Data and Metadata. 2026; 5:1289
doi: 10.56294/dm20261289

ORIGINAL

Financial predictors of sme failure: variable selection with lasso versus random forest

Predictores financieros del fracaso de las pyme: selección de variables con lasso frente a random forest

Amina Guennoun[1], Siham Ammari[2], Saad Saadouni[1], Souad Habbani[1]

[1]Interdisciplinary Research Laboratory in Economics, Finance, and Organizational Management (LIREFIMO), Faculty of Law, Economics and Social Sciences, University Sidi Mohammed Ben Abdellah. Fez, Morocco.
[2]Laboratory of Innovation in Management and Engineering for Enterprise (LIMIE), ISGA Business School. Fez, Morocco.

Cite as: Guennoun A, Ammari S, Saadouni S, Habbani S. Financial predictors of sme failure: variable selection with lasso versus random forest. Data and Metadata. 2026; 5:1289. https://doi.org/10.56294/dm20261289

Submitted: 08-09-2025 **Revised:** 14-10-2025 **Accepted:** 09-12-2025 **Published:** 01-01-2026

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Amina Guennoun

ABSTRACT

Introduction: the study aimed to identify the financial variables that best predicted the failure of small and medium-sized enterprises (SMEs). It addressed the need for reliable financial indicators capable of signaling early distress and supporting risk-management practices.
Methods: a quantitative methodology was adopted within a hypothetico-deductive framework. Two complementary variable-selection techniques were applied. First, the LASSO regression method introduced a regularization constraint to eliminate variables with weak explanatory power. Second, the Random Forest algorithm assessed the relative importance of financial variables in overall model performance. The two approaches were compared to determine their effectiveness in identifying the most relevant predictors of SME failure.
Results: the LASSO model produced a negative coefficient of determination (R^2 = -1,2179), demonstrating performance inferior to a simple mean-based prediction and indicating that LASSO was not suitable in this context. In contrast, the Random Forest model achieved a very high R^2 value (0,9571), reflecting strong predictive accuracy and robustness. Based on the Random Forest results, six key financial predictors of SME failure were identified: financial structure, return on assets, return on sales, return on equity, liquidity, and solvency.
Conclusions: the study demonstrated that Random Forest outperformed LASSO in selecting meaningful financial predictors of SME failure. The six identified variables offered a reliable analytical framework for understanding and anticipating financial distress. These findings provided valuable insights for academic research and practical applications in risk assessment and early warning systems for SMEs.

Keywords: Prediction SME Failure; Variable Selection; Financial Ratios; LASSO Regression; Random Forest.

RESUMEN

Introducción: el estudio tenía como objetivo identificar las variables financieras que mejor predecían el fracaso de las pequeñas y medianas empresas (pymes). Abordaba la necesidad de contar con indicadores financieros fiables capaces de señalar dificultades tempranas y respaldar las prácticas de gestión de riesgos.
Métodos: se adoptó una metodología cuantitativa dentro de un marco hipotético-deductivo. Se aplicaron dos técnicas complementarias de selección de variables. En primer lugar, el método de regresión LASSO introdujo una restricción de regularización para eliminar las variables con escaso poder explicativo. En segundo lugar, el algoritmo Random Forest evaluó la importancia relativa de las variables financieras en el rendimiento global

del modelo. Se compararon ambos enfoques para determinar su eficacia a la hora de identificar los predictores más relevantes del fracaso de las pymes.
Resultados: el modelo LASSO produjo un coeficiente de determinación negativo (R^2 = -1,2179), lo que demuestra un rendimiento inferior al de una simple predicción basada en la media e indica que LASSO no era adecuado en este contexto. Por el contrario, el modelo Random Forest alcanzó un valor R^2 muy alto (0,9571), lo que refleja una gran precisión predictiva y solidez. A partir de los resultados de Random Forest, se identificaron seis predictores financieros clave del fracaso de las pymes: estructura financiera, rendimiento de los activos, rendimiento de las ventas, rendimiento del capital, liquidez y solvencia.
Conclusiones: el estudio demostró que Random Forest superó a LASSO en la selección de predictores financieros significativos del fracaso de las pymes. Las seis variables identificadas ofrecieron un marco analítico fiable para comprender y anticipar las dificultades financieras. Estos hallazgos proporcionaron información valiosa para la investigación académica y las aplicaciones prácticas en la evaluación de riesgos y los sistemas de alerta temprana para las pymes.

Palabras clave: Predicción del Fracaso de las PYMES; Selección de Variables; Ratios Financieros; Regresión LASSO; Bosque Aleatorio.

INTRODUCTION

Business failure is a complex and multidimensional phenomenon that has attracted scholarly attention since the early twentieth century, with Fitzpatrick[1] among its pioneers. Research on this topic has developed along multiple perspectives, including economic, financial, strategic, organizational, and managerial dimensions. At the macroeconomic level, studies have highlighted the influence of economic cycles, monetary policy, and systemic conditions on bankruptcy risk.[2,3,4,5,6] At the microeconomic level, failure primarily affects firms with insufficient sales to cover costs, underinvestment, or inefficiencies arising from contractual and managerial factors.[7,8,9,10] Strategic choices, innovation, competition, and organizational factors further shape the vulnerability of firms.[11,12,13,14]

Financial distress constitutes a central dimension of failure. Firms in difficulty typically experience cash-flow tensions, declining profitability, and deterioration of key financial ratios.[15,16,17] From a legal standpoint, failure occurs when a firm cannot meet its obligations, potentially triggering judicial reorganization or liquidation, as codified in Moroccan law under Law No. 73-17 (2019), which introduced safeguard procedures to promote early restructuring and preserve employment.

The literature emphasizes that failure is rarely sudden; it is a gradual process, with early warning signals often detectable in financial data.[18] Financial ratios reflecting profitability, liquidity, and solvency have proven valuable in predicting bankruptcy, yet selecting the most relevant indicators remains a major methodological challenge due to high correlations and redundancy among variables.[19,20,21] In emerging markets, and particularly for SMEs, the relevance of these predictors may differ substantially due to structural financing constraints, market imperfections, and local institutional conditions.

Against this backdrop, a clear research gap emerges: although numerous predictors have been proposed, little consensus exists on which variables are most relevant for SMEs in emerging markets, nor on which modern statistical or machine-learning methods are best suited for variable selection in this setting. Against this backdrop, a central question arises: which financial factors are the most predictive of failure? This paper addresses this gap by empirically testing and comparing two contemporary variable-selection techniques LASSO regression and the Random Forest algorithm in the context of Moroccan SMEs, and identifying a validated set of key financial indicators that best predict failure in this environment. This contribution advances both methodological and empirical knowledge in bankruptcy-risk modeling.

METHOD

Hypotheses development

After defining business failure according to its various approaches, it becomes necessary to identify the key financial factors based on the literature review, in order to formulate the hypotheses to be tested.

In this perspective, liquidity reflects a firm's ability to meet its short-term obligations through its quickly mobilizable assets. In the case of SMEs, particularly in Morocco, this dimension is especially important due to the structural fragility of their financing and limited access to external capital. A deterioration in liquidity ratios disrupts the balance sheet equilibrium, reduces the coverage capacity of the operating cycle, and often forces the firm to resort to debt, thereby increasing financial charges. The literature generally considers these tensions as one of the earliest warning signals of potential failure.

Several empirical studies confirm the central role of liquidity in failure mechanisms. Some authors[22,23] emphasize that liquidity ratios assess an organization's ability to repay its debts at maturity. Persistent deficits can lead to default, with the firm becoming unable to meet its obligations.[24] Back et al.[25], using a sample of 74 Finnish firms, showed that liquidity is a major determinant of bankruptcy: including liquidity indicators in their models significantly reduced the classification error rate during the three years preceding failure. Bunn et al.[26] note that the current ratio reduces the probability of failure and represents one of the essential parameters for measuring this factor. Finally, the systematic review conducted by Kliestik et al.[27] on 103 predictive models developed in the Visegrad countries between 1993 and 2018 confirms that the current ratio remains the most frequently used measure in bankruptcy prediction studies. Based on the theoretical arguments and empirical evidence presented above, we propose the following hypothesis:

Hypothesis 1: "Liquidity is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

Furthermore, the financial structure reflects the balance between stable resources (equity and long-term debt) and the assets they finance. When imbalance occurs, it weakens the firm's stability and increases its exposure to failure risk. Empirical research shows that financing structure affects firms' resilience to economic shocks, particularly for SMEs, whose financial flexibility is limited.

Among the causes of failure frequently noted in the literature is an inadequate financial structure, revealing a deficit of stable resources to cover permanent assets.[18,28] The inability to mobilize external funds, often due to lenders' distrust, constitutes an aggravating factor, as highlighted by Argenti[29], Crutzen[18], Marco[30], and Ooghe et al.[31] This vulnerability is closely linked to the firm's financial autonomy, i.e., the relative weight of debt in its financial structure. Based on these strong theoretical foundations, we propose the following hypothesis:

Hypothesis 2: "An unbalanced financial structure is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

In addition, solvency refers to a firm's ability to meet its long-term financial obligations. It is generally measured by the ratio of total debt to equity: the higher this ratio, the greater the dependence on external financing, increasing insolvency risk, particularly during activity contractions or cash-flow tensions. In the context of Moroccan SMEs, this indicator is closely monitored by creditors as an early signal of structural weakness.

Insolvency can be defined as a firm's inability to meet its due liabilities with available assets.[32] It manifests as a persistent difficulty in making regular payments when short-term liabilities significantly exceed realizable assets, leading to repeated payment incidents toward creditors. This vulnerability is assessed using ratios comparing assets and liabilities with the same maturity.[33] A firm is truly in distress only when part or all of its solvency is compromised, i.e., when it can no longer meet its financial obligations.[8]

Numerous studies have demonstrated the relevance of solvency indicators to differentiate healthy firms from those in distress. The pioneering work of Altman,[6] as well as more recent studies[21,34,35,36,37,38,39], confirm that these variables are reliable tools for anticipating failure risk. Based on the theoretical justification and empirical evidence presented above, we formulate the following hypothesis:

Hypothesis 3: "Solvency is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

Moreover, empirical analyses show that firms with low or negative economic profitability are more exposed to structural difficulties that can lead to failure. Economic profitability measures how efficiently a firm uses its assets to generate profits. A prolonged decline in this ratio limits self-financing capacity and undermines investment viability, thus fostering financial tensions.

Insufficient economic asset profitability is frequently identified as a major explanatory factor of failure. In discriminant analyses, this factor is particularly significant, as healthy firms generally exhibit satisfactory profitability, unlike failing firms.[28,40] Therefore, the economic profitability ratio emerges as a relevant indicator for assessing economic performance and anticipating failure risk. Based on the theoretical and empirical evidence presented, we propose the following hypothesis:

Hypothesis 4: "Economic profitability is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

Financial profitability, complementary to economic profitability, measures the return obtained by shareholders on their invested capital. A decline in this ratio can reduce the firm's attractiveness to investors, limit its financing sources, and increase reliance on debt, thereby raising the risk associated with cash-flow tensions. It is thus a central indicator of a firm's financial strength and sustainability.

Lack of financial profitability reflects an inability to generate returns for capital providers. When they can no longer achieve satisfactory returns on their investment, they may choose to disengage or even liquidate the firm. This factor is particularly discriminant in detecting at-risk firms, as shown.[40] The financial profitability ratio is the most appropriate indicator for measuring this level of performance.[41] Considering the theoretical foundations and literature evidence, we propose the following hypothesis:

https://doi.org/10.56294/dm20261289 ISSN: 2953-4917

Hypothesis 5: "Financial profitability is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

Beyond economic and financial profitability, commercial profitability—generally measured by the margin generated on sales—reflects a firm's ability to generate surplus from operational activity. A sustained contraction of this margin constitutes a warning signal, indicating loss of competitiveness and difficulty in covering fixed costs. This situation reduces strategic flexibility and increases the risk of business cessation, particularly in competitive environments.

Several commercial factors can explain this failure, such as poor understanding of customer needs and expectations, inadequate product offerings, weak market positioning, or shortcomings in the sales force. These dysfunctions, by affecting commercial performance, directly contribute to low commercial profitability and, consequently, to increased exposure to failure risk.[31,42,43,44,45] Based on the observations above, we formulate the following hypothesis:

Hypothesis 6: "Commercial profitability is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

It is also important to consider financial balance. This reflects a firm's ability to finance fixed assets using stable resources, such as equity and long-term debt. Imbalance, characterized by excessive reliance on short-term resources to finance long-term assets, undermines cash-flow stability and weakens the balance sheet structure. This balance is generally measured using the ratio of permanent capital to fixed assets, a central indicator for assessing financial strength.

The mismatch between resources and the assets they finance is also identified as a risk factor for failure.[18,28] In this perspective, the financial balance ratio is a relevant tool to capture and quantify this situation. Based on these foundations, we propose the following hypothesis:

Hypothesis 7: "Financial balance is a key determinant of business failure risk and constitutes a reliable predictive indicator in forecasting models."

Finally, accounts receivable appears as a central component of working capital, reflecting a firm's ability to manage payment terms and collection efficiency. Excessive accumulation of receivables ties up financial resources, increases working capital needs, and reduces available cash, particularly exposing SMEs to tensions that can threaten their sustainability. Rigorous management of this item is therefore essential to preserve liquidity and solvency.

Some firms grant excessively large credits to clients to boost sales or encourage loyalty. This practice can harm cash flow and financial balance, generating additional financing needs and increasing default risk, potentially creating a domino effect.[28] Excessive receivables may also reflect weak negotiating power with clients, placing the firm in a vulnerable position relative to its partners. In our model, this hypothesis will be evaluated using the accounts receivable ratio to assess its role as a predictive indicator of failure risk. Based on the theoretical justification presented above, we propose the following hypothesis:

Hypothesis 8: "High levels of accounts receivable are a key determinant of business failure risk and constitute a reliable predictive indicator in forecasting models."

METHOD

In the context of our study, the focus is placed on causal explanation through the use of quantitative methods, which allow for the identification of significant relationships between explanatory variables particularly accounting and financial ratios and the occurrence of business failure. From this perspective, our epistemological positioning aligns with a moderated positivist stance, suited to the nature of our research object and the methodological requirements it entails.

Adopting this stance leads us to favor a hypothetico-deductive approach, based on the formulation of explicit hypotheses derived from an established theoretical framework, and their verification through the analysis of empirical data. This approach aims to highlight explanatory relationships between specific financial variables and the risk of SME failure. By ensuring rigor and reproducibility, this methodological positioning guarantees the internal coherence of our approach and enhances the credibility and robustness of the results, while accounting for uncertainties and limitations inherent in any modeling in management sciences.

Sample and variables of the study

Defining the sample constitutes a key step in any empirical research, as it directly affects the robustness and validity of the results. Prior to selecting the observational units, it is essential to rigorously define the phenomenon under study, in this case, business failure.

In this study, we adopted the legal approach to failure, considered the most robust from an institutional perspective. According to the Moroccan Commercial Code, a firm is deemed to be in failure when it is in a state of cessation of payments, that is, when it is unable to meet its due liabilities with available assets, a situation that triggers the opening of collective proceedings. This definition, which is objective and legally framed, allows for the precise identification of firms in a confirmed state of failure.

Based on this criterion, the cases studied were selected from the archives of the Fès commercial court and from information provided by trustees in charge of collective proceedings, ensuring the reliability of the dataset. Healthy firms, on the other hand, were selected with the assistance of local certified accountants, who provided an updated database.

The resulting sample consists of small and medium-sized enterprises located in the Fès-Meknès region, an economically dynamic area that has been relatively underexplored in studies on business failure risk. This choice addresses both data accessibility considerations and the ambition to offer a conceptualized understanding of the vulnerabilities specific to Moroccan SMEs.

Sample specifications

Initially, our investigation focused on Moroccan small and medium-sized enterprises (SMEs) established in the Fès-Meknès region. An initial census identified 110 entities, including both firms considered healthy and others facing difficulties.

To ensure methodological consistency and comparability of financial indicators, we restricted the selection to companies with a minimum age of five years. This criterion aimed to exclude newly established entities, whose financial statements do not yet reflect stabilized operations and whose potential problems are more likely related to start-up constraints than to confirmed economic failure.

For the sake of homogeneity, only limited liability companies (SARLs) were retained, as this legal form represents the majority of regional SMEs and provides more consistent access to financial statements. Entities with incomplete accounting information, larger firms, or those located outside the study area were excluded.

After this filtering process, the final sample comprises 60 SMEs: 30 healthy and 30 in a state of failure, totaling 180 observations covering the fiscal years 2018 to 2020. This rigorous selection ensured the homogeneity of the dataset as well as the reliability of the data used for empirical analysis.

The following table explains the reasons for excluding certain companies and the selection of those retained for the empirical study:

Table 1. Composition of the sample selected for the study

Total number of companies	110 companies
Companies not retained	50 companies not retained
Unavailable information	• 25 companies with incomplete balance sheets or income statements • 13 companies with inconsistent legal forms • 6 large companies • 6 newly established companies
Companies retained	60 companies: (30 failing; 30 healthy)
Characteristics of the retained sample	• Age: Over 5 years • Size: SME • Legal form: SARL • Availability of accounting information and documents • Region: Fès-Meknès.

Selection of Study Variables

Our literature review revealed a wide array of accounting and financial ratios employed in empirical studies on failure prediction. The most commonly used explanatory variables relate to key dimensions of financial analysis, including liquidity, profitability, financial structure, solvency, and management efficiency. In line with this tradition, the present study defined the target variable in binary form, assigning a value of 1 to failing firms and 0 to healthy firms. To ensure the robustness and generalizability of the predictive models, the dataset was divided into training and test sets, enabling an objective assessment of model performance on unseen data. Moreover, before applying the LASSO method, all financial variables were normalized to eliminate scale-related distortions and to allow the regularization process to operate under optimal conditions for variable selection and coefficient estimation.

Taking into consideration both data availability and the specific characteristics of Moroccan SMEs, eight accounting and financial ratios were initially selected to inform the statistical analysis. These indicators capture the main dimensions of the firms' financial and operational performance and are likely to offer meaningful explanatory insights into the mechanisms underlying business failure. The selected ratios are presented as follows:

https://doi.org/10.56294/dm20261289 ISSN: 2953-4917

Table 2. List of ratios used for the study

Ratios	Related Hypothesis	Definition	Calculation Formula	Coding
Liquidity Ratio	H1	Current ratio	Current Assets / Current Liabilities	LI
Structure Ratio	H2	Fixed assets coverage ratio by permanent funds	Equity / Total Assets	ST
Debt Ratio	H3	Solvency ratio	Total Debt / Equity	SLV
Profitability Ratio	H4	Economic profitability	Operating Income / Total Assets	ROA
	H5	Financial profitability	Net Income / Equity	ROE
	H6	Commercial profitability	Net Income / Sales Revenue	ROS
Financial Balance Ratio	H7	Fixed assets coverage ratio by permanent capital	Permanent Capital / Fixed Assets	EF
Management Ratio	H8	Accounts receivable coverage ratio	Accounts Receivable / Sales Revenue	CC

RESULTS

Descriptive statistical analysis of the sample

This section presents a descriptive statistical analysis of the sample studied over the period 2018-2020. All statistical processing and analysis were performed using Python software, enabling rigorous data manipulation and the production of reliable results (table 3).

The descriptive analysis of the financial variables reveals a marked dispersion of data, reflected in high standard deviations and wide ranges between minimum and maximum values. The skewness analysis reveals that the LI, ST, EF, and CC indicators show positive skewness, reflecting a concentration of observations around low values, accompanied by a few exceptionally high values. Conversely, ROE, ROA, ROS, and SLV show negative skewness, indicating a distribution centered on relatively high values, but punctuated by particularly low observations.

All variables also have kurtosis coefficients well above the reference value (kurtosis > 3), indicating leptokurtic distributions, characterized by a high concentration around the mean and the presence of extreme elements. The results of the Jarque-Bera test confirm, with zero p-values, the rejection of the normality hypothesis for all the series analyzed.

These findings highlight the non-Gaussian structure of the financial data studied and justify the use of robust econometric methods capable of handling heterogeneity and non-normal distributions in the context of default risk forecasting.

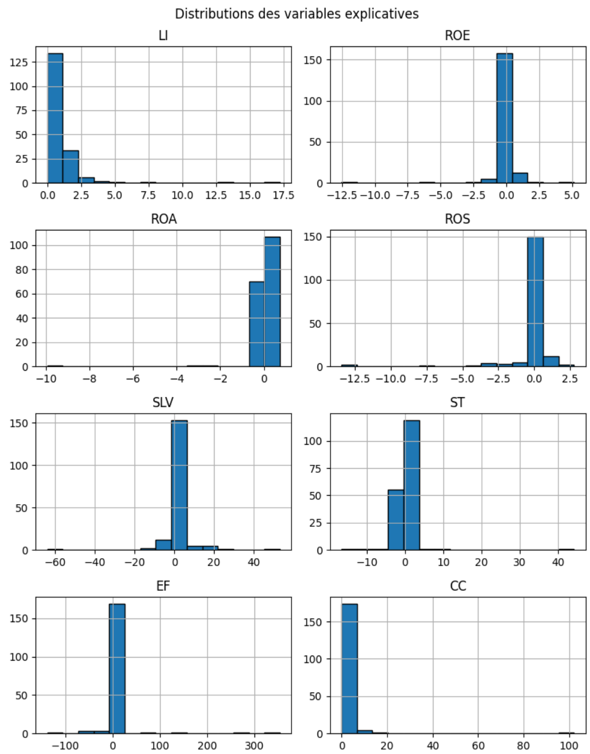

Figure 1. Distributions of explanatory variables

https://doi.org/10.56294/dm20261289 ISSN: 2953-4917

Table 3. Descriptive statistics for explanatory variables over the period 2018-2020

	LI	**ROE**	**ROA**	**ROS**	**SLV**	**ST**	**CF**	**CC**
Count	180,0	180,0	180,0	180,0	180,0	180,0	180,0	180,0
Mean	1,11031	0,008782	-0,032706	-0,149857	1,284998	0,026929	4,719066	1,422383
Std	1,760658	1,180169	0,823021	1,682407	7,606521	3,884097	38,181974	7,816375
Min	0,0	-12,530317	-9,93693	-13,435065	-64,014592	-16,608314	-138,302148	0,0
25 %	0,326207	-0,000937	-0,030597	-0,001122	0,230609	-0,572275	0,300783	0,004494
50 %	0,857386	0,128054	0,063677	0,079613	0,609519	0,004475	1,005269	0,237734
75 %	1,167406	0,255085	0,161055	0,252825	0,988163	0,542289	1,997709	0,758983
Max	17,22767	5,156858	0,733533	2,776787	53,15264	44,153701	354,385416	102,161402
Skewness	6,485849	-7,046357	-10,182099	-5,847899	-1,242666	7,762023	6,331862	12,15848
Kurtoisis	51,39614	76,454937	119,278534	40,654676	42,50319	95,66575	54,598984	156,444254
JB statistic	18828,384984	41956,742787	104515,484388	11659,99723	11750,091883	66209,529841	21171,187919	181023,402182
P-value	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Jaque-bera interpretation	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*	*Rejeter H0*

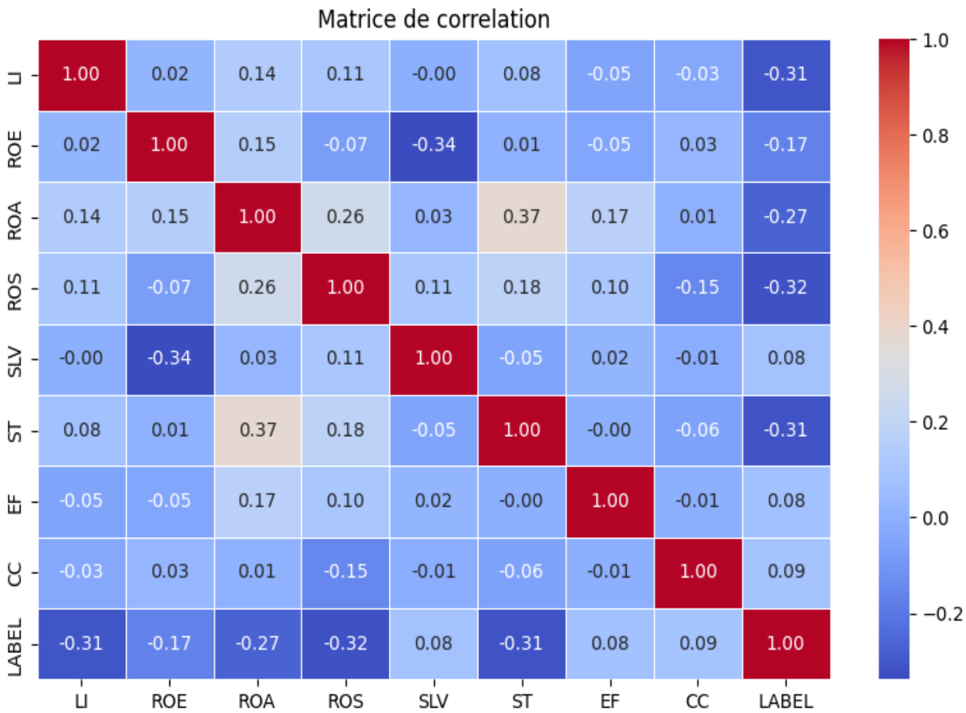

Figure 2. Correlation matrix between the explanatory variables and the target variable (LABEL) indicating the failure situation

The correlation matrix analysis highlights the statistical relationships between the main financial ratios and the target variable measuring the state of default of SMEs. The results indicate that several indicators show a moderate negative correlation with default, notably commercial profitability (ROS: -0,32), liquidity (LI: -0,31), financial structure (ST: -0,31), economic profitability (ROA: -0,27) and, to a lesser extent, financial profitability (ROE: -0,17). These results suggest that good performance in these areas helps reduce the risk of failure.

Conversely, variables such as solvency (SLV), financial autonomy (EF), and the customer receivables coverage ratio (CC) show a slightly positive correlation with default status, but their intensity remains low (approximately +0,08 to +0,09), thus limiting their explanatory power.

An examination of the cross-correlations between the ratios also highlights the absence of strong linear dependencies: the highest, observed between ROA and ST (0,37), remains well below the critical thresholds for multicollinearity. This relative independence of the variables reinforces the robustness of the empirical framework and strengthens the validity of the resulting predictive models.

Selection of predictors

In this article, the selection of explanatory variables is an essential step in improving the performance of business failure prediction. In order to reduce the dimensionality of the dataset while retaining the most relevant variables, two complementary approaches were implemented: the random forest method and lasso regression. The combined use of these techniques makes it possible to obtain a robust subset of explanatory variables, while limiting the risks of overfitting and strengthening the interoperability of the model.

Random forest method

The random forest method is a powerful tool for selecting explanatory variables. By measuring the relative importance of each indicator based on its effect on reducing impurity or prediction error within the trees, it identifies the most decisive attributes of the model. This approach is particularly well suited to complex or large datasets, thanks to its robustness in the face of outliers and possible correlations between variables.

Figure 3 shows the assessment of the relative importance of explanatory variables in predicting business failure, as estimated by the random forest algorithm. The results highlight the preeminence of financial structure (ST), which alone accounts for 34,7 % of the model's explanatory power, confirming its decisive role in the occurrence of failures.

Economic performance indicators, in particular commercial profitability (ROS) and economic profitability (ROA), also play a substantial role, with respective contributions of 23,1 % and 21,0 %. These results underscore that profitability and financial strength are key variables for anticipating corporate difficulties.

Conversely, certain variables such as financial equilibrium (EF) and accounts receivable (CC) are of marginal

https://doi.org/10.56294/dm20261289 ISSN: 2953-4917

importance, at less than 2 %. Their low weight suggests a limited influence on the probability of default, or possible redundancy with more structural indicators in the model.

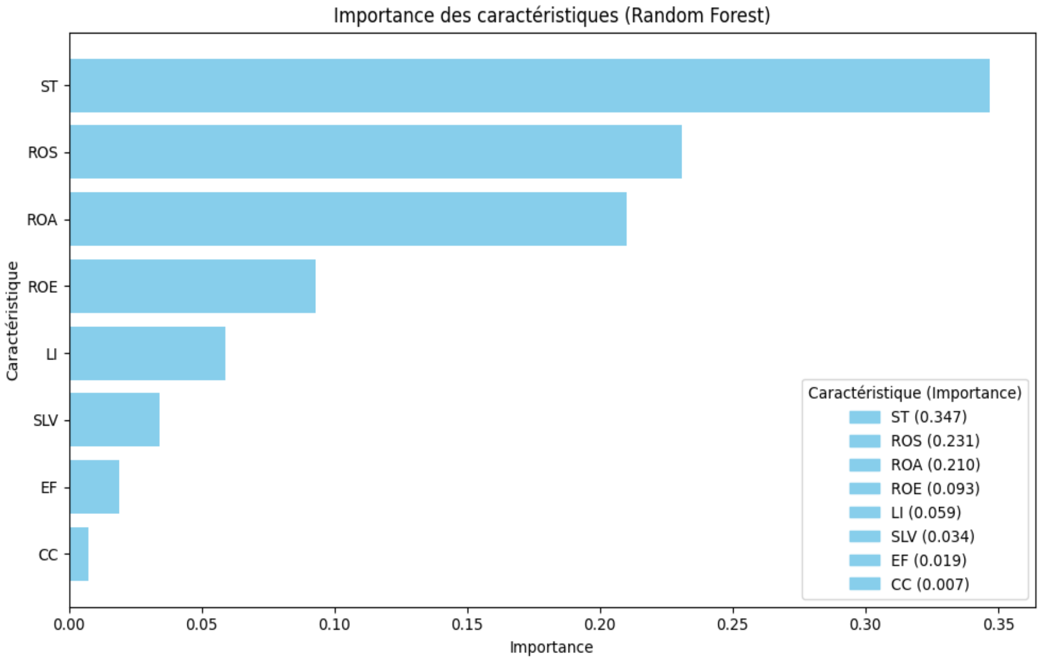

Figure 3. Importance of predictors according to the random forest algorithm

Lasso method

The acronym Lasso stands for "Least Absolute Shrinkage and Selection Operator." This method is frequently used to manage high-dimensional data, as it facilitates automatic feature selection. The Lasso principle is based on adding a penalty term to the sum of squared residuals (RSS), weighted by a regularization parameter denoted λ (lambda). This parameter controls the intensity of the regularization applied: high values of λ increase the penalty, causing several coefficients to be reduced to zero and, as a result, the automatic elimination of certain variables. Conversely, lower values of λ reduce the effect of the penalty, allowing a greater number of variables to be retained in the model.

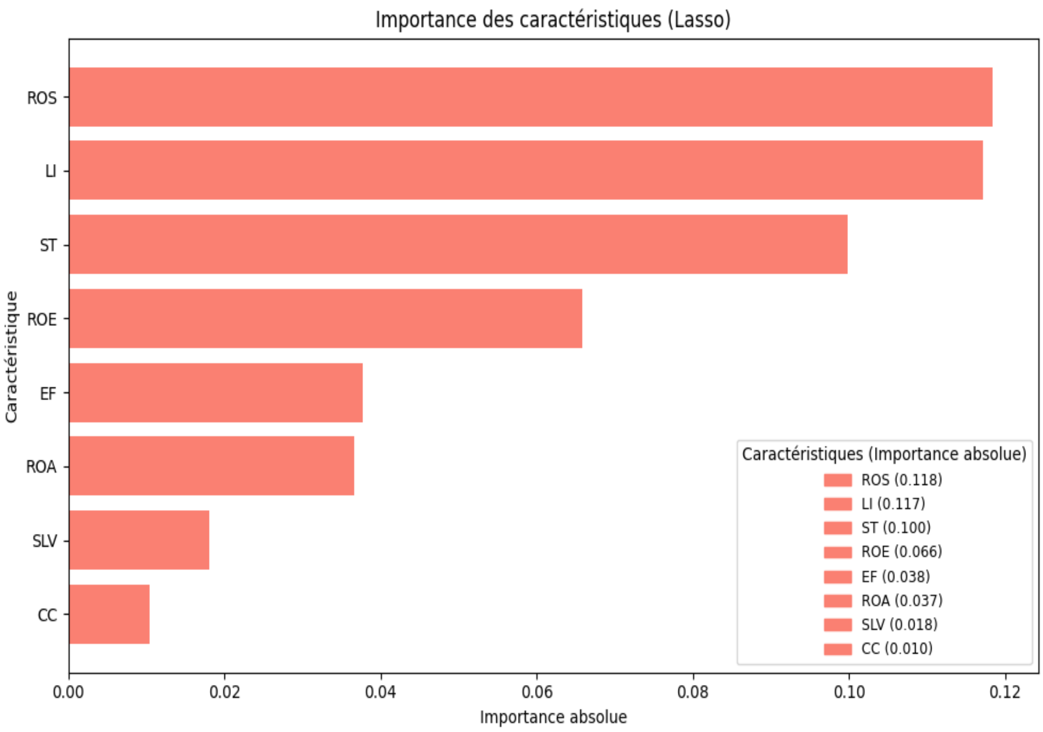

Figure 4. Importance of predictors according to the Lasso method

The results in figure 4 show that ROS (commercial profitability) and LI (overall liquidity) are the two most influential variables, with absolute importance scores of 0,118 and 0,117 respectively, closely followed by ST (financial structure) at 0,100. These variables therefore appear to be the most relevant for modeling. On the other hand, SLV (Solvency) and CC (Customer Management Coverage Capacity) have very low weights, suggesting a negligible contribution to the model.

Comparison between selection methods: Lasso and Random Forest

In order to evaluate the relevance of variable selection methods, we calculated the coefficient of determination (R^2) for models derived from Lasso and Random Forest. The coefficient of determination is defined by the following relationship:

$$\mathbf{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where y_i denotes the observed value, $\hat{y}_i$ the value predicted by the model, $\bar{y}$ the mean of the observed values, and n the total number of observations.

A R^2 close to 1 indicates an excellent fit of the model to the data, while a R^2 close to 0 suggests that the model explains very little of the observed variability. A negative value of R^2 indicates performance inferior to that of a trivial model based solely on the mean of the observations.

```
==========================================================================================
🔍 Comparaison des scores R² moyen obtenus par les méthodes de sélection Lasso et Random Forest.
------------------------------------------------------------------------------------------
R² moyen - Lasso : -1.2179531054354538
R² moyen - Random Forest : 0.9571347671138632
```

Figure 5. Comparison of (R^2) scores obtained using the Lasso and Random Forest selection methods

The results presented in figure 5 reveal that the Lasso model obtains a negative coefficient of determination (R^2 = -1,2179), indicating a performance inferior to that of a simple prediction based on the average of the observations. This observation suggests that the Lasso algorithm is not suitable, in this context, for effective selection of explanatory variables.

In contrast, the Random Forest model achieves a very high R^2 (0,9571), reflecting a remarkable ability to explain the variance of the target variable. These results confirm that the Random Forest approach is a more reliable and robust method for identifying relevant variables in our study.

On this basis, six explanatory variables were selected for further analysis: FS (Financial Structure), ROA (Economic Profitability), ROS (Commercial Profitability), ROE (Financial Profitability), LI (Liquidity), and SLV (Solvency). Their selection is based on their demonstrated relative importance in predicting corporate default. These variables will serve as the basis for the subsequent modeling steps, optimizing predictive performance while controlling the complexity of the model, thus ensuring an optimal compromise between accuracy and simplicity.

The results obtained are summarized in the table 4, highlighting the evaluation of the hypotheses formulated on the basis of the Random Forest model estimates.

The Random Forest model highlights that financial structure and profitability (commercial, economic, and financial) are the most predictive dimensions of business failure. Indicators such as financial equilibrium and accounts receivable, despite their theoretical relevance, appear to have little discriminating power in this empirical model.

Within the context of our study on Moroccan SMEs, the analysis of financial ratios reveals significant disparities between failing firms and those in good health. In light of the Random Forest algorithm's results, the variables with the highest predictive importance are financial structure (ST), operating profitability (ROS), and economic profitability (ROA). This ranking reflects the central role of operational performance and cash flow management in the survival dynamics of SMEs.

More specifically, non-failing firms are characterized by more effective cash management, reflecting a stronger capacity to meet short-term obligations. Conversely, failing enterprises display a clear imbalance between cash flows and short-term liabilities, indicating recurring financial stress. High operating profitability among healthy firms suggests better control of margins and more efficient management of operating costs, likely linked to sound strategic choices and a more relevant product or market orientation.

Furthermore, economic profitability (ROA) emerges as a key differentiating factor for failure. Robust SMEs succeed in generating an attractive return on assets, reflecting better utilization of available resources. In contrast, failing companies exhibit deteriorated ROA, often resulting from unprofitable investments or poor asset management.

Table 4. Validation of hypotheses based on variable importance analysis (Random Forest)

Variable	Importance	Associated Hypothesis	Analysis	Decision
LI (Liquidity)	0,059	Hypothesis 1	Low contribution, but higher than others – may be retained.	Validated
ST (Financial structure)	0,347	Hypothesis 2	Very high importance: an imbalance in financial structure is strongly linked to failure.	Validated
SLV (Solvency)	0,034	Hypothesis 3	Low importance, but not negligible; last one retained.	Validated
ROA (Economic profitability)	0,210	Hypothesis 4	Notable importance: plays a significant explanatory role in the model.	Validated
ROE (Financial profitability)	0,093	Hypothesis 5	Moderate but significant role: non-negligible impact.	Validated
ROS (Operating profitability)	0,231	Hypothesis 6	High importance: operating profitability is a clear discriminating factor.	Validated
EF (Financial equilibrium)	0,019	Hypothesis 7	Very low importance, close to the margin of error; negligible impact.	Rejected
CC (Accounts receivable)	0,007	Hypothesis 8	Very low impact; does not provide explanatory power to the model.	Rejected

It is also important to note that certain ratios traditionally emphasized in the literature such as financial profitability (ROE), liquidity (LI), and solvency (SLV) appear in our model with relatively low predictive weight. This suggests that, although important, these indicators alone are insufficient to explain failure in our sample and should be interpreted in conjunction with other financial dimensions.

In sum, our empirical analysis suggests that the failure of Moroccan SMEs is primarily associated with low operational profitability and inefficient cash management, whereas viable firms stand out for the strength of their operating margins and effective management of financial flows.

DISCUSSION

To address the questions surrounding the prediction of failure among Moroccan SMEs, hypotheses were formulated based on financial ratios likely to influence the occurrence of failure and to identify contexts of vulnerability within firms. These hypotheses could only be validated or rejected through empirical analysis. Accordingly, their evaluation relied on a comparison with the relative importance of explanatory variables obtained from the Random Forest model, which allowed for a rigorous assessment of each hypothesis and the identification of factors truly decisive in predicting bankruptcy risk.

The existing literature converges in identifying certain financial variables as key indicators of business failure. Since the pioneering work[6,46], indebtedness and financial structure (ST, SLV) have been recognized as central determinants. Economic and financial profitability (ROA, ROE), as well as operating profitability (ROS), have also been widely used to explain firms' vulnerability, particularly in contexts of high leverage or declining performance.[21,46,47] Moreover, liquidity (LI) and solvency (SLV) are consistently highlighted as essential factors for preventing insolvency and ensuring business continuity.[25,26,48] These studies confirm that satisfactory financial performance and prudent management of the financing structure are major levers for reducing the risk of failure.[49,50,51,52]

Based on this theoretical foundation, our study retains six key explanatory variables: ST (Financial Structure), ROA (Economic Profitability), ROS (Operating Profitability), ROE (Financial Profitability), LI (Liquidity), and SLV (Solvency). Identified as the most significant in the literature, these indicators form the basis of our predictive analyses, allowing us to optimize model performance while maintaining an economically interpretable structure.[53,54,55]

The superior performance of the Random Forest model can be largely explained by its ability to capture non-linear relationships and complex interactions among these financial variables. In the context of SME failure prediction, indicators often interact in intricate ways—for example, liquidity constraints may exacerbate solvency issues only under certain profitability levels, or financial leverage may have different effects depending on a firm's asset structure. As an ensemble learning method based on multiple decision trees, Random Forest naturally accommodates such interactions and non-linearities without requiring explicit specification. This flexibility enables it to model the heterogeneous and context-dependent patterns that characterize business failures more accurately than linear methods.[56,57]

In contrast, LASSO regression assumes linear relationships between predictors and the target variable.[58,59] While effective for dimensionality reduction and variable selection in linear contexts, LASSO may fail when underlying relationships are non-linear or involve higher-order interactions. Moreover, it is sensitive to multicollinearity, which can lead to important but correlated variables being assigned zero coefficients, further limiting its predictive power in complex datasets such as SME financial data.[60,61]

The identification of financial structure (ST) as the most significant predictor for Moroccan SMEs reflects the country's specific financial and economic context. Many SMEs in Morocco face structural financing constraints, including limited access to long-term credit, reliance on short-term debt, and strict banking requirements. Consequently, a firm's capital composition—its mix of equity and debt—strongly affects its resilience to financial stress and its likelihood of survival. The prominence of ST as a predictive factor underscores how local financial practices, banking policies, and the economic environment shape the determinants of business failure. This contextual insight represents a significant contribution, highlighting the importance of accounting for country-specific financial realities to achieve both accurate and practically relevant predictive models.

CONCLUSION

In conclusion, the empirical analysis conducted in this study clarified the main financial determinants of SME failure in the Fès-Meknès region of Morocco. Based on a balanced sample of 60 firms monitored over three accounting periods, the results show that financial structure, along with indicators of economic, operating, and financial profitability, as well as liquidity and solvency, play a central role in predicting the risk of failure. The combined use of Random Forest and Lasso variable selection methods allowed us to distinguish the most significant predictors, with Random Forest exhibiting particularly strong predictive power.

These findings underline that SME survival depends as much on the robustness of their financial structure as on their ability to generate profits and manage cash flows effectively. Variables traditionally emphasized in the literature, such as financial equilibrium or accounts receivable, appear less discriminating in this context, underscoring the importance of considering the most relevant financial dimensions in an integrated manner.

However, this study has certain limitations. The relatively small sample size and the focus on a single region restrict the generalization of the findings to all Moroccan SMEs. In addition, the number of financial ratios analyzed remains limited. The study also relies exclusively on quantitative data, without directly incorporating qualitative factors such as management practices, strategic choices, or managerial skills that may influence the risk of failure.

These limitations open several avenues for future research. It would be relevant to extend the analysis to other Moroccan regions and include larger samples to test the robustness of the results. Adding new financial ratios and qualitative or behavioral indicators would also provide a more nuanced understanding of the vulnerability mechanisms affecting SMEs. Finally, the application of advanced predictive methods, such as deep learning or hybrid models, could further improve the accuracy of predictions and enrich the analysis of business failure.

REFERENCES

1. Fitzpatrick F; 1932. A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms. Certified Public Accountant; 6:727-731.

2. Juglar C; 1862. Des crises commerciales et leur retour périodique en France, en Angleterre et aux États-Unis (rééd. 1935).

3. Kondratieff ND; 1935. The long waves in economic life. Review of Economics and Statistics; 17(6):105-115.

4. Keynes JM; 1937. The general theory of employment. Quarterly Journal of Economics; 51(2):209-223. https://doi.org/10.2307/1882087

5. Samuelson P; 1939. Interaction between the Multiplier Analysis and the Principle of Acceleration. Review of Economic Statistics; 4:75-78. https://doi.org/10.2307/1927758

6. Altman EI; 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance; 23:589-609. https://doi.org/10.2307/2978933

7. Michaux B; 1978. Profil et signification économique des faillites en Belgique. In A.-M. Kumps et al. (Éds.), Entreprises en difficulté et initiative publique (1). PUSL. https://doi.org/10.4000/books.pusl.9041

8. Malécot JF; 1991. Analyses théoriques des défaillances d'entreprises : une revue de la littérature. Revue d'Économie Financière, 19. https://doi.org/10.3406/ecofi.1991.1746

9. Jacquemin A; 1985. Économie d'entreprise. Paris : Dalloz.

10. Leibenstein H; 1966. Allocative Efficiency vs. X-Efficiency. American Economic Review; 56:392-415.

11. Sharma S, Mahajan V; 1980. Early Warning Indicators of Business Failure. Journal of Marketing; 44:80-89. https://doi.org/10.1177/002224298004400412

12. Picory C; 1994. PME, incertitude et organisation industrielle. Revue d'économie industrielle; 67(1):40-58.

13. Baldwin JR; 1998. Les faillites d'entreprise au Canada.

14. Lamontagne E, Thirion B; 2000. Les facteurs de survie : les qualités du projet priment sur celles du créateur. Insee Première, 703.

15. Casta JF, Zerbib JP; 1979. Prévoir la défaillance des entreprises ? Revue française de gestion, 97.

16. Baldwin CY, Mason SP; 1983. [Titre manquant]. American Finance Association; 38(2):505-516. https://doi.org/10.2307/2327985

17. Wruck KH; 1990. Financial distress, reorganization, and organizational efficiency. Journal of Financial Economics; 27(2):419-444. https://doi.org/10.1016/0304-405X(90)90063-6

18. Crutzen N, Van Caillie D; 2007. L'enchaînement des facteurs de défaillance de l'entreprise : Une réconciliation des approches organisationnelles et financières. Comptabilité et Environnement.

19. Bellovary JL, Giacomino DE, Akers MD; 2007. A review of bankruptcy prediction studies: 1930 to present. Journal of Financial Education; 1:1-42.

20. Altman EI, Iwanicz-Drozdowska M, Laitinen EK, Suvas A; 2017. Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. Journal of International Financial Management & Accounting; 28:131-171. https://doi.org/10.1111/jifm.12053

21. Mselmi N, Lahiani A, Hamza T; 2017. Financial distress prediction: The case of French SMEs. International Review of Financial Analysis; 50:67-80. https://doi.org/10.1016/j.irfa.2017.02.004

22. Daubie M, et al; 2005. Utilisation de variables non financières dans le cadre de la prédiction de faillite d'entreprises de moins de cinq ans : une approche multicritère pour le cas belge. GET/ENST Bretagne, 68-82.

23. Valaskova K, Kliestik T, Svabova L, Adamko P; 2018. Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis. Sustainability; 10:2144. https://doi.org/10.3390/su10072144

24. Charalambous C, Charitou A, Kaourou F; 2000. Comparative analysis of artificial neural network models: application in bankruptcy prediction. Annals of Operations Research; 99:403-425.

25. Back B, Laitinen T, Sere K, van Wezel M; 1996. Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms. Turku Centre for Computer Science Technical Report; 40:1-18. https://doi.org/10.1016/S0957-4174(96)00055-3

26. Bunn P, Redwood V; 2003. Company Accounts-Based Modelling of Business Failures and the Implications for Financial Stability. Bank of England Working Paper No. 210.

27. Kliestik T, Valaskova K, Lazaroiu G, Kovacova M, Vrbka J; 2020. Remaining financially healthy and competitive: The role of financial predictors. Journal of Competitiveness; 12:74-92. https://doi.org/10.7441/joc.2020.01.05

28. Ben Jabeur S; 2011. Statut de la faillite en théorie financière: approches théoriques et validations empiriques dans le contexte français. Thèse de doctorat, Université du Sud Toulon-Var.

29. Argenti J; 1976. Corporate Collapse: The Causes and Symptoms. Holsted Press, McGraw-Hill, London. https://doi.org/10.12691/ijbrm-2-1-4

30. Marco L; 1984. Les défaillances d'entreprises et la crise en France (1974-1983). Revue d'économie politique; 94(5):691-724.

31. Ooghe H, Van Wymeersch C; 1986. Modèles prévisionnels de la faillite. Annales de Droit de Liège; 3:183-196. https://doi.org/10.12691/ijbrm-2-1-4

32. Ohlson JA; 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 109-131. https://doi.org/10.2307/2490395

33. Camerer C, Lovallo D; 1999. Overconfidence and Excess Entry: An Experimental Approach. American Economic Review; 89:306-318. https://doi.org/10.1257/aer.89.1.306

34. Kaplan R, Norton D; 1992. The Balanced Scorecard—Measures That Drive Performance. Harvard Business Review, 79.

35. Lawrence PR, Lorsch JW; 1967. Differentiation and Integration in Complex Organizations. Administrative Science Quarterly; 12:1-47. http://dx.doi.org/10.2307/2391211

36. Mueller SA; 1991. The Opportunity Cost of Discipleship: Ethical Mutual Funds and Their Performance. Sociological Analysis; 52:111-124. https://doi.org/10.2307/3710719

37. Oglhe H, Van Wymeersch C; 1996. Traité d'analyse financière (6e éd.). Namur : Presses Universitaires de Namur.

38. Baldwin R; 1989. Measurable Dynamic Gains from Trade. National Bureau of Economic Research Working Paper No. 3147. https://doi.org/10.3386/w3147

39. Lukason O, Laitinen EK; 2019. Firm failure processes and components of failure risk. Journal of Business Research; 98:380-390. https://doi.org/10.1016/j.jbusres.2018.06.025

40. Myers SC; 1977. Determinants of Corporate Borrowing. Journal of Financial Economics; 5:147-175. https://doi.org/10.1016/0304-405X(77)90015-0

41. Holder M, Loeb J, Portier G; 1984. Le score de l'entreprise. Paris : Nouvelles Éditions Fiduciaires.

42. Brilman J; 1986. Gestion de crise et redressement d'entreprises. Paris : Éditions Hommes et Techniques.

43. Jaminon R; 1986. Facteurs explicatifs de faillites. Revue de la Faculté de Droit de l'Université de Liège; 3:197-207. https://doi.org/10.12691/ijbrm-2-1-4

44. Bardos M; 1995. Les défaillances d'entreprises dans l'industrie: ratios significatifs, processus de défaillances, détection précoce. Banque de France, Centrale des bilans, Collection Entreprises, janvier

45. Ooghe H, Waeyaert N; 2004. Oorzaken van faling en falingspaden. Economisch en Sociaal Tijdschrift; 57(4):367-393. https://doi.org/10.12691/ijbrm-2-1-4

46. Altman EI; 1984. The success of business failure prediction models: an international survey. Journal of Banking and Finance; 8(2):171-198. https://doi.org/10.1016/0378-4266(84)90003-7

47. Yim J, Mitchell H; 2002. A Comparison of Corporate Failure Models in Australia. RMIT Business School Working Paper, 10. https://doi.org/10.1007/3-540-45034-3_35

48. Gresse C; 1994. Les entreprises en difficulté. Paris : Economica.

49. Marchesnay M; 1988. La mercatique de la petite entreprise. Revue internationale PME; 1(3-4):258-276. https://doi.org/10.7202/1007884ar

50. Sharifabadi M, Mirhaj M, Izadinia N; 2017. The impact of financial ratios on the prediction of bankruptcy of SMEs. QUID: Investigación, Ciencia y Tecnología; 1:164-173.

51. Liou D, Smith M; 2007. Macroeconomic Variables and Financial Distress. Journal of Accounting, Business & Management; 14:17-31.

52. Svabova L, Michalkova L, Durica M, Nica E; 2020. Business failure prediction for Slovak SMEs. Sustainability; 12:4572. https://doi.org/10.3390/su12114572

53. Barron DN, West E, Hannan MT; 1994. A Time to Grow and a Time to Die: Growth and Mortality of Credit Unions in New York City (1914-1990). American Journal of Sociology; 100(2):381-421.

54. Crucifix F, Derni A; 1992. Symptômes de défaillance et stratégies pour le redressement d'entreprise. Louvain-La-Neuve : Académia.

55. Franks J, Sussman O; 2005. Financial innovations and corporate bankruptcy. Journal of Financial Intermediation; 14(3):283-317.

56. Lalonde RN, Moorcroft R; 1985. The Role of Attitudes and Motivation in Second Language Learning. Language Learning; 35:207-227. https://doi.org/10.1111/j.1467-1770.1985.tb01025.x

57. Liefhooghe B; 1997. Causes et mécanismes des faillites d'entreprises : une synthèse bibliographique. Cahiers de la Faculté des Sciences Économiques, Sociales et de Gestion; 189:1-45.

58. Koening G; 1985. Entreprise en difficulté : des symptômes aux remèdes. Revue Française de Gestion; 50:1-8.

59. Bescos PL; 1987. Défaillance et redressement des Pmi: Recherche des indices et des causes de défaillance. Cahier de Recherche du Cereg, 8701.

60. Hambrick DC, D'Aveni RA; 1988. Large Corporate Failures as Downward Spirals. Administrative Science Quarterly; 33:1-23. https://doi.org/10.2307/2392853

61. Kamaluddin A, Ishak N, Mohammed NF; 2019. Financial distress prediction through cash flow ratios analysis. International Journal of Financial Research; 10:63-76. https://doi.org/10.5430/ijfr.v10n3p63

FINANCING

No financing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Amina Guennoun, Saad Saadouni, Siham Ammari.
Data curation: Amina Guennoun.
Formal analysis: Amina Guennoun.
Research: Amina Guennoun, Souad Habbani.
Methodology: Amina Guennoun, Souad Habbani.
Project management: Souad Habbani.
Resources: Amina Guennoun, Siham Ammari, Saad Saadouni, Souad Habbani.
Software: Amina Guennoun, Saad Saadouni.
Supervision: Souad Habbani.
Validation: Amina Guennoun, Souad Habbani.
Display: Siham Ammari.
Drafting - original draft: Amina Guennoun.
Writing - proofreading and editing: Amina Guennoun, Saad Saadouni, Siham Ammari.