

Article

Exploring the Role of Global Value Chain Position in Economic Models for Bankruptcy Forecasting

Mélanie Croquet ¹, Loredana Cultrera ¹, Dimitri Laroutis ², Laetitia Pozniak ¹ and Guillaume Vermeyleen ^{3,*}

¹ Belgium-Soci&ter and Risk Research Institutes, Warocqué School of Business and Economics, University of Mons, BE-7000 Mons, Belgium; melanie.croquet@umons.ac.be (M.C.); loredana.cultrera@umons.ac.be (L.C.); laetitia.pozniak@umons.ac.be (L.P.)

² ESC Amiens, Centre de Recherche en Risk Management and Soci&ter Research Institute, FR-80000 Amiens, France; dimitri.laroutis@esc-amiens.com

³ CEBRIG and DULBEA Research Institutes, Warocqué School of Business and Economics, Belgium-Soci&ter, University of Mons, BE-7000 Mons, Belgium

* Correspondence: guillaume.vermeyleen@umons.ac.be; Tel.: +32-65373114

Abstract: This study addresses a significant gap in the literature by comparing the effectiveness of traditional statistical methods with artificial intelligence (AI) techniques in predicting bankruptcy among small and medium-sized enterprises (SMEs). Traditional bankruptcy prediction models often fail to account for the unique characteristics of SMEs, such as their vulnerability due to lean structures and reliance on short-term credit. This research utilizes a comprehensive database of 7104 Belgian SMEs to evaluate these models. Belgium was selected due to its unique regulatory and economic environment, which presents specific challenges and opportunities for bankruptcy prediction in SMEs. Our findings reveal that AI techniques significantly outperform traditional statistical methods in predicting bankruptcy, demonstrating superior predictive accuracy. Furthermore, our analysis highlights that a firm's position within the Global Value Chain (GVC) impacts prediction accuracy. Specifically, firms operating upstream in the production process show lower prediction performance, suggesting that bankruptcy risk may propagate upward along the value chain. This effect was measured by analyzing the firm's GVC position as a variable in the prediction models, with upstream firms exhibiting greater vulnerability to the financial distress of downstream partners. These insights are valuable for practitioners, emphasizing the need to consider specific performance factors based on the firm's position within the GVC when assessing bankruptcy risk. By integrating both AI techniques and GVC positioning into bankruptcy prediction models, this study provides a more nuanced understanding of bankruptcy risks for SMEs and offers practical guidance for managing and mitigating these risks.

Keywords: financial econometrics; artificial intelligence; forecasting; GVC



Citation: Croquet, M., Cultrera, L., Laroutis, D., Pozniak, L., and Vermeyleen, G. 2024. Exploring the Role of Global Value Chain Position in Economic Models for Bankruptcy Forecasting. *Econometrics* 12: 31. <https://doi.org/10.3390/econometrics12040031>

Academic Editor: Julien Chevallier

Received: 14 August 2024

Revised: 5 October 2024

Accepted: 28 October 2024

Published: 5 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bankruptcy significantly impacts not only the affected firm but also its broader network of stakeholders, including employees, customers, creditors, and suppliers (Weitzel and Jonsson 1989; Refait 2000; Daubie and Meskens 2001). The importance of predicting bankruptcy is therefore crucial, especially for small and medium-sized enterprises (SMEs), which form the backbone of most economies but are inherently more vulnerable to financial distress. Despite this, most bankruptcy prediction models have historically focused on larger firms, leaving a gap in understanding the specific challenges that SMEs face (Peel and Peel 1987; Storey et al. 1987; Keasey and Watson 1987; Altman and Sabato 2007; Crutzen and Van Caillie 2010; Zoricák et al. 2020; Papik and Papíková 2023).

SMEs operate under distinct conditions that set them apart from large corporations. Their lean organizational structures, centralized management, and dependence on external factors—such as customers, suppliers, and financial providers—make traditional

bankruptcy prediction models, which primarily rely on financial ratios, less effective for SMEs (Beaver 1966; Argenti 1976; Ohlson 1980; Altman 1984; Morris 1997; Julien 1997; Guilhot 2000; Van Caillie 2000; Daubie and Meskens 2001; Ooghe and De Prijcker 2008; Ciampi and Gordini 2008, 2009; Xu and Zhang 2009; Ogachi et al. 2020; Charalambous et al. 2021; Kitowski et al. 2022). Additionally, SMEs frequently rely on short-term credit and face difficulties in accessing long-term financing, leading to higher failure rates and added financial pressure (Peacock 2004). These factors necessitate specialized bankruptcy prediction models that are tailored to the complexities of SME operations.

A particularly critical gap in the literature is the limited attention to the effectiveness of different variable selection processes for predicting bankruptcy in SMEs. Moreover, few studies consider how a firm's position within the Global Value Chain (GVC) can influence bankruptcy risk. Firms that operate upstream in the GVC, such as manufacturers, may be more exposed to the financial failures of downstream firms, making traditional models insufficient (Fujiwara 2008; Hua et al. 2011; Delli Gatti et al. 2009). This raises the need for prediction models that can account for these contagion effects and other non-financial risk factors.

To address these gaps, this study proposes the application of artificial intelligence (AI) techniques to enhance bankruptcy prediction models for SMEs. AI has proven highly effective in handling complex, real-world data in various fields, including finance, where it can identify patterns that traditional statistical methods may overlook. For example, the study of Shetty et al. (2022) successfully applied machine learning algorithms to predict financial distress, showing that AI can process diverse, multi-dimensional data and deliver accurate forecasts in challenging environments.

Building on this foundation, this study evaluates and compares the effectiveness of two broad categories of variable selection methods: (i) traditional statistical techniques such as backward and forward selection, and (ii) AI-based methods, including the lasso algorithm and Classification and Regression Trees (CART). The goal is to determine which methods are most effective in predicting bankruptcy for SMEs, thereby addressing the specific data-selection challenges that arise from their unique operational environments. The ability of AI techniques to analyze large data sets, identify non-obvious patterns, and handle nonlinear relationships makes them particularly well-suited for tackling the complexities of SME bankruptcy prediction.

Additionally, we explore the role of a firm's GVC position in shaping bankruptcy risk. By integrating a firm's position within the GVC into the model, we aim to improve the accuracy of bankruptcy predictions, particularly for upstream firms that may be vulnerable to downstream financial failures. This analysis contributes to a more comprehensive understanding of how sector-specific variables influence bankruptcy prediction (Chava and Jarrow 2004), especially in industries where firms operate in interdependent relationships.

Finally, it is worth mentioning that this paper is focusing on the Belgian framework, and that Belgium's bankruptcy laws, particularly as they relate to SMEs, have important differences from those in other countries that have been studied. These differences in legal frameworks, creditor protections, cultural attitudes, and procedural efficiency can affect both the likelihood of a firm entering bankruptcy and the outcomes of those proceedings. SMEs in Belgium may face higher regulatory burdens, slower bankruptcy processes, and fewer restructuring opportunities than in more debtor-friendly jurisdictions like the U.S., UK, or other European nations with streamlined SME-specific insolvency processes. These factors are important to consider when studying Belgian bankruptcies and comparing them to findings from other countries.

This study aims to fill the existing gaps in the literature by focusing on the unique characteristics of SMEs and leveraging AI techniques to improve bankruptcy prediction. By comparing traditional statistical methods with AI-based approaches, and by incorporating the GVC position into our analysis, we aim to provide actionable insights that are relevant not only for the academic community but also for practitioners and stakeholders working with SMEs.

The paper is structured as follows: Section 2 discusses the variable selection methods and data used. Section 3 presents the empirical results, comparing the effectiveness of different methods. Finally, Section 4 discusses the implications of our findings for both theory and practice.

2. Materials and Methods

2.1. Data

The empirical data for this study are collected from the Bureau Van Dijk database (hereafter, BVD), namely Belfirst (i.e., Financial Reports and Statistics on Belgian and Luxembourg Companies), which gathers detailed information over 32,000 Belgian firms and 4000 firms established in Luxembourg. We first select a balanced sample of Belgian SMEs covering the period 2002–2012. Surrounding the 2008 financial crisis in a paper on bankruptcy offers several compelling arguments:

- The 2008 crisis represents a key period of global economic turmoil, leading to a sharp increase in bankruptcies across multiple sectors. By focusing on this period, the paper can highlight how extreme economic conditions test the resilience of firms;
- Firms faced unique stressors during the 2008 crisis, including sudden liquidity crunches, declining consumer demand, and restricted access to credit. The crisis offers a rich context for exploring the factors that led to firm bankruptcies, and which companies were more vulnerable;
- The 2008 crisis reshaped many financial and economic theories regarding risk, leverage, and the sustainability of business models. Addressing this period allows a discussion of how these theoretical shifts have influenced bankruptcy prediction models and risk assessment tools.

By analyzing bankruptcy during this significant period, the paper can contribute to a deeper understanding of how firms respond to economic shocks, how policies shape outcomes, and how future crises might be better managed.

After removing firms with missing or inappropriate variables (such as negative net sales, for example), the number of firms registered as bankrupt is 3552¹. Then, the same number of non-failing firms is also included in the sample, based on a random selection, leading the total balanced sample to 7104 Belgian SMEs². This way of selecting firms has several advantages: it helps balance the data set, prevent model bias, and ensures robust predictions, making the model more effective at predicting bankruptcies.

2.2. Variables

2.2.1. Bankruptcy Indicator

In order to build our prediction model, a bankruptcy indicator is needed as a dummy variable that equals 1 if the firm has experienced corporate bankruptcy over the investigated period, 0 otherwise. For the set of explanatory variables, and following [Shumway \(2001\)](#), we extract a time-varying panel data set where each firm-year observation in our sample is treated as a separate observation. The final data set therefore contains all the predictor variables used in this study and the binary response variable indicating the firm's bankruptcy status. This panel data structure allows us to consider all the information of the firms as a potential predictor of the future bankruptcy and should provide consistent and accurate out-of-sample prediction ([Mai et al. 2019](#)).

2.2.2. Predictors

Using firms' balance sheets and income statements, we calculate 50 financial ratios commonly used in the literature when predicting corporate bankruptcy ([Beaver 1966](#); [Altman 1968](#); [Ohlson 1980](#); [Dimitras et al. 1996](#); [Dimitras et al. 1999](#); [Balcaen and Ooghe 2006](#); [Agarwal and Taffler 2008](#); [Amendola et al. 2011](#); [Jackson and Wood 2013](#); [Wang et al. 2014](#); [Altman and Branch 2015](#); [Kim et al. 2016](#); [Veganzones and Séverin 2018](#)). We then restricted the number of variables to the most noteworthy 30 variables. The reduction of financial ratios was primarily driven by the need to address multicollinearity,

a common issue in regression models where predictor variables are highly correlated. To tackle this, the study used Variance Inflation Factor (VIF) analysis to identify and remove variables with VIF values higher than 10, which indicated problematic multicollinearity (O'Brien 2007). By eliminating such variables, the model's stability and reliability were improved, making the regression coefficients more interpretable and reducing the risk of overfitting. The remaining variables were further narrowed down based on statistical significance, ensuring that only those with a strong predictive relationship with bankruptcy were retained. Additionally, variables were selected for their relevance to the research question, prioritizing ratios that were known to be important indicators of financial health, such as profitability, liquidity, and leverage ratios. This process aimed to strike a balance between model parsimony and predictive power, ensuring that the final model was efficient without compromising accuracy, which are reported in Table 1. Still, and since including all 30 financial ratios would lead to a very high-dimensional feature space, reducing the model's predictive power (Veganzones and Séverin 2018), we perform variable selection processes allowing us to robustly reduce the predictors to a subset of the most relevant financial ratios.

Table 1. List of Potential Corporate Bankruptcy Predictors.

Variable	Abbreviation
Firm age	Firm age
Log of total employees	Firm size
Shareholder equity/Total assets	Solvency
EBIT/Total assets	Profitability
Current ratio	Current
Accruals/Total assets	Accruals
Value added/Total workers	VA/TW
Acid test	Acid
Log of sales	LnS
Number of days of client credit	ClientDays
Number of days of supplier credit	SupplierDays
Long term debts/Total assets	LTD/TA
Gross sales margin	GSM
Net sales margin	NSM
Net income/Sales	NI/S
Net income/Total assets	NI/TA
Cash-flow/Equity	CF/E
Cash-flow/Total debts	CF/TD
Net Working Capital/Sales	NWC/S
Net Working Capital/Total assets	NWC/TA
Long term debts/Equity	LTD/E
Total debts/Total assets	TD/TA
Net income/Current debts	NI/CD
EBITDA/Total debts	EBITDA/TD
Cash-flow/Current assets	CF/CA
Net income/Total debts	NI/TD
Net income/Current assets	NI/CA
Current debts/Sales	CD/S
Tax expense/Total assets	Tax/TA

2.3. Variables Selection Methods

The variable selection method used may influence the performance of prediction models (Veganzones and Séverin 2018; du Jardin et al. 2019). Therefore, we implement four different selection techniques in order to select the best predictors among those listed above. These techniques can be gathered in two main categories: statistical techniques, with the backward and forward procedures; and artificial intelligence-based techniques, with the machine learning Lasso (Least Absolute Shrinkage and Selection Operator) and the ensemble classification and regression trees (CART) procedures. These four methods were

chosen because they complement one another's strengths. While backward and forward selection offer simplicity and interpretability in smaller data sets, Lasso and CART handle complexity and multicollinearity, making them ideal for tackling more challenging aspects of SME bankruptcy prediction. By applying a mix of these techniques, the study can assess which variables matter most under different modeling approaches, offering a more robust and comprehensive understanding of bankruptcy risk factors.

2.3.1. Statistical Techniques

Backward selection consists in starting with all potential predictors, testing the deletion of each potential predictor using a chosen model fit criterion, deleting the potential predictor (if any) whose loss gives the most statistically insignificant alteration of the model fit, and repeating this process until no further potential predictors can be deleted without a statistically significant loss of fit. It is thus an iterative search procedure that identifies which independent variables, previously thought to be of some importance, have the strongest predictive power of the dependent variable (Cultrera and Brédart 2016).

Econometrically, we start with a full model, which includes all predictors:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k \quad (1)$$

where:

- \hat{y} is the predicted value (dependent variable);
- β_0 is the intercept;
- β_ix_i is the set of variables i , i going from 0 to k .

It then removes the least significant predictor, i.e., the predictor with the least significance (based on the highest p -value or smallest improvement in R2):

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{k-1}x_{k-1} \quad (2)$$

It finally repeats the process, removing predictors step by step until all remaining variables significantly improve the model based on the chosen criterion:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{k-n}x_{k-n} \quad (3)$$

As mentioned, several criteria can be used to select the predictors during the backward selection. We use a p -value threshold of 0.05 as a criterion to exclude the weakest predictors from the model at each step. Once variables have been dropped, they cannot reenter the equation. Accordingly, the software automatically drops one by one variable with the highest p -value until only those that are significant at the 0.05 level are remaining.

The forward procedure, on the other hand, starts from the null model, and step by step, chooses the best model with one additional predictor. At each step, the predictor satisfying the entry criterion is added to the model (Tsai 2009).

Econometrically, we start with no predictors in the model and add them one by one, based on some criterion (e.g., p -value, F-Stat, or adjusted R2):

$$\hat{y} = \beta_0 \quad (4)$$

where:

- \hat{y} is the predicted value (dependent variable);
- β_0 is the intercept.

It then adds predictor variables one at a time. For each predictor variable x_i , the model becomes:

$$\hat{y} = \beta_0 + \beta_ix_i \quad (5)$$

The model is evaluated based on the selected criterion. The predictor with the best improvement is added to the model. More precisely, the selection method sequentially

includes variables based on the F-stat until adding more variables is not able to improve the model, with the final model including only significant predictors (Miller 2002). In order to select our predictors during the process, we followed the Tsai's (2009) criteria, i.e., a variable is included in the model if the F stat is less than 0.05 and removed from the model if the F stat is more than 0.10 (Bauweraerts 2016).

2.3.2. Artificial Intelligence Techniques

There is a cost to including lots of regressors, and we can reduce the objective function by throwing out those that contribute little to the model's fit. Clearly, fitting a model with a higher polynomial degree is useful only if it significantly reduces the error compared to a simpler model. Penalized logistic regression imposes a penalty to the logistic model for having too many variables, leading to selection variables processes to be even more crucial. This model, also known as regularization, shrinks the coefficients of the less contributive variables toward zero (James et al. 2013; Kassambara 2018). As a consequence, the variance of the model is reduced. The mostly used penalized regression methods are the ridge regression and the Lasso regression (Tibshirani 1996; Park and Casella 2008), with the later having the advantage of involving in the final model only a subset of the predictors, which in turn improves the model interpretability by eliminating irrelevant variables not associated with the response variable, reducing overfitting (James et al. 2013; Fonti and Belitser 2017). Controlling the trade-off between error and complexity of the model, the Lasso procedure penalizes the model by setting the coefficients for some variables to zero. It provides a coefficients matrix which is a continuous linear function of a tuning parameter λ that controls the strength of the penalty and is computed through cross-validation process. As λ increases, variables are settled to zero and are removed from the model. The only need is to settle the penalty level, i.e., the λ . Several ways of selecting λ are available and follow the information criteria approach developed by Chen and Chen (2008), who introduce the Extended BIC (Bayesian Information Criterion) computation, imposing an additional penalty on the number of parameters.

Econometrically, the Lasso method solves the following optimization problem:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (6)$$

where:

- $\hat{y}_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$;
- y_i is the actual value for observation i ;
- β_0 is the intercept;
- β_j are the coefficients for each predictor x_j ;
- λ is the tuning parameter that controls the strength of the penalty.

The economic reasoning is that without penalty ($\lambda = 0$), the Equation (6) becomes a standard linear regression. However, with penalty ($\lambda > 0$), the Lasso strategy shrinks the coefficients, and some of them may become exactly zero, effectively selecting a subset of the most important predictors.

Classification and regression tree is a recursive algorithm in data mining which explores the structure of a data set and develops visualised decision rules for predicting a categorical variable, i.e., classification tree, and a continuous variable, i.e., regression tree (Brezigar-Masten and Masten 2012; Singh et al. 2015; Choubin et al. 2018). One advantage in relying on CART compared to other artificial intelligence methods is that it provides easily understandable decision rules (Li et al. 2010). Classification trees use a binary tree to recursively partition the predictor space into subsets. The terminal nodes of the tree correspond to the distinct regions of the partition, and the partition is determined by splitting rules associated with each of the internal nodes. By moving from the root node through to the terminal node of the tree, each observation is then assigned to a unique terminal node. CART analysis is nonparametric and can detect complex relationships

between dependent variable and explanatory variables (Brezigar-Masten and Masten 2012). Therefore, CART analysis is particularly suited for discovering nonlinear structures and variables interactions in data sets with a large number of potential explanatory variables.

Econometrically, CART aims to minimize the sum of squared errors (SSE) for each split. The splitting rule is based on reducing the variance in the target variable. For a split at node t , the goal is to minimize:

$$\sum_{i \in R_1(t)} (y_i - \bar{y}_{R_1})^2 + \sum_{i \in R_2(t)} (y_i - \bar{y}_{R_2})^2 \tag{7}$$

where:

- $R_1(t)$ and $R_2(t)$ are the regions (subsets of data) defined by the split at node t ;
- \bar{y}_{R_1} and \bar{y}_{R_2} are the mean values of the target variable in the respective regions.

The economic reasoning starts with the entire data set at the root node. At each step, we split the data into two child nodes using the predictor x_j and split value s that minimize the loss function (i.e., the sum of squared errors). We then repeat the process recursively until a stopping criterion is met (i.e., the minimum number of observations in a node).

2.4. The Position in the GVC

In order to take into account the potential influence of the position of the firm in the GVC, we isolate from the data set the group of firms we may define as upstream, i.e., those being part of the following industrial sectors (NACE Rev 2): Agriculture, forestry and fishing (NACE A), Mining and Quarrying (NACE B), Manufacturing (NACE C), Electricity, gas, steam and air conditioning supply (NACE D) and Water supply (NACE E), representing a balanced subsample of 882 firms (441 firms registered as bankrupt and 441 non-failing firms).

2.5. The Logistic Model and Misclassification Evaluation

Since our dependent variable, i.e., bankruptcy, is a dichotomous variable, trying to estimate the relation through a linear regression may lead to estimations qualified as aberrant and biased, that is to say a dependant variable exceeding the limits of 0 and 1. We then use as estimation techniques a logistic regression model, fitted to binomial variables. This model is particularly used in the majority of studies focusing on the occurrence of the bankruptcy phenomenon (Ohlson 1980; Premachandra et al. 2009). A logistic regression ties a link between a dependant variable that takes two different values: 1 if the firm is bankrupt and 0 if it is healthy, and k other explanatory variables. These variables can be quantitative or qualitative. The first step is to identify that our response variable is binary, and then defined as:

$$y_i = \begin{cases} 1 & \text{if (the } i^{th} \text{ firm is bankrupt)} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

y_i can be seen as a realization of the random variable that takes the value one or zero with the probability P_i and $1 - P_i$, respectively, and with the distribution of Y_i known as a Bernoulli distribution. This distribution can be written for $y_i = 0, 1$ as:

$$Pr[Y_i = y_i] = P_i^{y_i} (1 - P_i)^{1-y_i} \tag{9}$$

Then, the probability to go bankrupt, P_i , depends on observed covariates X_i . Given that ordinary least squares techniques do not ensure that the predicted value of the dependent variable will be in the correct range between zero and one unless complex restrictions are settled, we need as a second step to transform the probability in order to remove these range restrictions. To do so, we first compute the odds as the ratio of favourable to unfavourable cases, that is:

$$odds_i = \frac{P_i}{1 - P_i} \tag{10}$$

Secondly, we transform the odds into logarithms, which gives the logit such as:

$$\text{logit}(P_i) = \log \frac{P_i}{1 - P_i} \quad (11)$$

This leads to the fact that if the probability to go bankrupt goes down to zero, the odds approach zero and the logit approaches $-\infty$. Conversely, if the probability to go bankrupt approaches one, the odds approach $+\infty$ and the logit approaches $+\infty$ also.

The logistic regression model therefore assumes a linear function of the predictors, with a vector of covariates and the related coefficients, such as:

$$\text{logit}(P_i) = x_i \beta \quad (12)$$

From that point, coefficients may be interpreted as in linear models, but we may be careful that the left-hand-side is a logit rather than a mean as it is in ordinary least squares for example. Practically, each represents the change in the logit of the probability associated to a one-unit change in the given predictor. In turn, we are only able to know, through the observation of the regression coefficients, the direction and the significance of the relation between the dependent variable and the independent variables but not the magnitude of this variation. And so, a positive coefficient will show that the independent variable increases the probability that y_i takes the value 1, but we have no information about the way the magnitude of this probability increase. To know this, we compute the marginal effects:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \frac{\partial P(y_i = 1|x_i)}{\partial x_{ik}} = \frac{e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} \beta_k \quad (13)$$

The marginal effects measure the impact of a change in the variable x_i on the probability for the dependent variable y_i to take the value of 1. In other words, the marginal effects of a regressor represents how much the (conditional) probability of the outcome variable changes when we change the value of a regressor by one unit, all other regressors remaining constant at given values.

In order to evaluate the prediction capacity of variables selected through the aforementioned procedures, we select four evaluation metrics: sensitivity, specificity, overall classification, and area under the receiver operating characteristic (ROC) curve (AUC). Classification metrics are computed with a 10-fold cross-validation method that has been repeated 50 times, representing an average performance of 500 individual testing subsets. This out-of-sample prediction is also in line with the current BASEL III recommendations for default model validation purpose (Mai et al. 2019). We therefore randomly separate the data set by selecting 70% of the data as the training set and the remaining 30% as the testing set, which is common in previous researches (du Jardin 2016; Cultrera and Brédart 2016; Cultrera and Vermeylen 2018, among others).

Sensitivity and specificity metrics focus on evaluating a type of firm sensitivity for failed firms and specificity for non-failed firms, such as (Veganzones and Séverin 2018):

$$\text{Sensitivity} = \frac{\text{Class}_{\text{Bankrupt}}}{\text{Class}_{\text{Bankrupt}} + \text{MissClass}_{\text{Bankrupt}}} \quad (14)$$

$$\text{Specificity} = \frac{\text{Class}_{\text{Non_Bankrupt}}}{\text{Class}_{\text{Non_Bankrupt}} + \text{MissClass}_{\text{Non_Bankrupt}}} \quad (15)$$

where *Sensitivity* represents the percentage of bankrupt firms correctly classified, *Specificity* is the percentage of non-bankrupt firms correctly classified, $\text{Class}_{\text{Bankrupt}}$ represents bankrupt firms correctly classified, $\text{MissClass}_{\text{Bankrupt}}$ represents bankrupt firms misclassified; $\text{Class}_{\text{Non_Bankrupt}}$ represents non-bankrupt firms correctly classified, and $\text{MissClass}_{\text{Non_Bankrupt}}$ represents non-bankrupt firms misclassified.

Classical models of classification usually compute sensitivity and specificity metrics on the bases of an average risk score (cutoff) of 0.5. This article also improves the accuracy evaluation through computing the best cutoff points that fit the data, by providing graphs of sensitivity versus risk score and specificity versus risk score. This allows us to compute sensitivity and specificity statistics with respect to the more precise cutoff points.

Then, we compute the AUC metric, allowing us to evaluate overall performance. AUC provides a graphical representation of the trade-off between a failed firm that has been correctly classified and a failed firm that has been incorrectly classified. For a classifier, the ROC curve measures how well a model discriminates between bankrupt and non-bankrupt firms and needs to be graphically as far to the top left corner as possible, where its value will be close to 1. A value of 0.5 means a random model with no predictive ability. At the extreme, a value of 1 means perfect discrimination (Chava and Jarrow 2004). Current literature on bankruptcy prediction referring to ROC analysis shows a ROC statistic ranging between 0.7392 (Zmijewski 1984) and 0.9113 (Shumway 2001), with the pioneer Altman's model (Altman 1968) standing in the middle with a ROC area of 0.8662.

The calculations, modelling, and estimation procedures were performed using STATA/SE 18.0 software, ensuring robust and reliable statistical analysis throughout the study.

3. Results

3.1. Variable Selection

Applying the aforementioned variable selection procedures allows us to select a short list of accurate predictors among the initial set of 30. The Table 2 gathers the information extracted from each selection procedure. Focussing on the CART procedure (see Figure 1) and following Cho et al. (2010), the variables identified enter as key variables into our bankruptcy prediction model (Model 1 as denoted in the fourth column of Table 2). But this estimation strategy does not take advantage of the information given by the classification tree as nonlinearities in the relation between each predictor selected and the probability of bankruptcy as highlighted in each node. A way to capture this information is to create a set of dummy variables taking the value 1 if the values of the variable that defines a branch fall into the region above the CART threshold and zero otherwise. All dummy variables are therefore implemented in the prediction Model 2 of the Table 2.

Table 2. Key predictors selected by each selection procedure—Overall sample.

Backward	Forward	Lasso	CART (Model 1)	CART (Model 2) *
Profitability	Profitability	Profitability	Profitability	Profitability_D1
Solvency	Solvency	Solvency	Solvency	Profitability_D2
Current	Current	Current	VA/TW	Solvency_D1
VA/TW	VA/TW	VA/TW	VA/TAX	Solvency_D2
Firm size	Firm Size	CF/TD	CF/TD	Solvency_D3
Firm age	Accruals	ClientDays		Solvency_D4
	NWC/TA	SupplierDays		VA/TAX_D
		LTD/TA		VA/TW_D1
				VA/TW_D2
				CF/TD_D1
				CF/TD_D2

* With the corresponding thresholds used: Profitability_D1 = 1 if Profitability \leq 21.24; 0 otherwise; Profitability_D2 = 1 if Profitability \leq 20.43; 0 otherwise; Solvency_D1 = 1 if Solvency \leq 20.69; 0 otherwise; Solvency_D2 = 1 if Solvency \leq 24.04; 0 otherwise; Solvency_D3 = 1 if Solvency \leq 40.78; 0 otherwise; Solvency_D4 = 1 if Solvency \leq 8.63; 0 otherwise; VA/Tax_D = 1 if VA/Tax \leq 5.08; 0 otherwise; VA/TW_D1 = 1 if VA/TW \leq 71.02; 0 otherwise; VA/TW_D2 = 1 if VA/TW \leq 63.13; 0 otherwise; CF/TD_D1 = 1 if CF/TD \leq 0.07; 0 otherwise; CF/TD_D2 = 1 if CF/TD \leq -0.035; 0 otherwise.

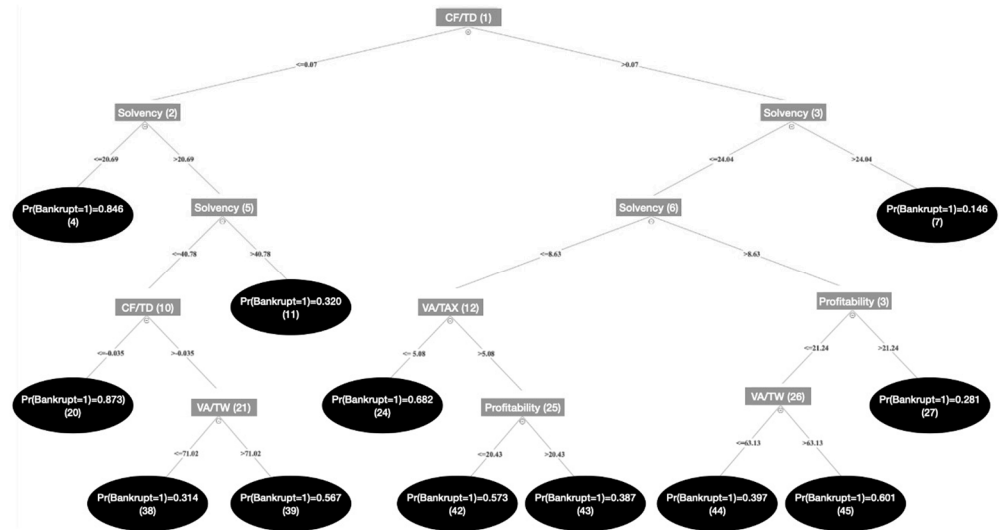


Figure 1. Classification Tree—Overall sample.

3.2. The Position in the GVC

In a second step, we reimplement the estimation strategy according to whether the firm belongs to an industrial sector located upstream in the GVC. The four variable selection techniques have therefore been run on the subsample of 882 firms, leading to the subsets of predictors presented in Table 3 and Figure 2. Globally, the predictors selected are the same as for the whole sample of firms, though the selection processes seem to be more restrictive.

Table 3. Key predictors selected by each selection procedure—Upstream firms.

Backward	Forward	Lasso	CART (Model 1)	CART (Model 2) *
Profitability	Profitability	Profitability	CF/TD	CF/TD_D
Solvency	Solvency	Solvency	Solvency	Solvency_D
Current	Current	Current		
VA/TW	VA/TW	VA/TW		
Firm size		CF/TD		
		Firm size		

* With the corresponding thresholds used: CF/TD_D = 1 if CF/TD \leq 0.085; 0 otherwise; Solvency_D = 1 if Solvency \leq 23.38; 0 otherwise.

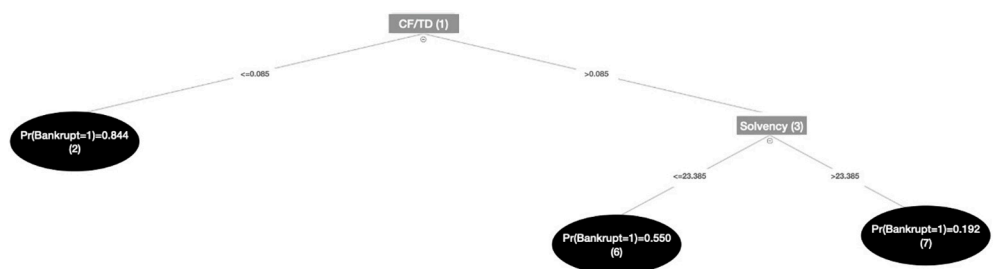


Figure 2. Classification Tree—Upstream firms.

3.3. Accuracy and Misclassification Evaluation

Table 4 presents the four evaluation metrics (sensitivity, specificity, overall classification, and area under the receiver operating characteristic-ROC curve-AUC) obtained by running the logit model on each variable selection procedures, which allow us to compare performance associated with each variable selection process⁴. First, and in order to be more accurate in the evaluations, we compute the cutoff points to be implemented in classification analysis (fifth row). Then, ROC areas are all significant and imply models with a

concrete predicting ability (ROC statistic ranging between 0.7021 and 0.8662 according to the selection method used, fourth row of Table 4). Evaluation metrics also show that, on the whole, prediction performance increases when intelligence selection techniques (both Lasso and CART models) are implemented rather than statistical techniques (third row). Within the intelligence techniques, and regarding sensitivity analysis presented in the first row, which represents the major statistic since it represents the capability of our model to recognize bankrupt firms, results clearly show that the CART selection procedure (Model 2) outperforms other selection methods. It therefore seems that the CART selection procedure better fit the specific characteristics of Belgian SMEs, generating discrimination rules that correspond to the data.

Table 4. Accuracy evaluation rates achieved with prediction model—Overall sample.

	Backward	Forward	Lasso	CART (Model 1)	CART (Model 2)
Sensitivity	49.45%	62.48%	75.37%	78.56%	84.28%
Specificity	94.67%	89.86%	78.38%	80.24%	77.72%
Correctly classified	74.79%	74.96%	76.76%	79.22%	80.70%
ROC	0.8489	0.7021	0.8425	0.8365	0.8662
Cutoff	0.60	0.60	0.60	0.60	0.625

In a second step, we measure the influence of the position of the firm in the GVC on the prediction capacity of our models. As mentioned before, this allows us to test whether financial ratios may be as good predictors of bankruptcy for upstream firms as for the overall sample. The results presented in Table 5 first show that ROC areas (fourth row) still confirm the predicting power of our models (ROC statistic ranging between 0.7248 and 0.8521 depending on the selection method used). Then, sensitivity (first row) is sharply decreasing, suggesting that the position in the GVC may have an effect on the capability for financial ratios to predict bankruptcy. Therefore, for such firms, there seems to exist some noise in the forecasting strategy, providing support to the presence of other variables affecting bankruptcy for these upstream firms such as contagion effect that could be more pronounced in such upstream industrial sectors (Fujiwara 2008; Delli Gatti et al. 2009; Hua et al. 2011).

Table 5. Accuracy evaluation rates achieved with prediction model—Upstream sample.

	Backward	Forward	Lasso	CART (Model 1)	CART (Model 2)
Sensitivity	42.42%	55.88%	47.37%	66.10%	69.84%
Specificity	79.19%	71.03%	96.15%	83.22%	62.22%
Correctly classified	65.81%	67.38%	75.56%	73.75%	66.67%
ROC	0.7391	0.7248	0.8473	0.8521	0.7278
Cutoff	0.65	0.65	0.75	0.625	0.75

3.4. Non-Linearities Further Investigations

In order to better capture complex and non-linear relationships, we further incorporate Principal Component Analysis (PCA) for dimensionality reduction and multicollinearity, together with Extreme Gradient Boosting (XGBoost)⁵.

For a set of variables X_1, X_2, \dots, X_p , a principal component is a linear combination of these variables that maximizes the variance in the data. If all the variables are represented in matrix X , the first principal component is defined as:

$$W_1^T X \quad (16)$$

where the weight coefficient vector W_1 is chosen to maximize the variance. The second principal component, $W_2^T X$, is constructed to be uncorrelated with the first and to have the

second highest variance, and this process continues for subsequent components. Using the Lagrange multiplier technique, it is shown that the optimal weight vector W_i corresponds to the eigenvectors of the covariance matrix of X , with the i -th eigenvector associated with the i -th largest eigenvalue λ_i .

In general, the transformation from the original variable matrix X to the principal component matrix Y is expressed as $Y = W X$, where W is the matrix of eigenvectors from the covariance matrix of X . The position of each observation in the principal component space, known as the score, is given by $Z = XW^T$. The total variance explained by the first k principal components, V_k , is:

$$V_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \quad (17)$$

In our analysis, we also refer to X as the data matrix rather than the variable matrix. Following Gower and Dijksterhuis (2004), the number of principal components to retain is often based on criteria such as capturing 70–90% of the total variance, which preserves most of the information in X . Another criterion is the magnitude of the eigenvalues; components with variance less than one ($\lambda_k < 1$) are considered less informative and can be excluded.

Applied to our data, the results show that retaining components with eigenvalues higher or equal to one suggest we can limit our analyses to five components⁶. Computing the scores of the five first components allows us to next implement XGBoost on a restricted sample.

Boosting is an ensemble learning technique that builds a strong classifier by combining multiple weak classifiers, each of which is relatively simple to train (Schapire 1999). A widely used variant of this approach is Extreme Gradient Boosting (XGBoost), which enhances the standard gradient boosting method by incorporating a more regularized model to prevent overfitting (Chen and Guestrin 2016). XGBoost, along with deep learning, ranks among the most effective algorithms for large-scale data classification and has been key to many winning solutions in Kaggle machine learning competitions. It is versatile and implemented in various programming languages, including Python, where it is integrated into the Scikit-Learn library.

XGBoost has gained significant attention in recent years, especially in the areas of credit scoring and bankruptcy prediction (Carmona et al. 2019; Shetty et al. 2022; Son et al. 2019). According to Chen and Guestrin (2016), XGBoost combines regression trees and gradient boosting, with each tree in the training process improving upon the residual errors of the previous one to optimize the objective function. This iterative process reduces model complexity and helps avoid overfitting. The final model is the result of combining all trees, and the prediction output is given by the formula:

$$Z = G(X_i) = \sum_{j=1}^K g_j(X_i) \quad (18)$$

where X_i is the financial ratio and $g_j(X_i)$ the output function of each tree.

In addition to its strong performance, XGBoost provides several practical advantages, such as parallelization for faster execution and built-in handling of missing data, making it a favored approach in structured data tasks across various domains. The method's flexibility and efficiency have positioned it as a valuable tool in both academic research and practical applications in finance and beyond. Results in Table 6 present a slightly but significant increase in the model accuracy to predict bankruptcy. More precisely, we can expect a 1.61%-point increase in the number of firm correctly classified compared to the best model CART (Model 2) for the overall sample, and a 2.42% point increase compared to Lasso estimates when focusing on upstream firms.

Table 6. Accuracy evaluation rates achieved with prediction model.

	XGBoost-Overall	XGBoost-Upstream
Sensitivity	77.45%	65.89%
Specificity	88.56%	84.22%
Correctly classified	82.31%	77.98%

3.5. Accuracy or Computational Efficiency? A Trade-Off

Improving efficiency often requires a time investment. Thus, a computational complexity experiment is necessary to compare the performance of machine learning models with traditional econometric methods, focusing on the trade-off between execution time and gains in accuracy. Specifically, we measure the time required for each model to run and assess its classification accuracy. By doing so, we can evaluate the balance between computational efficiency and predictive performance.

To implement this experiment, we first measure the computational complexity of our models using the timer command to capture execution times. We then compare these results with the accuracy rates (as presented in the previous section), providing insights into the trade-offs between accuracy and computation time.

Table 7 presents the execution times for each model applied to our data. When compared to the accuracy rates shown in Table 4 for the overall sample, we observe a clear trade-off: improving model accuracy comes at the cost of increased computational time. Specifically, the higher accuracy achieved by the XGBoost model requires significantly more time to run, highlighting the balance between precision (i.e., correct firm classification) and computational demands.

Table 7. Accuracy evaluation rates achieved with prediction model—Overall sample.

	Backward	Forward	Lasso	CART (Model 1)	CART (Model 2)	XGBoost
Accuracy	74.79%	74.96%	76.76%	79.22%	80.70%	82.31%
Execution time	2.87 s	3.30 s	4.22 s	5.24 s	30.06 s	45.23 s

4. Discussion and Conclusions

The current literature on bankruptcy prediction has predominantly concentrated on models designed for larger firms, with limited attention given to those tailored for smaller, country-specific contexts (Crutzen and Van Caillie 2007, 2010). In recent years, only a few studies have laid the groundwork for investigating bankruptcy prediction specifically for SMEs (Altman and Sabato 2007; Ciampi and Gordini 2008, 2009; Yazdanfar 2011; Cultrera and Brédart 2016; Zoricák et al. 2020; Papik and Papíková 2023). In Belgium, the growing number of SMEs, along with the country's unique political landscape, regional institutional differences, and the uncertainties in the current economic environment, highlight the need to conduct more in-depth analyses focused on bankruptcy prediction for Belgian SMEs. That is, the Belgian context makes bankruptcy laws, particularly for SMEs, differ significantly from those in other studied countries. Variations in legal frameworks, creditor protection, cultural perceptions, and procedural efficiency can influence both the likelihood of firms declaring bankruptcy and the outcomes of those cases. SMEs in Belgium often face higher regulatory challenges, slower bankruptcy proceedings, and fewer opportunities for restructuring compared to more debtor-friendly systems like those in the U.S., UK, or other European nations with specialized SME insolvency processes.

However, most studies trying to predict corporate bankruptcy mainly focus on robust and sometimes innovative prediction methods to enhance the models' performance, without sufficiently considering other model elements, such as an upwind variables selection. Therefore, the added value of this paper has wide-ranging implications for managers in the improvement of bankruptcy prediction models.

As we show, the variable selection method can significantly influence the performance of the model. This paper demonstrates that intelligence techniques, such as machine learning models, outperform more classical statistical approaches in terms of predictive accuracy. More precisely, decision trees like CART appear to be highly efficient and are not employed as much as they deserve in the context of bankruptcy prediction. However, one must carefully consider the trade-off between the time required to implement complex computations and the resulting accuracy. While more sophisticated models can yield higher predictive accuracy, they often come at the cost of longer processing times, which may not always justify the incremental gains in performance.

Also, we show that specific variables related to the position of the firm within the production process may disrupt debates. That is, variables selection processes presented in this research are found to be less extensively efficient regarding financial-oriented variables for bankruptcy prediction of firms located upstream in the production process, such as it is for their downstream counterparts. Therefore, for such upstream firms, there seems to exist some other influences, i.e., some *noise*, in the forecasting strategy. This provides support to other variables affecting bankruptcy for these firms such as contagion effects that could be more pronounced in upstream industrial sectors (Fujiwara 2008; Delli Gatti et al. 2009; Hua et al. 2011). At the end of the day, we hope this research will lead other researchers to pay more attention to this element in their efforts to develop more accurate models as well.

Besides the contributions of this study to the academic literature on variables selection strategies and bankruptcy prediction models, investors and managers may rely on such a new mode of information technology deployment in their evaluation and anticipation of their firms' bankruptcy risks. This research may also offer inputs to financial institutions in their evaluation of the risk profiles of SMEs, since such variables selection models may help them to better identify firms with a higher risk of failure in their lending decisions.

This paper suffers from some limitations, including its focus on Belgium, which limits the applicability of the findings to other countries, and its emphasis on SMEs, making it less relevant to larger firms. Additionally, the model lacks integration of important factors like macroeconomic variables and a detailed analysis of contagion effects among firms.

Finally, future research could explore cross-country comparisons to understand how bankruptcy laws differ internationally, potentially offering more nuanced insights into SME financial distress. Integrating external economic factors could further enhance the accuracy of prediction models. Additionally, expanding the scope to include larger firms or developing sector-specific models would increase the generalizability of the findings. Longitudinal studies could also refine these predictive models over time, improving their robustness in forecasting SME bankruptcy risks in various economic contexts.

Author Contributions: All authors contributed equally to the conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, original draft preparation, review and editing, visualization, supervision, and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The research data for this article was sourced from the Bureau Van Dijk Belfirst database.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Descriptive statistics show that the firms in the sample are relatively young, averaging about 14 years of age, suggesting a mix of both established and newer firms. Then, the mean number of employees (8) means that firms are small to medium-sized enterprises (SMEs). This aligns with the study's focus on SMEs, which tend to have more limited resources and distinct challenges compared to larger firms. Also, the equity-to-asset ratio shows low shareholder equity relative to total assets, reflecting either a reliance on debt

financing or modest capital bases. This could be a sign of vulnerability, as SMEs with lower equity reserves may struggle to absorb financial shocks. Then, with an average negative EBIT-to-total-assets ratio, firms in this data set are not generating positive operating profits relative to their asset base. This could signal financial distress or weak operational efficiency. Then, the current ratio is almost two, indicating that firms generally have enough current assets to cover their short-term liabilities. This suggests relatively good liquidity, although some firms may face challenges in turning these assets into cash quickly. Also, a negative ratio of accruals to total assets points to the fact that many firms are operating under accrual accounting with a conservative approach to revenue recognition, or that they are recognizing more liabilities than assets through accruals. Next variable, the relatively low value added per worker could imply that the labor productivity of the firms in the sample is modest, which could be related to the smaller size and scope of SME operations. Then, the acid test, which measures liquidity by excluding inventory from current assets, suggests that these firms have strong liquidity, surpassing the usual benchmark of 1.0. This indicates a solid ability to cover short-term liabilities without relying on inventory sales. The average log of sales implies variability in the size of firms, but it aligns with expectations for SMEs, suggesting moderate levels of revenue generation. Then, the number of days firms are extending credit to clients (75 days) is almost the same as the number of days they take to pay suppliers (74 days), which points to a balance in managing receivables and payables. However, these are relatively long credit periods, which could create liquidity pressure if collection is delayed. Then, statistics show that firms are financing a significant portion of their assets with long-term debt (58%), which is high but potentially manageable for growth, depending on their ability to service this debt. Gross sales margin (16.45%) and net sales margin (7.97%) indicate that while firms are making a reasonable gross profit, the net margins are narrower, suggesting that operating and other expenses are absorbing a significant portion of revenue. Concerning net income/sales and net income/total assets, statistics suggests that overall, firms are struggling to generate sufficient income relative to their asset base. Firms are also generating low cash flow relative to both equity and debt, indicating limited ability to cover liabilities with operational cash flow. Next, while there is significant working capital relative to sales, the negative ratio of working capital to total assets suggests that many firms are operating with tight cash flows and short-term financial strains. Also, the high level of long-term debt compared to equity (nearly three times) suggests that firms are heavily leveraged, which could amplify financial risk, particularly if they face a downturn. Then, descriptive statistics show that firms are financing 85% of their assets with debt, a very high level of leverage that could increase their vulnerability in periods of financial stress. Then, while firms have modest earnings before interest, taxes, depreciation, and amortization relative to their total debt, the ratio EBITDA/total debts suggests that firms may still be able to service their debts under normal conditions. Also, the positive but low net income/current debts and net income/total debts ratios indicate that firms generate enough net income to cover a portion of both their short-term and total debt obligations, but with only modest coverage. The relatively high current debt compared to sales also indicates that firms may face liquidity challenges, as current debts represent a significant portion of their revenue. Concerning tax expense, the ratio indicates that, on average, tax expenses account for a significant portion of the firms' total assets. A ratio this high suggests that taxes are a substantial financial burden for these SMEs, which could impact their overall profitability and financial flexibility. Finally, the regional and sectorial distributions show that the majority of the firms are based in Flanders and that the construction sector and intellectual and administrative activities dominate the data.

Table A1. Descriptive Statistics of Selected Variables.

Variable	Mean	Std. Dev.
Firm age	13.91	12.25
Number of employees	8.14	14.32
Shareholder equity/Total assets (€)	1825.85	10,497.80
EBIT/Total assets (k€)	−0.03	6.30
Current ratio	1.94	5.24
Accruals/Total assets (k€)	−0.30	0.63
Value added/Total workers (k€)	92.86	81.18
Acid test	2.89	8.02
Log of sales	7.88	2.29
Number of days of client credit	74.97	102.40
Number of days of supplier credit	74.26	110.02
Long term debts/Total assets	0.58	5.61
Gross sales margin (%)	16.45	27.61
Net sales margin (%)	7.97	24.37
Net income/Sales (k€)	6.75	69.87
Net income/Total assets (k€)	−0.15	6.51
Cash-flow/Equity (k€)	0.03	0.41
Cash-flow/Total debts (k€)	0.29	3.96
Net Working Capital/Sales (€)	159.73	1643.89
Net Working Capital/Total assets (€)	−1.78	47.42
Long term debts/Equity (k€)	2.92	101.63
Total debts/Total assets (k€)	0.85	0.39
EBITDA/Total debts (k€)	1.34	42.79
Cash-flow/Current assets (k€)	0.34	16.26
Net income/Current debts (k€)	0.29	4.58
Net income/Total debts (k€)	1.04	32.87
Net income/Current assets (k€)	0.29	13.49
Current debts/Sales (k€)	6.31	49.16
Tax expense/Total assets (k€)	11.68	31.30
Region		
Wallonia (%)	23.56	42.44
Flanders	14.36	35.07
Brussels	62.08	37.65
Industrial Sector (NACE)		
Agriculture and industries	17.84	32.57
Energy and water production	0.53	5.92
Construction	28.76	39.57
Food service industry	11.58	26.86
Intellectual and administrative activities	27.94	39.14
Activities primarily in the public and social sector	8.34	23.08
Others	5.01	18.10

Notes

- ¹ The exclusion of firms with missing or inappropriate variables may lead to a potential selection bias. While this is a common issue in empirical studies, further robustness check has been conducted to assess the potential impact of this exclusion. That is, instead of excluding firms with missing variables, data imputation techniques through mean imputation have been used to estimate and fill in missing values. Applied to our sample, the imputed data set yields similar results, which are available on request. This provides confidence that missing data did not significantly affect the results.
- ² Descriptive statistics of the main variables (Table A1), together with a discussion, are presented in Appendix A.
- ³ Note that in the case where $y_i = 1$, we obtain P_i , and if $y_i = 0$, we obtain $1 - P_i$.
- ⁴ Note that we rely on a logit model with controls for unobserved fixed characteristics of firms. That is, we add as controls: (i) firm size, since larger firms may have more financial stability compared to smaller firms, affecting their bankruptcy risk; (ii) industry sector, which can significantly impact the risk of bankruptcy due to differing market risks, regulations, and competition levels and which is usually fixed over a medium term; (iii) and the geographical location, since firms located in economically stronger regions may have access to better resources and markets, affecting their financial resilience.

⁵ We thank an anonymous referee for these insightful suggestions.

⁶ The mix of variables extracted from Table 1 in each component is the following: Component 1 relies on Profitability, Solvency, CF/E, and CF/TD. Component 2 relies on Firm age, Solvency, CF/TD, and TD/TA. Component 3 relies on Current, VA/TW, CF/TD, and EBITDA/TD. Component 4 relies on Profitability and CF/TD. Component 5 relies on Firm age, Solvency, and TD/TA.

References

- Agarwal, Vineet, and Richard Taffler. 2008. Comparing the Performance of Market-based and Accounting based Bankruptcy Prediction Models. *Journal of Banking and Finance* 32: 1541–51. [CrossRef]
- Altman, Edward. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609. [CrossRef]
- Altman, Edward. 1984. The success of business failure prediction models: An international survey. *Journal of Banking and Finance* 8: 171–98. [CrossRef]
- Altman, Edward, and Ben Branch. 2015. The Bankruptcy System's Chapter 22 Recidivism Problem: How Serious is It. *The Financial Review* 50: 1–26. [CrossRef]
- Altman, Edward, and Gabriele Sabato. 2007. Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus* 43: 332–57. [CrossRef]
- Amendola, Alessandra, Marialuisa Restaino, and Luca Sensini. 2011. Variable Selection in Default Risk Models. *The Journal of Risk Model Validation* 5: 3–19. [CrossRef]
- Argenti, John. 1976. *Corporate Collapse: The Causes and Symptoms*. London: Holsted Press McGraw-Hill.
- Balcaen, Sophie, and Hubert Ooghe. 2006. 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *The British Accounting Review* 38: 63–93. [CrossRef]
- Bauweraerts, Jonathan. 2016. Predicting Bankruptcy in Private Firms: Towards a Stepwise Regression Procedure. *International Journal of Financial Research* 7: 147–53. [CrossRef]
- Beaver, William H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71–111. [CrossRef]
- Brezigar-Masten, Arjana, and Igor Masten. 2012. CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications* 39: 10153–59. [CrossRef]
- Carmona, Pedro, Francisco Climent, and Alexandre Momparler. 2019. Predicting failure in the U. S. banking sector: An extreme gradient boosting approach. *International Review of Economics and Finance* 61: 304–23. [CrossRef]
- Charalambous, Chris, Spiros H. Martzoukos, and Zenon Taoushianis. 2021. Estimating corporate bankruptcy forecasting models by maximizing discriminatory power. *Review of Quantitative Finance and Accounting* 58: 297–328. [CrossRef]
- Chava, Sudheer, and Robert A. Jarrow. 2004. Bankruptcy Prediction with Industry Effects. *Review of Finance* 8: 537–69. [CrossRef]
- Chen, Jiahua, and Zehua Chen. 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95: 759–71. [CrossRef]
- Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Paper presented at the International Conference on Knowledge Discovery and Data Mining Proceedings, San Francisco, CA, USA, August 13–17.
- Cho, Sungbin, Hyojung Hong, and Byoung-Chun Ha. 2010. A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Application* 37: 3482–3488. [CrossRef]
- Choubin, Bahram, Hamid Darabi, Omid Rahmati, Farzaneh Sajedi-Hosseini, and Bjørn Kløve. 2018. River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Science of The Total Environment* 615: 272–81. [CrossRef]
- Ciampi, Francesco, and Niccolo Gordini. 2008. *Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: An Empirical Analysis*. Oxford: Oxford Business and Economics Conference Proceedings.
- Ciampi, Francesco, and Niccolo Gordini. 2009. Default Prediction Modeling for Small Enterprises: Evidence from Small Manufacturing Firms in Northern and Central Italy. *Oxford Journal* 8: 13–32.
- Crutzen, Nathalie, and Didier Van Caillie. 2007. The Prevention of Business Failure: A State of the Art. HEC Working Paper. Available online: <https://orbi.uliege.be/bitstream/2268/30309/1/NC%20-%20Accounting%20history%20-%20Oct%2007.pdf> (accessed on 11 April 2024).
- Crutzen, Nathalie, and Didier Van Caillie. 2010. Towards a Taxonomy of Explanatory Failure Patterns for Small Firms: A Quantitative Research. *Review of Business and Economics* 55: 438–63.
- Cultrera, Loredana, and Guillaume Vermeylen. 2018. Prédire la faillite des PME belges? Une modélisation statistique. In *Risques Financiers. Mesures et conséquences*. Edited by Sebastien Galanti. Rennes: Presses Universitaires de Rennes.
- Cultrera, Loredana, and Xavier Brédart. 2016. Bankruptcy prediction: The case of Belgian SMEs. *Review of Accounting and Finance* 15: 101–19. [CrossRef]
- Daubie, Mickaël, and Nadine Meskens. 2001. Bankruptcy Prediction: Literature Survey of the Last Ten Years. *Belgian Journal of Operations Research, Statistics and Computer Science* 41: 43–58.
- Delli Gatti, Domenico, Mauro Gallegati, Bruce C. Greenwald, Alberto Russo, and Joseph E. Stiglitz. 2009. Business fluctuations and bankruptcy avalanches in an evolving network economy. *Journal of Economic Interaction and Coordination* 4: 195–212. [CrossRef]

- Dimitras, Augustinos I., Stelios H. Zanakis, and Constantin Zopounidis. 1996. A survey of business failures with an emphasis on failure prediction methods and industrial applications. *European Journal of Operational Research* 90: 487–513. [CrossRef]
- Dimitras, Augustinos I., Roman Slowinski, Robert Susmaga, and Constantin Zopounidis. 1999. Business failure prediction using rough sets. *European Journal of Operational Research* 114: 263–80. [CrossRef]
- du Jardin, Philippe. 2016. A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research* 254: 236–52. [CrossRef]
- du Jardin, Philippe, David Veganzones, and Eric Séverin. 2019. Forecasting corporate bankruptcy using accrual-based models. *Computational Economics* 54: 7–43. [CrossRef]
- Fonti, Valeria, and Eduard Belitser. 2017. Feature Selection Using Lasso. VU Amsterdam Research Paper in Business Analytics. Available online: <https://vu-business-analytics.github.io/internship-office/papers/paper-fonti.pdf> (accessed on 5 May 2024).
- Fujiwara, Yoshi. 2008. Chain of Firms' bankruptcy: A macroscopic study of link effect in a production network. *Advances in Complex Systems* 11: 703–17. [CrossRef]
- Gower, John C., and Garmt B. Dijksterhuis. 2004. *Procrustes Problems*. Oxford: Oxford University Press.
- Guilhot, Bernard. 2000. Défaillance d'entreprise: Soixante-dix ans d'analyses théoriques et empiriques. *Revue Française de Gestion* 26: 52–67.
- Hua, Zhongsheng, Yanhong Sun, and Xiaoyan Xu. 2011. Operational causes of bankruptcy propagation in supply chain. *Decision Support Systems* 51: 671–81. [CrossRef]
- Jackson, Richard H. G., and Anthony Wood. 2013. The Performance of Insolvency Prediction and Credit Risk Models in the U.K: A Comparative Study. *The British Accounting Review* 45: 183–202. [CrossRef]
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Berlin: Springer.
- Julien, Pierre André. 1997. *Les PME: Bilan et Perspectives*. Paris: Economica Publishing.
- Kassambara, Alboukadel. 2018. *Machine Learning Essentials—Practical Guide in R*. Scotts Valley: CreateSpace Independent.
- Keasey, Kevin, and Robert Watson. 1987. Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypotheses. *Journal of Business Finance & Accounting* 14: 335–54. [CrossRef]
- Kim, Hyun-Jung, Nam-Ok Jo, and Kyung-Shik Shin. 2016. Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications* 59: 226–34. [CrossRef]
- Kitowski, Jerzy, Anna Kowal-Pawul, and Wojciech Lichota. 2022. Identifying Symptoms of Bankruptcy Risk Based on Bankruptcy Prediction Models—A Case Study of Poland. *Sustainability* 14: 1416. [CrossRef]
- Li, Hui, Jie Sun, and Jian Wu. 2010. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications* 37: 5895–904. [CrossRef]
- Mai, Feng, Shaonan Tian, Chihoon Lee, and Ling Ma. 2019. Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research* 274: 743–58. [CrossRef]
- Miller, Alan. 2002. *Subset Selectin in Regression*. London: Chapman & Hall/CRC.
- Morris, Richard. 1997. *Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence*. Farnham: Ashgate Publishing Limited.
- O'Brien, Robert M. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity* 41: 673–90. [CrossRef]
- Ogachi, Daniel, Richard Ndege, Peter Gaturu, and Zeman Zoltan. 2020. Corporate Bankruptcy Prediction Model, a Special Focus on Listed Companies in Kenya. *Journal of Risk and Financial Management* 13: 47. [CrossRef]
- Ohlson, James A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–31. [CrossRef]
- Ooghe, Hubert, and Sofie De Prijcker. 2008. Failure processes and causes of company bankruptcy: A typology. *Management Decision* 46: 223–42. [CrossRef]
- Papik, Mário, and Lenka Papíková. 2023. Impacts of crisis on SME bankruptcy prediction models' performance. *Expert Systems with Applications* 214: 119072. [CrossRef]
- Park, Trevor, and George Casella. 2008. The Bayesian Lasso. *Journal of the American Statistical Association* 103: 681–86. [CrossRef]
- Peacock, Rolffe W. 2004. *Understanding Small Business: Practice, Theory and Research*. Adelaide: Scarman Publishing.
- Peel, Michael J., and David A. Peel. 1987. Some Further Empirical Evidence on Predicting Private Company Failure. *Accounting and Business Research* 18: 57–66. [CrossRef]
- Premachandra, Inguruwatt M., Gurmeet Singh Bhabra, and Toshiyuki Sueyoshi. 2009. DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research* 193: 412–24. [CrossRef]
- Refait, Catherine. 2000. Estimation du risque de défaut par une modélisation stochastique du bilan: Application à des firmes industrielles françaises. *Finance* 21: 103–29.
- Schapiro, Robert E. 1999. A brief introduction to boosting. *Ijcai* 99: 1401–406.
- Shetty, Shekar, Mohamed Musa, and Xavier Brédart. 2022. Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management* 15: 35. [CrossRef]
- Shumway, Tyler. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74: 101–24. [CrossRef]
- Singh, Smita, Patricia Doyle Corner, and Kathryn Pavlovich. 2015. Failed, not finished: A narrative approach to understanding venture failure stigmatization. *Journal of Business Venturing* 30: 150–66. [CrossRef]

- Son, Hwijae, Chongseok Hyun, Du Phan, and Hyung Ju Hwang. 2019. Data analytic approach for bankruptcy prediction. *Expert Systems with Applications* 138: 112816. [[CrossRef](#)]
- Storey, David J., Kevin Keasey, Robert Watson, and Pooran Wynarczyk. 1987. *The Performance of Small Firms: Profits, Jobs and Failures*. London: Routledge.
- Tibshirani, Robert. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B* 58: 267–88. [[CrossRef](#)]
- Tsai, Chih-Fong. 2009. Feature selection in bankruptcy prediction. *Knowledge-Based Systems* 22: 120–27. [[CrossRef](#)]
- Van Caillie, Didier. 2000. La détection des signaux financiers annonciateurs de faillite en contexte PME: Une approche méthodologique spécifique. Paper presented at 5th International French Speaking Conference on SME, Lille, France, October 25–27.
- Veganzones, David, and Eric Séverin. 2018. An investigation of bankruptcy prediction in imbalances datasets. *Decision Support Systems* 112: 111–24. [[CrossRef](#)]
- Wang, Gang, Jian Ma, and Shanlin Yang. 2014. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications* 41: 2353–61. [[CrossRef](#)]
- Weitzel, William, and Ellen Jonsson. 1989. Decline in Organizations: A literature Integration and Extension. *Administrative Sciences Quarterly* 34: 91–109. [[CrossRef](#)]
- Xu, Ming, and Chu Zhang. 2009. Bankruptcy prediction: The case of Japanese listed companies. *Review of Accounting Studies* 14: 534–58. [[CrossRef](#)]
- Yazdanfar, Darush. 2011. Predicting bankruptcy among SMEs, Evidence from Swedish firm-level data. *International Journal of Entrepreneurship and Small Business* 14: 551–64. [[CrossRef](#)]
- Zmijewski, Mark E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 22: 59–82. [[CrossRef](#)]
- Zoricák, Martin, Peter Gnip, Peter Drotar, and Vladimír Gazda. 2020. Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets. *Economic Modelling* 84: 165–76. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.