

SCORING: FAILURE RISK MANAGEMENT TOOL FOR SMES IN ALGERIA

Lamine TARHLISSIA

l_tarhlissia@esc-alger.dz

ÉCOLE SUPÉRIEURE DE COMMERCE, ALGIERS, ALGERIA

ABSTRACT:

Algerian public banks need to implement credit risk management techniques tailored to the specific characteristics of SMEs to prevent the deterioration of the banks' solvency due to the degradation of the quality of their SME portfolios. In this regard, our main objective is to highlight the interest that credit risk management will have within the People's Credit of Algeria by developing a Credit Scoring model based on the logistic regression technique, using a sample of 226 SMEs. The study results demonstrate the importance of the logistic regression model in classifying companies and its predictive ability for default, with a good classification rate of 91.2%.

KEYWORDS:

Banks, SMEs, scoring, risk, credit, Algeria

JEL Classifications: G21, C13

1. Introduction

Risk in general is considered an element inherent in all banking activities, and any financial institution faces various types of risks in its operations. Due to the strong relationship between return and risk, banks are key economic actors concerned with risk-taking, as their primary objective is to achieve maximum returns (Zamore, 2018). Given that the primary function of banks is credit provision and banking intermediation, the evident outcome of this activity is credit risk. In practice, credit risk cannot be completely eliminated, but it can be mitigated through effective management and analysis by financial institutions.

The bank's objective is not to eliminate this risk but to find the best “risk-return” combination (Egger, 2020). In this context, banks are urged to enhance their banking risk management measures, particularly credit

risk, and strengthen their analytical tools to accurately detect the various factors influencing and exacerbating this risk (Zhou, 2022).

Beyond the conventional management implemented by banks, which relies mainly on financial diagnosis and the ratio approach, and the requirement of banking guarantees, banks continuously attempt to implement other management methods based on sophisticated mathematical models tailored to each bank's strategy. Among these tools, the credit scoring method is one of the most well-known approaches to assessing counterparty risk, allowing the bank to create a scoring system tailored to its specific requirements. This leads us to pose the following issue: “*How can the bank improve both the prevention of credit defaults and its service quality to its*

clientele simultaneously?” The objective of our article is to present a modern tool for banks capable of addressing credit risk and providing a decision criterion that facilitates the selection of good clients while rejecting the bad ones: the logistic credit scoring method. As Algeria is a developing country with an economy characterized by a significant presence of small and medium-sized enterprises (SMEs), these enterprises generally rely on public banks for financing. Therefore, our article will focus on the management of the default risk of 226 SMEs that have obtained credit from the People's Credit of Algeria from 2015 to 2020, using the “Credit Scoring” tool.

2. Credit risk and evaluation methods

2.1. Definition of credit risk

We will define the concept of risk before discussing the notion of credit risk. So, risk is defined as: *“Risk refers to the uncertainty that weighs on results and potential losses when environmental developments are adverse”* (Churchill & Coster, 2001). Thus, *“risk corresponds to an unforeseeable event that affects the company's activity; it complicates the task of companies and financial institutions in achieving their objectives and realizing expected results”* (Sampson, 1982). Therefore, we can say that credit risk is the main risk to which a bank is exposed, and it is also the primary cause of difficulties and bankruptcies of banks, *“It is a potential loss resulting from the inability of borrowers to meet their financial commitments”* (Jacolin, Paquier, Mérieux & Marchand, 1995). According to Sampson (1982) credit risk is *“the tension that inhabits bankers, inseparable from their profession”*. They watch over the savings of others and yet make profits by lending them to others, which inevitably involves risks. A banker who does not take risks is not a banker (Azzouz, 2009).

2.2. Typology of credit risk

Credit risk consists of three types of risks: default risk, spread deterioration risk, and recovery uncertainty risk.

●Default Risk:

Default risk refers to the borrower's inability to meet its obligations at maturity, representing the most dangerous and common risk for a bank. It is also known as the risk of non-repayment or counterparty risk (Kénitra, 2009). There are three elements in measuring default risk:

– **Probability of Default (PD):** It measures the likelihood that a debtor defaults, regardless of the reason. Determining this probability requires estimation through statistical models such as Credit Scoring.

– **Exposure at Default (EAD):** This is the amount of outstanding repayments on the loan. It represents the maximum loss that the lender can incur in the event of immediate default.

– **Loss Given Default (LGD):** It represents the portion of the exposure at default that will not be recovered after the occurrence of default. It is measured by considering all direct and indirect costs related to the recovery of the loan. From these three parameters that constitute the default risk, the bank can calculate the amount of loss incurred at the time of default:

– **Expected Loss (EL):** It is the product of the three characteristic risk parameters of a loan – probability of default, exposure at default, and loss given default. This expected loss is used to estimate the anticipated risk that should be covered according to Basel Committee forecasts.

– **Unexpected Losses (UL):** These losses must be covered, according to the Basel Committee, by regulatory capital known as “economic capital” to maintain the bank's solvency.

●Spread Deterioration Risk:

The credit spread signifies the risk premium associated with it, indicating a

direct relationship between risk and premium. Therefore, an increase in the risk level results in a change in the premium. The spread risk translates into the deterioration of credit quality (degradation of the debtor's rating), leading to an increase in the risk premium and consequently a reduction in the interest margin (Mathlouthi, 2010).

● **Recovery Uncertainty Risk:**

This risk is associated with the recovery rate of the debt, determining the amount that will be recovered in case of borrower default after all legal procedures. This rate is a source of uncertainty for the bank, depending on the duration of legal procedures, the real value of collateral, and the bank's position in the list of creditors (Dietsch, 2010).

3. Scoring as a predictive tool for SME default risk

3.1. Concept of credit scoring

In common language, *“the term score can mean ranking, result, grade, etc”*. In statistics, the idea of 'ranking' is especially emphasized' (Boujelbène & Khemakhem, 2013).

Another definition states that: *“Scoring corresponds to a method of financial analysis that attempts to synthesize a certain number of variables in the form of a single indicator capable of distinguishing healthy companies from failing ones”* (Dietsch, 2008). Credit scoring is a technique for automated customer selection, generally based on statistical analysis. It aims to determine, from a study sample, the best classification of customers based on their risk profile by assigning each credit applicant a score inversely proportional to their probability of default (Siddiqi, 2012). The score is a scientific approach to a decision-making process. This rating assigned to a financing request synthesizes information contained in several financial ratios, especially those most predictive of default and most

discriminating between the two classes of 'good' and 'bad' clients. This is achieved through a statistical approach to data aggregation and processing.

3.2. Techniques of credit scoring

Constructing a scoring model implies assigning a risk score calculated from a linear combination of default explanatory variables. Numerous techniques exist to calculate this linear combination and obtain the coefficients (a_i). The most commonly used methods in the banking sector are linear methods due to their simplicity and robustness (Viganò, 1993).

3.2.1. Parametric approaches

In this category of approaches, two methods are distinguished:

Linear Discriminant Analysis (LDA): Discriminant analysis is a multidimensional statistical method of scoring that aims to discriminate and differentiate between observations based on their individual characteristics. It seeks to determine the contribution of quantitative explanatory variables that explain a qualitative dependent variable (the membership or non-membership of borrowers to two or more groups). The empirical application of discriminant analysis began in the 1930s through the research conducted by (Fisher & Mahalanobis, 1936).

Logistic Regression: It is a statistical method of scoring that aims to predict a qualitative characteristic (binary dependent variable). In other words, this method allows studying the relationship between a qualitative dependent variable (the variable to be explained) and explanatory independent variables that can take both quantitative and/or qualitative forms. Moreover, this technique enables the establishment of a scoring function by treating dichotomous variables as quantitative variables (Luo & Wu, 2017).

3.2.2. Non-parametric approaches

Unlike the parametric approach, this type of method imposes no distributional conditions on the sample under study or on the explanatory variables. This feature makes them widely used, although they lack a scoring function. Among the methods in this approach, we will focus on presenting the technique of neural networks (Hand & Henley, 1997).

Neural Networks: Neural networks are sophisticated and flexible non-parametric techniques, derived from artificial intelligence. They are inspired by the functioning of the human biological brain and are considered powerful forecasting tools. Neural networks measure complex relationships between explanatory variables and a dependent variable to be explained. Typically, these relationships are nonlinear, unlike classical credit scoring methods such as discriminant analysis and logistic regression, which seek a linear relationship. This technique originated from the work of Warren McCulloch and Walter Pitt in the 1940s on the formal neuron, demonstrating the capability of calculating any arithmetic or logical function with this approach. However, it took advancements in computer science and powerful mathematical tools to renew interest in these techniques (Jensen, 1992).

4. Methodology and variable selection

To construct a scoring function, three main steps must be followed: the creation of the initial sample, the selection of discriminatory variables, and statistical analysis. For the development of a credit scoring model, four main steps need to be followed: the choice of the sample to be analyzed and the default criterion, the selection of explanatory variables, the choice of the statistical technique to apply, and the validation of the constructed model. The following is a presentation of the approach followed in our work.

4.1. Sample and default criterion selection

Our study, focusing on credit risk management for SMEs using the scoring method, comprises small and medium-sized enterprises operating in the commercial and industrial sectors affiliated with the Credit Populaire d'Algérie, having obtained credit during the period from 2015 to 2020. Our database consists of 282 companies randomly selected from credits granted to SMEs in the commercial and industrial sectors by the CPA. This population includes 220 performing (Solvent) and 62 non-performing (failing) enterprises. We will divide our base sample into two sub-samples: the construction sample for building the model and a validation sample for validating the model. As for the default criterion, we have chosen the reason for any payment delay of 90 days from the date of the payment incident, in accordance with Basel Committee requirements and Regulation 2014-03 on the classification and provisioning of receivables.

4.2. Variable selection

The choice of explanatory variables is an essential step that determines the effectiveness of the score function. The initially chosen explanatory variables must undergo statistical tests to retain the most significant ones that are most related to the default of companies. The variables selected in our study are divided into two groups: quantitative variables (financial ratios, liquidity ratios, profitability ratios, liquidity ratios, and financial structure ratios) and qualitative variables. In our case, a list of 26 ratios coded from R1 to R26 represents the quantitative variables, and the table below summarizes the ratios selected in this study:

Table no. 1
Presentation of Selected Ratios in the Study

R1	Added value / Turnover
R2	Working Capital / Turnover
R3	Working Capital Requirement on days / Turnover
R4	Days sales outstanding / Turnover
R5	Days payable outstanding / Turnover
R6	Inventory/short term debt
R7	Cash / Assets
R8	Cash / short term debt
R9	Financial expenses/ Earnings before interest, taxes and depreciation
R10	short term debt /Total debt
R11	Rotation inventory
R12	Income / Assets
R13	Net income / shareholder's equity
R14	Self-financing capacity / turnover
R15	Added value/ turnover
R16	Earnings before interest, taxes and depreciation / turnover
R17	Operating income / turnover
R18	Financial income / turnover
R19	Income / turnover
R20	Income / turnover
R21	Total debt / Assets
R22	Financial expenses / Fonds propres
R23	Shareholders equity / non-current liabilities
R24	Shareholders equity / Assests
R25	Financial expenses/ net income
R26	Long term debt/ self-financing capacity

(Source: Compiled by the author)

For qualitative variables, we deemed it useful to transform these variables into dichotomous variables. In the following, we present the variables that we have derived:

Legal Form: Our sample consists of 4 types of companies: LLC, OPLC, SNC, and PC. The table below shows the modalities of the legal form.

Table no. 2
Company Modalities

LLC (limited liability company)	1
OPLC (one-person limited liability company)	2
JSC (Joint-stock company)	3
PC (partnership company)	4

– **Unpaid bills from colleagues:** This variable allows us to determine if the company has recorded a payment incident on a loan taken from another local bank. The variable takes on the modalities presented below:

Modality 1	None 01
Modality 2	Existing 02

– **Risk office:** This variable allows us to know whether the company had already obtained credit from a colleague. It takes two modalities presented below:

Modality 1	None 01
Modality 2	Existing 02

– **Movements in entrusted turnover:** This variable allows us to understand the movements of the customer account in relation to its turnover. It takes the following modalities:

Partial (0) If the movements entrusted to the bank are below 80% of the turnover achieved
Almost full (1) If the movements entrusted to the bank exceed 80% of the turnover achieved.

After identifying the variables, we will now perform some statistical tests on them to eliminate variables that do not explain the default and keep the most significant ones:

➤ **Correlation Test between Quantitative (Explanatory) Variables:**

One of the conditions for creating a scoring function is the independence of variables used to explain default. To check the dependence between the variables, we used the Pearson correlation test, which tests the correlation relationship between metric variables.

➤ **Dependency Test between Quantitative Variables and Default:**

To determine metric variables that have a dependence relationship with business failure, we conducted the ANOVA test, which detects the existence or non-existence of a dependence relationship between a quantitative variable (financial ratios) and another qualitative variable (default).

➤ **Dependency Test between Qualitative Variables and Default:**

At this stage, we need to test the dependence of each variable with the default individually. We performed the Chi-square test, which measures the dependence between two qualitative variables. The application of these three tests in the SPSS software allowed us to retain the following 10 ratios:

Table no. 3
List of Retained Variables

Quantitative Variables	Qualitative Variables
R2: working capital/Turnover	Unpaid bills from colleagues
R3: working capital requirement/Turnover	Risk office
R8: cash/short term debt	Movements in entrusted turnover
R9: financial expenses/Earnings before interest, taxes and depreciation	
R12: Income before interest/Total Assets	
R22: Financial expenses/shareholders equity	
R26: long term debt/self-financing capacity	

(Source: Compiled by the author)

4.3. Choice of statistical technique used

In our work, we will opt for the logistic regression method to estimate the scoring function. Logistic regression is a statistical scoring method that aims to predict a qualitative characteristic (binary dependent variable). In other words, this method allows the study of the relationship between a qualitative dependent variable (the variable to be explained) and explanatory independent variables that can take quantitative and/or qualitative forms.

4.4. Validation of the constructed model

Model validation methods are based, on the one hand, on statistical inference methods and robustness testing procedures on control samples. On the other hand, they involve checking the signs of the coefficients of financial ratios for their consistency with accounting and financial logic.

In our case, we used the confusion matrix as a validation method. This matrix provides a set of indicators, including:

- Percentage of correct classification of defaulting borrowers.
- Percentage of correct classification of healthy borrowers.
- Overall percentage of correct classification.
- Overall error rate.

5. Discussion and interpretation of results

There is a direct relationship between the company's score and its probability of default. The higher the company's score, the higher the risk of default. Therefore, a company with a high score will be more exposed to the risk of default. Consequently, our assignment rule is as follows:

If the score function $Y < 0$, the company is healthy ($P < 5\%$).

If the score function $Y > 0$, the company is in default ($P > 5\%$).

To establish the scoring function, we chose the method that allows for maximum

likelihood estimation (descent). This approach helps eliminate non-significant variables. The relationship between the dependent variable (default) and the explanatory variables selected by the logistic regression method resulted in the following scoring function:

$$\hat{Y} = -1,584 + 2,002 IC + 2,647 MC - 2,015 \hat{R}8 + 0,006 \hat{R}9 - 1,988 \hat{R}12 + 3,006 \hat{R}22 + 10,636 \hat{R}26 \quad (1)$$

According to the constructed model, we observe that the variable central risk situation and the ratios R3 and R4 have been systematically eliminated from the model due to their low level of significance.

Based on the coefficients of the variables, it appears that the “Long-term debt/CAF” ratio is the factor that most influences a company's default on repayment. We will demonstrate the consistency of these signs by analysing each variable individually:

●**Unpaid bills from colleagues:** This variable indicates whether the company has previously taken a loan from a peer and has recorded a payment incident on this loan. The estimation results show that the coefficient sign for this variable is positive (+2.002); this means that as payment incidents increase, the level of the company's score will also rise. Therefore, we observe that the higher this score, the weaker the company's repayment capacity.

●**Entrusted movements (EM):** This variable allows us to understand the movements in the company's account with the bank. The estimation results show that the coefficient of this variable is positive (+2.647), which is not consistent with financial logic. According to the coding we have performed, if entrusted movements are greater than 80% of the turnover, it means that the company's account is active, indicating good financial health. If the movements are less than 80% of the turnover, there is a possibility that the company may be facing financial difficulties that could lead to a default on repayment. In fact, this variable should be

inversely related to default, which is not the case in our study.

●**Long-term/Self-financial capacity:** (R26) This ratio measures the company's ability to repay its debts. The results show that the coefficient sign of this ratio is positive (+10.636). This implies long-term debt exceed Self-financial capacity, indicating an insufficient Self-financial capacity compared to long term debt. We can infer that the company's score is strongly and positively influenced by this ratio, representing the level of indebtedness relative to Self-financial capacity. The higher the value of this ratio, the lower the company's ability to repay its loans.

●**Financial expenses/Shareholder's equity (R22):** This ratio measures the degree of funding of the company's activities by equity or debt. The results show that this variable has a positive impact on the company's score, as indicated by its positive coefficient (+3.006). An increase in the value of this ratio raises the score and the likelihood of the company defaulting on payments. This aligns with financial logic because as this ratio increases, the company becomes less financially independent (relying on debt), increasing the possibility of facing repayment difficulties if its debts become excessive.

●**Cash/Short-term debt (R8):** This ratio indicates the extent to which available funds can finance debt or expresses the company's ability to meet its short-term commitments. The results show that the coefficient sign of this ratio is negative (-2.015). Therefore, there is an inverse

relationship between the company's score and this variable. A decrease in this ratio will inevitably increase the chances of a company defaulting because when a company's cash position tends to deteriorate continuously, it will be unable to repay its loans obtained from the bank.

●**Income/Total Assets (R12):** This ratio assesses the economic profitability of the company. The results show that the coefficient sign of this variable is negative (-1.988), indicating an inverse relationship between this ratio and the company's score. Therefore, the higher this ratio, the lower the company's exposure to default.

●**Financial expenses/Earnings before interest, taxes and depreciation (R9):** his ratio expresses the company's ability to repay debt service. The results show that the coefficient of this ratio is positive (+0.006), indicating a positive relationship between this ratio and the company's score. Therefore, the higher this ratio, the higher the proportion of Earnings before interest, taxes and depreciation absorption for the payment of loan interest, and consequently, the higher the likelihood of a company defaulting. After estimating the parameters of the score function, we now move on to its validation, through the confusion matrix.

– The confusion matrix:

In the following, we will present the results of the reassignment of "good/bad" companies that make up the construction sample. The results of the confusion matrix are presented in the table below:

Table no. 4
Confusion Matrix

	Company considered Solvent	Company considered failing	Total	Correct Percentage
Solvent company	172	6	178	96,6%
Failing companies	14	34	48	70,8%
Good classification rate				91,2%

(Source: Compiled by the author)

This method allows us to calculate the correct classification rate and the error rate:

- Correct percentage for solvent companies: $172/178 = 96.6\%$;

- Correct percentage for failing companies: $34/48 = 70.8\%$;

- Overall correct classification rate: $(172+34) / 226 = 91.2\%$;

- Error rate for considering a failing company as healthy: $14/48 = 29.2\%$;

- Error rate for considering a healthy company as failing: $6/178 = 3.4\%$.

According to the results of the confusion matrix, which allowed us to determine the rate of correct classification of companies in the construction sample, we can say that our model classified 206 companies into their original class out of 226, with an error in classification of 20 companies. Based on this, we can conclude that our model provides very satisfactory results. Comparing the correct classification rate of healthy companies (96.6%) with the correct classification rate of failing companies (70.2%), we observe that our model better classified healthy companies than failing companies.

6. Conclusion

The importance of credit risk and its weight in determining the level of banks' solvency encourages them to develop more techniques to ensure better management of this risk, through a more precise definition and a more reliable estimation.

In this context, we focused on credit risk management by presenting the most adopted techniques by Algerian banks, such as credit scoring. This technique allows us to determine, through the estimation of risk scores incurred by the bank, their own default probabilities. The importance of this parameter lies in allowing us to estimate the successive losses resulting from the potential default of borrowers.

Practically, our work focused on building a credit scoring model. The construction methodology we adopted involved collecting a database consisting of 226 companies divided into two subgroups: defaulting and healthy. For each of these companies, we had their own financial statements as well as the study of the file previously conducted by the study officer.

From there, we calculated a set of ratios, and we were able to collect other non-financial information that we considered explanatory of the default of these companies. We then proceeded to estimate the score function by resorting to logistic regression by maximum likelihood, once all appropriate statistical tests were conducted.

The application of logistic regression allowed us to obtain good results in predicting financial distress in companies, characterized by a good classification rate of 91.2%. It allowed us to identify the following default indicators:

- An increase in defaults among peers increases the chances of non-repayment by the borrower.

- A significant repayment capacity leads to a lower risk of default.

- An increase in the financial leverage of the company negatively influences credit repayment.

- The increase in the debt service capacity leads to a lower risk of default.

- The company's ability to repay short-term debts, the higher it is, the lower the risk of default.

- Entrusted movements positively impact credit risk, which is not consistent with financial logic, as a busy bank account means that the company is financially healthy and therefore is certainly capable of repaying its debts.

REFERENCES

- Azzouz, E. (2009). La gestion du risque crédit par la méthode du scoring: cas de la Banque Populaire de Rabat-Kénitra. *Revue marocaine de recherche en management et marketing*, CNRST-IMIST, 291.
- Boujelbène, Y., & Khemakhem, S. (2013). *Prévision du risque de crédit: Une étude comparative entre l'Analyse Discriminante et l'Approche Neuronale*. Tunisie.
- Churchill, C., & Coster, D. (2001). *Manuel de gestion des risques en microfinance*.
- Dietsch, M., & Petey, J. (2008). *Mesure et gestion du risque de crédit dans les institutions financiers, édition 2*. RB Edition.
- Egger, D.J., Gutiérrez, R.G., Mestre, J.C., & Woerner, S. (2020). Credit risk analysis using quantum computers. *IEEE Transactions on Computers*, Vol. 70, Issue 12, 2136-2145. Available at: [10.1109/TC.2020.3038063](https://doi.org/10.1109/TC.2020.3038063).
- Hand, D.J., & Henley, W.E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 160, Issue 3, 523-541. Available at: <https://doi.org/10.1111/j.1467-985X.1997.00078.x>.
- Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law & Technology*. Available at: https://openyls.law.yale.edu/bitstream/handle/20.500.13051/7808/Hurley_Mikella.pdf.
- Jacolin, L., Paquier, O., Mérieux, A., & Marchand, C. (1995). La tarification des crédits. *Revue d'économie financière*, no. 35, 119-140.
- Jensen, H.L. (1992). Using Neural Networks for Credit Scoring. *Managerial Finance*, Vol. 18, Issue 6, 15-26. Available at: <https://doi.org/10.1108/eb013696>.
- Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, Vol. 65, 465-470. Available at: <https://doi.org/10.1016/j.engappai.2016.12.002>.
- Mathlouthi, H. (2010). *Cours de méthodes de scoring*. Tunisie: Université de Carthage.
- Sampson, A. (1982). *Les banquiers dans un monde dangereux*. R. Laffont.
- Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring (Vol. 3)*. John Wiley & Sons.
- Viganò, L. (1993). A credit scoring model for development banks: An African case study/un modele de "credit scoring" pour les banques de developpement: une etude de cas Africain. *Savings and development*, Vol. 17, Issue 4, 441-482.
- Zamore, S., Ohene Djan, K., Alon, I., & Hobdari, B. (2018). Credit risk research: Review and agenda. *Emerging Markets Finance and Trade*, Vol. 54, Issue 4, 811-835.
- Zhou, X.Y., Caldecott, B., Hoepner, A.G., & Wang, Y. (2022). Bank green lending and credit risk: an empirical analysis of China's Green Credit Policy. *Business Strategy and the Environment*, Vol. 31, Issue 4, 1623-1640. Available at: <https://doi.org/10.1002/bse.2973>.