiting validation now. To do so, we have recourse to the residuals Analysis of deviance and ot the hosmer-Lemeshow test.

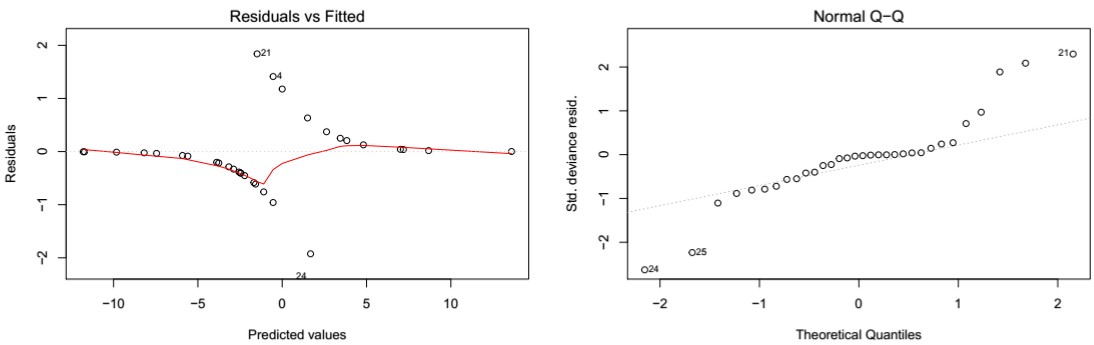

Figure 3. Curve of the residuals of deviance

Residuals of deviance analysis On the Residuals vs Fitted, we notice that no atypical point is found; all the points, are well situated in the interval $[-2, +2]$ moreover, the normal shows that the Residuals of deviance follow a law of Gauss (since the majority of points are situated around the dotted line). Hence, we conclude that the elaborated model is valid.

Hosmer Lemeshow test The result of the Hosmer Lemeshow Test is the following: At the threshold of significance of 5%, the adjustment of the model is good as the p-value of the statistic of the test of chi square to 8 degrees of liberty is lower than 5%. Therefore, the hypothesis $\mathbf{H_1}$: The considered model is not appropriate, is rejected we then conclude that the model is appropriate so valid.
Being given that the Fischer analysis has not been conclusive, it follows that the best model for our database is that obtained by logistic regression.

5 Conclusion

In this article, we have shown how to model a helping tool to the decision making by the Fischer discriminant analysis and the logistic regression, which are techniques founded on statistics and probabilistic methods of the method of Scoring. We have presented the theoretical and the practical construction of the

Table 5. Coefficients of canonic discriminating functions

Fonction 1	
R12	0,75
R7	0,458

Table 6. M Box test of the equality of matrices of covariances

	M de Box	2.987
F	Approximativement	0.92
	ddl1	3.00
	ddl2	17832.40
	Signification	0.43

Table 7. Canonic correlation and proper values

Function	proper.value	% of the variance	% cumulated	canonic correlation
1	0.28	100.00	100.00	0.47

Table 8. Wilk's Lambda test

Test of the function(s)	lambda of Wilk's	khi-deux	ddl	signification
1	0.78	7.15	2	0.03

Table 9. Results of the logistic regression model

Ratios retenus	Estimate	Std. Error	z value	Pr(> |z|)	$IC_{90\%}$
(Intercept)	-20.94	11.02	-1.90	0.06	-39.06 | -2.81
R2	61.59	40.47	1.52	0.13	-4.99 | 128.16
R3	-16.11	12.21	-1.32	0.19	-36.19 | 3.98
R4	1.78	0.90	1.99	0.05	0.31 | 3.26
R5	2.61	1.25	2.09	0.04	0.55 | 4.66
R7	9.14	5.51	1.66	0.10	0.08 | 18.20
R8	18.10	11.61	1.56	0.12	-1.00 | 37.20
R10	-23.87	18.05	-1.32	0.19	-53.56 | 5.82
R11	-11.61	7.36	-1.58	0.11	-23.72 | 0.49
R13	19.54	9.64	2.03	0.04	3.69 | 35.39

Table 10. Hosmer Lemeshow Test at the level of 5%

X-Squared	df	p-value
32	8	$9{,}3.10^{-5}$

score function for each technique. The validation of the regression model has been carried out by the HosmerLemeshow Test and the Residuals of Deviance Analysis. For the Fischer discriminant Analysis, this latter has not been conclusive at the term of the M Box Test and the Canonic Correlation. The obtained score function by the logistic regression is:

$$S(x) = -20,94 + 1,78\ R4 + 2,61\ R5 + 19,54\ R13. \tag{37}$$

It comes out the ratios R4, R5, R13 are the most discriminating ratios enabling to explain the membership of an individual to a modality of Y. The basic hypothesis according to which the risk of non-repayment depends on the characteristics of the client is found then checked by the logistic regression technique.

Acknowledgements. We thank all our colleagues for the discussions and the suggestions which have helped us to improve the presentation of this work.

Conflict of interests. The authors declare no conflict of interest.

References

1. K. Zaghdoudi, N. Djebali, and M. Mezni, "Credit Scoring and Default Risk Prediction," *University of Jendouba*, 2016.
2. G. Hein, "Le contrôle interne du risque de crédit bancaire," *Université de Nice-Sophia Antipolis*, pp. 18, 2002.
3. T. Roncalli, "La gestion des risques financiers," Economica, pp. 32-325, 390-393, 2004.
4. L. Thomas, C. Edelman, and N. Crook, "Credit Scoring and Its Applications," *Philadelphia, society for Industrial and Applied Mathematics*, 2002.
5. W. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, pp. 71-111, 1966.
6. E. Altman, "Financial ratios, discriminant analysis and predictor of corporate Bankruptcy," *The Journal of finance*, pp. 589-609, 1968.
7. A. Saunder, L. Allen, "Credit rist measurements: New approach to value at risk and the other paradigms," New York, 2002.
8. M. Schervish, "Theory of statistics," Springer-Verlag, pp. 459, 1995.
9. M. Chavent, "Scoring," 2014.
10. M. Chavent, "Analyse discriminante linéaire et quadratique," 2015.
11. E. Azzouz, "Credit risk management by the Scoring method: Case of the popular bank of Rabat-Kenitra," Remadem, 2009.
12. C. Chesneau, "Modèles de régression," *Université de Caen*, pp. 69 -87, 2017.
13. R. Rakotamolala, "Pratique de la régression logistique: version 2.2," *Université Lumière de Lyon 2*, 2015.
14. S. Menard, "Applied logistic regression analysis," *University of Colorado, Boulder*, pp. 82, 2002.
15. L. Rouviere, *"Régression logistique avec R," Universités Rennes 2*, 2015.
16. H. Calvet, "Méthodologie de l'analyse financière des établissements de crédit: 2^{e} édition," Economica, Paris, 2002.
17. A. Charbonneau, "La mise en place d'un modèle d'évaluation du risque de crédit dans le cadre de la reforme solvabilité 2," *Université d'Orleans*, page 25-34. 2003.
18. A. Chebill, A. Anderson, "On the existence of maximum likelihood estimates in logistic régression models," Vol. 71, $N°1$, pp. 1-10, 1984.
19. E. Deakin, "A discriminat analysis of predictors of business failure," *Journal of Accounting Research*, Spring, page 167-179, 1972.
20. F. William, M. Carthy, and N.Guo, "The existestence of maximun likelihood estimateds for the logistic regression model," *Maryland Medical Research Institute*, 2009.
21. M. Schreiner, "Scoring the next break through in microcrédit," CGAP, pp. 102, 2002.
22. A. Rencher, "Interpretation of Canonical Discrimant Functions, Canonical variates, and Principal Components," Vol. 4, $N°3$, pp. 217-225, 1992.
23. O. Kindi, and A. Achonu, "A credit Scoring model for the savings and crédit mutual of the patout zone (MECZOP)/ Sénégal," *The Journal of Sustainable Development*, vol. 7, pp. 17-33, 2012.
24. C. Gourieroux, "Courbe de performance, de sélection et de discrimination," *Anales d'économie et de statistique*, $N°28$, page 108-115, 1992.
25. C. Gourieroux, "Économétrie de la finance: L'exemple du risque de crédit," *Actualité de l'économie*, vol. 79, $N°4$, pp. 399-418, 2003.

Appendix

Proof of the Theorem 2.2

We know that $g_k = argmax\ \mathbb{P}_k f_k(x)$. Maximise $\mathbb{P}_k f_k(x)$ is equivalent to maximise $\ln(\mathbb{P}_k f_k(x))$;

$$\begin{aligned}\ln(\mathbb{P}_k f_k(x)) &= \ln(\mathbb{P}_k) + \ln[f_k(x)] \\ &= \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum{}^{-1}(x-\mu_k) - \ln\left[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}\right]\end{aligned}$$

Since $\ln\left[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}\right]$ does not depend on k, then maximizing $\ln(\mathbb{P}_k f_k(x))$, is equivalent to maximise $d_k(x) = \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum^{-1}(x-\mu_k)$. i.e maximizing $\mathbb{P}_k f_k(x)$ is to maximize $d_k(x)$ ■

Proof of the Proposition 2.4

It is important to show that:

$$\mathbb{P}(Y = k \mid X = x) = \frac{exp[d_k(x)]}{\sum_{k=0}^{1} exp[d_k(x)]} \tag{38}$$

with $d_k(x) = \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum{}^{-1}(x-\mu_k)$.
From the theorem of Bayes, $\mathbb{P}(Y = k \mid X = x) = \dfrac{\mathbb{P}_k f_k(x)}{\sum_{k=0}^{1} \mathbb{P}_k f_k(x)}$ or $\ln(\mathbb{P}_k f_k(x)) = d_k(x) - \ln[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}]$.

Thus

$$\begin{aligned}\mathbb{P}_k f_k(x) &= exp(d_k(x))exp(-\ln[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}]) \\ &= exp(d_k(x))\frac{1}{(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}}\end{aligned}$$

By replacing $\mathbb{P}_k f_k(x))$ by its value in the equation **5** the term $(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}$ is found on the numerator to the denominator and by simplifying we find the equation **38** ■

Proof of the Proposition 3.1

It is important to show that $\mathbb{P}(Y = 1 \mid X = x) = \dfrac{exp(x^\top \beta)}{1 + exp(x^\top \beta)}$.
From the relation 22, we have: $\ln\left(\dfrac{\mathbb{P}(Y = 1 \mid X = x)}{1 - \mathbb{P}(Y = 1 \mid X = x)}\right) = x^\top \beta$. By composing by the exponential function, we obtain $\dfrac{\mathbb{P}(Y = 1 \mid X = x)}{1 - \mathbb{P}(Y = 1 \mid X = x)} = exp(x^\top \beta)$ and by developing, then $\mathbb{P}(Y = 1 \mid X = x) = \dfrac{exp(x^\top \beta)}{1 + exp(x^\top \beta)}$. ■

