# A machine learning-based analysis on the causality of financial stress in banking institutions

**João Gabriel de Moraes Souza · Daniel Tavares de Castro · Yaohao Peng · Ivan Ricardo Gartner**

**Abstract** In this paper, we applied machine learning techniques to analyze the default probability in financial institutions using a large dataset of variables collected from 2,325 banks over 17 years, extracting the most relevant variables using a feature selection method (Lasso), predicting default and systemic risk with random forest and XGBoost algorithms, and finally investigating the contributions of each relevant feature to the overall financial stress of banking institutions using Explainable Artificial Intelligence (XAI) techniques. According to this methodology, we found that the most important variables for the default risk predictions are the probability of a bailout calculated, the market share in terms of assets, the market-to-book ratio, total liabilities, and the number of banks in the market (a measure of concentration and competition). For the systemic risk predictions, the most important variables are the number of banks in the country, the level of interest rates, the market share of the top 5 largest banks, and the region of the bank (in North America, Europe, and Central Asia). The findings of this research provide an empirical assessment of the main factors that explain the presence of financial stress in banking institutions, conciliating the versatility of machine learning models with practical interpretability and causal inference, being of potential interest to researchers in quantitative finance and market practitioners.

**Keywords** Big Data · Quantitative Finance · Explainable Artificial Intelligence · Ensemble Methods · Supervised Machine Learning · Banking

## 1 Introduction

The investigation of bankruptcy and financial distress in financial institutions has been one of the most prominent research topics of quantitative corporate finance (Gruszczyński, 2020). As discussed in Carvalho et al (2015), financial stresses in financial institutions are often associated with posterior equity valuation losses and investment cuts to borrower firms and can be transmitted onto non-financial firms, with the losses being concentrated in firms with the greatest information asymmetry problems and weakest financial positions. Likewise, the impacts of financial distress can be propagated within the financial system, as banks can make more efficient liquidations if they are regionally active and have closer relationships with the firm, while banks with higher exposure to firms that undergo financial distress have larger negative announcement-period returns, which results in wealth losses for the bank shareholders and also induces a negative effect on the borrower's returns (Dahiya et al, 2003; Höwer, 2016).

Given the variety of variables that may potentially explain the causes of financial distress in banks, it becomes relevant to provide measures of their respective contributions to this phenomenon. These aspects are particularly pertinent with the emergence and popularization of data science methods in both academic works and business solutions, thus allowing different settings of training samples, predictors, and hyperparameters for the same empirical analysis, all elements which may decisively affect the final outcome of the models. As pointed out in Hassani et al (2018), the financial system has been substantially influenced by data mining and big data techniques in terms of decision-making support and strategic management. On the other hand, a comprehensive understanding of the causes of financial stress and their implications may require additional caution to not only identify the important factors that maximize the model's overall explainability but also be able to reveal underlying patterns from the data, such as the variables' importance to individual observations and their joint behavior in terms of partial dependence. Iwanicz-Drozdowska and Ptak-Chmielewska (2019), for instance, analyzed an unbalanced panel of European banks with more than three thousand banks over 25 years applying regional grouping and investigating the relationships between macroeconomic variables and the occurrence of distress events, evidencing the dynamic role of the macroeconomic outlook and its systemic impacts on the banking sector.

In this sense, this paper employs state-of-the-art machine learning techniques to analyze the probability of default in financial institutions using a large dataset of variables collected from 2,325 banks over 17 years, tackling not only the predictive aspects but also the explanatory aspects of the phenomenon, by evaluating the most relevant variables using feature selection methods and investigating the contributions of each relevant feature to the overall financial stress of banking institutions using Explainable Artificial Intelligence (XAI) for enhanced interpretability of the models' results at a practical level, addressing gaps from the literature of quantitative banking finance in a data-driven approach.

This paper is structured as follows, Section 2 addresses the theoretical background of modeling bank risk, Section 3 presents the data and methods covered, Section 4 presents the main results of the research and its analysis, and finally presents the final considerations in Section 5.

## 2 Theoretical background

### 2.1 Default models

In this section, the calculation measures of default banking are presented. In particular, this section considers the Merton (1974) probability of default method, also known as the Merton model, the default model KMV from Moody's, and the Z-score model of Lown et al (2000) and of Tabak et al (2013), which is an adaptation of the Altman (1968) model.

#### *2.1.1 Merton Model*

The Merton (1974) model aims to find the values of assets and their volatilities in a dynamic process following Black and Scholes (1973). In the Merton model, it is assumed that the total value of the firm follows a geometric Brownian motion process.

$$dV = \mu V dt + \sigma_V V dW \tag{1}$$

where $V$ is the total value of the firm's assets (random variable), $\mu$ is the expected continuous return of $V$, $\sigma_V$ is the firm value volatility, and $dW$ is the standard process of Gauss–Wiener.

The Merton model uses the Black and Scholes (1973) model of options in which the firm's equity value follows the stipulated process of Black and Scholes (1973) for call options. A call option on the underlying assets has the same properties as a caller has, namely, a demand on the assets after reaching the strike price of the option. In this case, the exercise price of the option equals the book value of the firm's obligations. If the value of the assets is insufficient to cover the firm's obligations, then shareholders with a call option do not exercise their option and leave the firm to their creditors.

$$E = V\mathcal{N}(d_1) - e^{-rT}F\mathcal{N}(d_2) \tag{2}$$

where $E$ is the market value of the equity of the firm (or free cash flow to the shareholder), $F$ is the face value of the debt securities, $r$ is the risk-free interest rate, and $\mathcal{N}(.)$ is the standardized cumulative normal distribution; $d_1$ is given by

$$d_1 = \frac{ln(\frac{V}{F}) + (r + 0.5\sigma_V^2)T}{\sigma_V\sqrt{T}} \tag{3}$$

and $d_2$ is simply $d_1 - \sigma_V\sqrt{T}$.

Applying the Itô Lemma in the dynamic process of $V$ and manipulating the terms of Equation 3, we obtain the following equation of the variability of the free cash flows of the shareholders ($\sigma_E$) (Bharath and Shumway, 2008).

$$\sigma_E = \left(\frac{V}{E}\right)\frac{\partial E}{\partial V}\sigma_V \tag{4}$$

Given that Merton (1974), it can be shown that $\frac{\partial E}{\partial V} = \mathcal{N}(d_1)$; then, Equation 4 can be written as follows Bharath and Shumway (2008):

$$\sigma_E = \left(\frac{V}{E}\right)\mathcal{N}(d_1)\sigma_V \tag{5}$$

In essence, the algorithm works with Equations 2 and 5 to find the value terms of the asset $V$ and the volatility of the asset value $\sigma_V$. In this study, we use the Newton method to solve Equations 2 and 5, the same algorithm that was used by Anginer and Demirguc-Kunt (2014).

Equations 2 and 5 have numerical solutions only for the values of $V$ and $\sigma_V$. Once the numerical solution is found, the distance of default is calculated as follows:

$$DD = \frac{ln\left(\frac{V}{F}\right) + (\mu - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}} \tag{6}$$

According to Bharath and Shumway (2008), the distance to the default model of Merton (1974) accurately measures the probability of firms defaulting.

$$\pi_{Merton} = \mathcal{N}(-DD) \tag{7}$$

The default probability measure of Merton (1974) is simply the probability function of the normal minus the distance to default, Equation 6. According to Bharath and Shumway (2008), this probability of default (Equation 7) should be a sufficient statistic for the default prognostic.

The starting point of the algorithm follows an adaptation of Bharath and Shumway (2008) and Anginer and Demirguc-Kunt (2014), where the initial kicks of $V$ assume $V = market\ capitalization + total\ liabilities$ and $\sigma_V$ assume $\sigma_V = \sigma_{asset\ price\ return} \cdot (market\ capitalization + total\ liabilities)$.

### *2.1.2 KMV Model*

The KMV model is calculated from the total value of the firm's assets $V$ and the volatility of the asset value $\sigma_V$ from the iteration between Equations 2 and 5. We can observe the same process in Merton (1974) Model.

$$DD_{KMVit} = \frac{(V_{it} - TL_{it})}{(V_{it} \cdot \sigma_{Vit})} \tag{8}$$

where $V_{it}$ represents the market value of asset $i$ in period $t$, $\sigma_{Vit}$ represents the volatility of asset value $i$ in period $t$, and $TL_{it}$ is total liabilities to asset $i$ in period $t$. As indicated by Equation 8, the higher $DD_{kmvit}$, the higher the distance to default from bank $i$ in period $t$.

To normalize the variable to have a parallel effect and correlation analysis with the risk premium and risk exposure variables, normalization similar to that elaborated by Merton (1974) is used. In other words, the probability of default of the KMV model is given by equation 9:

$$\pi_{kmv} = \mathcal{N}(-DD_{KMV}) \tag{9}$$

The default KMV measure is the normal probability function minus the default distance, as well as the Merton (1974) model.

*2.1.3 Z-score Model*

Another way to measure the default risk is the Z-score indicator, similar to Lown et al (2000) and Tabak et al (2013). According to Lown et al (2000) and Tabak et al (2013), this indicator represents the probability of bank failure. The Z-score measure has the following formulation:

$$Z - score_{it} = \frac{ROA_{it} + EQUAS_{it}}{\sigma_{ROA_i}} \tag{10}$$

where $EQUAS = \left(\frac{E_{it}-E_{it-1}}{TA_{it}-TA_{it-1}}\right)$. In this model, $E_{it}$ represents bank equity $i$ in period $t$, $Eit-1$ represents bank equity $i$ in period $t-1$, $TA_{it}$ represents total assets of bank $i$ in period $t$, and $TA_{it-1}$ represents total assets of bank $i$ in period $t-1$.

Parameter $ROA_{it}$ is expressed as the following relation:

$$ROA_{it} = \frac{2\pi_{it}}{(TA_{it} - TA_{it-1})} \tag{11}$$

$ROA_{it}$ is the return on assets in period $t$ for bank $i$ and $\sigma_{ROAit}$ is the standard deviation of $ROA$ of bank $i$ in period $t$. As the formula indicates, the higher the Z-score value, the lower the probability of bank $i$ failure. For Tabak et al (2013), the Z-score is a risk default measure accepted by the literature. The Z-score measures the number of standard deviations of $ROA$ that must decrease for banks to become insolvent, which can be interpreted as the inverse of the probability of insolvency Tabak et al (2013).

Both the default probability measure of Merton (1974) and the KMV measure have a direct relationship with the default, that is, the higher their values, the greater the likelihood of a financial institution failing. Meanwhile, the Z-score model has an inverse relationship with the default bank, that is, the higher its values, the more distant the bank is from the default. This inverse relationship between the Merton, KMV, and Z-score models occurs because the first two represent a normalization of the distances from the default, making them directly linked to the default. In the Z-score model, on the other hand, the relationship is inverse because it measures the distance from the default, that is, the probability of banks being further from the default and not the direct probability of default.

The Merton (1974) model and the KMV model are default proxies that do not directly calculate the probability of default, but rather measure it implicitly, by looking at bank liabilities and how the market prices these liabilities (Milne, 2014). According to Wang et al (2017), to measure the default risk, a common proxy should not be used, but a flexible enough measure to quantify most firms in the market.

The Z-score model, despite being an accepted measure in the risk measurement literature, especially of bank risk, does not express the market relationship, but the banks' accounting relationship.

## 2.2 Systemic Risk Measures

In order to measure systemic stability, we apply an approach developed by Anginer et al (2014). As a measure of systemic risk, we use the $R^2$ obtained by regressing changes in bank default risk in relation to the average default risk for each of the 92 countries in the sample. The default models used were: the Merton (1974) Model, the KMV Model, and the Z-score model of Tabak et al (2013).

When calculating systemic risk, for each bank $i$ in the country $j$ in the year $t$, firstly we computed quarterly default probabilities. So, for each bank $i$ in the country $j$ time regressions are generated, observing the relationship between the bank variation $i$ and the country average $j$ in the period $t$, which in this case is annual.

$$\pi_{measure,\ i,j,t} = \alpha_{i,j,t} + \beta_{i,j,t}\frac{1}{n}\sum_{k=1,k\neq i}^{n} \pi_{measure,\ i,j,t} + \varepsilon_{i,j,t} \tag{12}$$

So, we compute the logistic transformation of the $R^2$ of the regression in Equation 12, which is equal to $log\left(\frac{R^2_{i,j,t}}{(1-R^2_{i,j,t})}\right)$. This measure expresses the systemic risk that the bank $i$ faces in country $j$.

Since the data were computed on an annual basis, in order to use them in the quarterly model, we applied cubic splines interpolation.

According to Anginer et al (2014), high $R^2$ values of the Equation 12 suggest that the bank $i$ is exposed to risks similar to the risks of other banks in the country $j$. That is, these figures suggest that there is a channel of the interdependence of banks in each country $j$. This common risk mechanism of banks in each country $j$ makes the banking sector more exposed to the individual risks of each bank $i$.

It is worth adding that, according to Acharya (2009), banks that have incentives to connect their risks through guarantees from the State - even if implicit - end up prolonging and affecting systemic risk.

## 2.3 Competition Measure, Bank Competition, and Bailout Probability Perception

In this section, one measure of competition is developed: the Lerner index, of Amidu and Wolfe (2013); Fu et al (2014); Anginer et al (2014). Three concentration averages are also addressed: the $HHI$ index, the $Market\text{–}Share$ index, and the $CR5$ index by Akins et al (2016). Finally, the perception of the probability of a bailout by Gropp et al (2011) is developed.

### *2.3.1 Competition Measure*

To measure the Lerner Index, it is necessary to find the Marginal Costs functions. Since the marginal cost estimates are not straightforward, this study uses the translog cost functions of each bank $i$ in the period $t$ (quarter) for each country $j$, an approach similar to that used by Tabak et al (2012).

The translog cost function is given by:

$$log\left(\frac{C}{w_2}\right)_{i,t} = \delta_0 + \sum_p logy_{p,i,t} + \frac{1}{2}\sum_p\sum_k \delta_{p,k} logy_{p,i,t} logy_{k,i,t}$$
$$+\beta_1 log\left(\frac{w_1}{w_2}\right)_{i,t} + \frac{1}{2}\beta_{1,1} log\left(\frac{w_1}{w_2}\right)_{i,t} log\left(\frac{w_1}{w_2}\right)_{i,t} + \sum_p \theta_p logy_{p,i,t} log\left(\frac{w_1}{w_2}\right)_{i,t}$$
$$+Dummies_t + \varepsilon_{i,t} \qquad (13)$$

Where $C$ represents the total costs of banks and $y$ represents three outputs: total loans; total deposits and non-interest income. This output is a measure that, according to Tabak et al (2012), represents a non-traditional activity of banks. These outputs, according to term $\frac{1}{2}\sum_p\sum_k \delta_{p,k} logy_{p,i,t} logy_{k,i,t}$, interact with each other in order to measure the translog costs. $w$ consists of 2 price inputs: financial expenses in relation to total deposits (the price of funds) and non-financial expenses on total assets (capital price).[1]

When estimating Equation 13 of cost translog, this study proceeded to estimate fixed-effect models corrected by the variance matrix of Newey and West (1987). Sargan and Hausman endogeny tests were also used in order to find possible endogenies in the estimation of Equation 13. However, the endogenous process was not found, which led to the estimation by fixed effects with correction of autocorrelation and heteroscedasticity via Newey and West (1987).

To obtain marginal costs ($MC$), the dependent variable of Equation 13 is derived in first order in relation to the output $y_{i,l,t}$ (loans):

$$MC_{i,l,t} = \left(\frac{C_{i,t}/w_2}{y_{i,l,t}}\right)\left(\delta_{j=l} + 2\delta_{l,l} logy_{i,l,t} + \sum_{k=1,\ldots,K;k\neq l} \delta_{l,k} logy_{i,k,t} + \theta_l log\left(\frac{w_1}{w_2}\right)\right) \qquad (14)$$

[1] Tabak et al (2012); Hasan and Marton (2003) use non-financial spending as *proxy* for personnel expenses, due to the restriction on obtaining this variable.

After calculating the $MC$ by Equation 14, the Lerner index and the Boone index were measured. The Lerner index of Amidu and Wolfe (2013); Fu et al (2014), and Anginer et al (2014) is given by:

$$Lerner_{i,t} = \frac{Price_{i,t} - MC_{i,l,t}}{Price_{i,t}} \tag{15}$$

Where the term $Price_{i,t}$ represents the price of the total asset in relation to total revenues (financial and non-financial income) of the bank $i$ in the period $t$ (quarter). The $MC_{i,l,t}$ represents the marginal cost given by Equation 14.

The Lerner index provides a direct relationship with the level of market power because it represents a marginal variation in prices in relation to the marginal cost Amidu and Wolfe (2013); Fu et al (2014); Anginer et al (2014).

### *2.3.2 Concentration Measures*

The first concentration measure used in this work is the Herfindahl Index ($HHI$), based on total bank deposits. This approach was used by Akins et al (2016). For each bank $i$ in quarter $t$, the total sum of deposits in the country $j$. Another measure used by Akins et al (2016), adopted in this study, is the ratio of the concentration of the five largest banks in the country $j$, in terms of deposits. For each country $j$ there is a $CR5$ measure per period $t$. The last measure utilized in this study is the market share of each bank $i$ in the period $t$ in the country $j$, in terms of deposits. These concentration measures seek to present the relationship between the most or the less concentrated with the measures of default and systemic stability.

### *2.3.3 Bailout Probability Perception Measures*

According to Gropp et al (2011), one of the most important and difficult measurements is that of the public guarantees that the State can give to banks. The objective of this section is, therefore, to use an approach similar to that of Gropp et al (2011) in order to construct a variable that measures the perception of the probability of a bailout in banks. This measure, referred to in the literature as the market share of insured competitor banks, is constructed as follows:

$$MSI_{-k,j} = \sum_{i \neq k}^{N_j} \pi_{i,j} \frac{a_{i,j}}{A_j} \tag{16}$$

Where $N_j$ represents the total number of banks in the country $j$, $a_{i,j}$ is the bank $i$ 's total assets in the country $j$, and $A_j$ is the sum of the bank's total assets in the country $j$ (($A_j = \sum_i^{N_j} a_{i,j}$). As Gropp et al (2011) note, if all banks have the probability to engage in a bailout as zero or one, then this variable will present only the market share in terms of bank $i$ assets in the country $j$. It is observed that the variable $MSI$ not only varies between countries but also between banks in each country.

With a simple transformation, the $MSI$ variable can be written as the product of the competitor's bailout probability averages in the total market share.

$$MSI_{-k,j} = \pi_{-k,j} \frac{A_{-k,j}}{A_j} \tag{17}$$

Wherein $\pi_{-k,j} = \sum_{i \neq k}^{N_j} \pi_{i,j} \frac{a_{i,j}}{A_j}$ is the average of the bailout probabilities weighted on market-shares and $A_{-k,j} = A_j - a_{k,j}$ is the total assets of competitors in the country $j$. In this sense, the bigger the average probability of bailout and total *market–share*, the greater the distortion of competition in the banking market (Gropp et al, 2011).

According to Gropp et al (2011), a major challenge is to estimate the probability of a bailout. In this context, the authors use the default probabilities calculated by the risk agencies. In this study, the default measures of Merton (1974), the KMV model, and the Z-score model by Tabak et al (2013) are adopted.

### 2.4 Nonlinearity and machine learning in finance

Machine learning models are characterized to be "inductive" in the sense that they are flexible to the data collected in the sample, yielding decision functions based on the patterns that the data show, instead of fixing functional forms (like a linear or quadratic one) or holding restrictive assumptions about the distribution of the data. Regarding these elements, studies like Burns and Moosa (2015) and Hsu et al (2016) argue in favor of the use of machine learning methods in traditional financial economics problems, given their better empirical predictive power and better capability to generalize nonlinear relationships, as well as providing a broader view for well-established research topics in the finance agenda beyond classic econometrics; indeed, the relationships between the specified variables and the effects they induce occur in a nonlinear way, which makes the use of linear models in these contexts generate potentially biased results that can decisively hinder the decision-making process (Mama, 2017).

Concerning the effects of high dimensionality in finance, Kozak et al (2020) tested a number of well-established asset pricing factor models – including the CAPM model and the five-factor model of Fama and French (2015) introducing nonlinear interactions between 50 anomaly characteristics and 80 financial ratios up to the third power, applying dimensionality reduction and regularization with $\ell_1$ and $\ell_2$ penalties to increase the model's sparsity. The results showed that a very small number of principal components were able to capture almost all of the out-of-sample explanatory power, resulting in a much more parsimonious and easy-to-interpret model; moreover, the introduction of additional regularized principal components does not hinder the model's sparsity but does not improve predictive performance either.

In fact, the flexibility of machine learning models allows the development of extended cases of many well-known models in the financial literature: for instance, Feng et al (2018) proposed a nonlinear feature extraction to map the most informative components that explain the patterns of financial assets over time, treating the sorting of securities as an activation function of a deep neural network. The authors showed that the well-known Fama-French models with three (Fama and French, 1992) and five factors (Fama and French, 2015) are particular cases of the proposed deep learning approach; those factors were compared the "deep factors", with the deep learning cases slightly outperforming the 3-factor case, but showing higher mean square error than the 5-factor case, while ordinary least square showed high levels of out-of-sample error.

Similarly, Gu et al (2020) compared machine learning methods like boosted regression trees, random forests, and artificial neural networks with linear models like ridge regression and LASSO to measure the risk premium of financial assets using data of nearly 30,000 financial assets from the New York Stock Exchange and NASDAQ using monthly data from 1957 to 2016. The results indicated that machine learning models yielded better fitness in comparison to traditional statistical methods and that all models converged to a similar set of relevant predictors composed mainly by variations in the assets' liquidity and volatility, arguing in favor of parsimonious models over complex ones.

### 2.5 Explainable AI: conciliating complexity and interpretability

As pointed out by Croxson et al (2019), interpretability is a key element for stakeholders to understand the process of model-driven decision-making and to evaluate its potential implications, especially when the outcome is obtained under some degree of automated analysis. While machine learning models have the potential to better approximate complex patterns, usually the interpretability of the models is harder as their complexity increases, as the estimated parameters may not have a straightforward real-world implication. For example, for the class of Kernel methods, like the Support Vector Machine (Cortes and Vapnik, 1995) and its extensions, the Kernel function is able to generalize polynomial interactions up to an arbitrarily high dimension, making the estimation of the model feasible, but also encapsulating all nonlinear interaction between the explanatory features inside an inner product operator, thus being unable to separate the marginal effect that a single interaction term has on the dependent variable.

In this sense, the explainable Artificial Intelligence (XAI) framework allows conciliating the advantages of machine learning methods and the possibility to interpret the impacts of each variable in the target variable ("global interpretation"), as well as in the predicted outcome of a specific observation ("local interpretation"). As Fisher et al (2019) pointed out, a metric of global feature importance can be obtained by a permutation-based approach, such that, for a relevant feature to model the target variable, the predictive performance of the learner is expected to suffer sharper drops when that feature is permuted; in this sense, the difference between the observed values for the loss function before and after the permutations on a feature can be regarded as a proxy for its overall importance.

For example, the paper of Gan et al (2020) applied deep learning methods for the pricing of arithmetic and geometric average options with gains in both empirical accuracy and computation speed, without taking assumptions present in traditional pricing models, which may be sometimes unrealistic. However, interpretability tends to be hard for deep learning models, since the estimated parameters for the deeper hidden layers represent abstract hierarchies that influence the target variable but are hard to be expressed in terms of the original predictors. Therefore, a question like "the expected impact on $y$ given a change in some preditor $x$" – which would be easy to answer in a linear model – tends to be non-trivial for most machine learning models. Based on the findings of Leo et al (2019)'s survey paper, conciliating efficient algorithms with intelligible interpretations at the practical level is one of the main challenges not only for banking risk management but for machine learning applications in finance in a broad sense.

This challenge is also evident in the application of Beutel et al (2019), which applied machine learning-based early warning systems for systemic banking crises and found out that the machine learning methods were outperformed by the simpler logistic regression when tested recursively for out-of-sample data. Besides evidencing the importance of parameter-tuning for machine learning models to avoid overfitting (Peng and Nagata, 2020), these results draw attention to the relevance of intuitive prediction outcomes and interpretability in machine learning models, since the logit model, albeit much more restrictive in terms of assumptions and generalization ability, has a very easy interpretation, making it still a widely used tool for real-world decision-making.

Bussmann et al (2021) proposed an XAI model for credit risk management, aggregating predictions of machine learning models in terms of similarity of their respective values for Shapley Additive Explanation (SHAP). The results suggest that the groupings revealed sets of characteristics for clients of different risk profiles, which in turn can be used to model the credit score of future individuals. Likewise, Xiaomao et al (2019) analyzed the application of SHAP values for the feature selection task in finance-related classification problems, performing an empirical exercise on the modeling of bankruptcy for Polish companies combined with random forest and gradient boosting.

## 3 Data and methods

Our data comprises quarterly observations of 2,325 banks from all five continents collected from 2000 to 2016, totaling 155,775 observations. As previously mentioned in sections 2.1 and 2.2, dependent variables refer to the probability of default of banks (or default risk) calculated according to the Z-score, Merton, and KMV models (variables z_score, merton and kmv, respectively), and to systemic risk measures calculated according to these three models (r2_z_score, r2_merton and r2_kmv, respectively), totaling six dependent variables.

In the default risk calculated with the Z-score model, 99% of the values are concentrated in the 0 to 300 range, with only less than 400 observations with negative values. With the Merton and KMV models, values are concentrated around 0 and 1, in approximately equal proportions. The left-hand side of figure 1 shows the behavior of the three default risk-dependent variables. In the systemic risk-dependent variables, most of the values are negative, but with a wider distribution. The right-hand side of figure 1 shows the behavior of these variables.

Independent variables include the variables described in section 2.3, generally employed to forecast default risk. These variables include accounting variables (interest and non-interest income, interest and non-interest expenses, long-term, cash and savings deposits, total liabilities and deposits, equity, among others), market-related variables (market capitalization of the bank, meaning historical stock returns of the bank and of the main index of the country, the standard deviation of these returns, market to book ration, historical t-bill return, etc.), performance variables (ROA, ROE, ROA and ROE to total assets, and others), competition and concentration-related variables (market-share

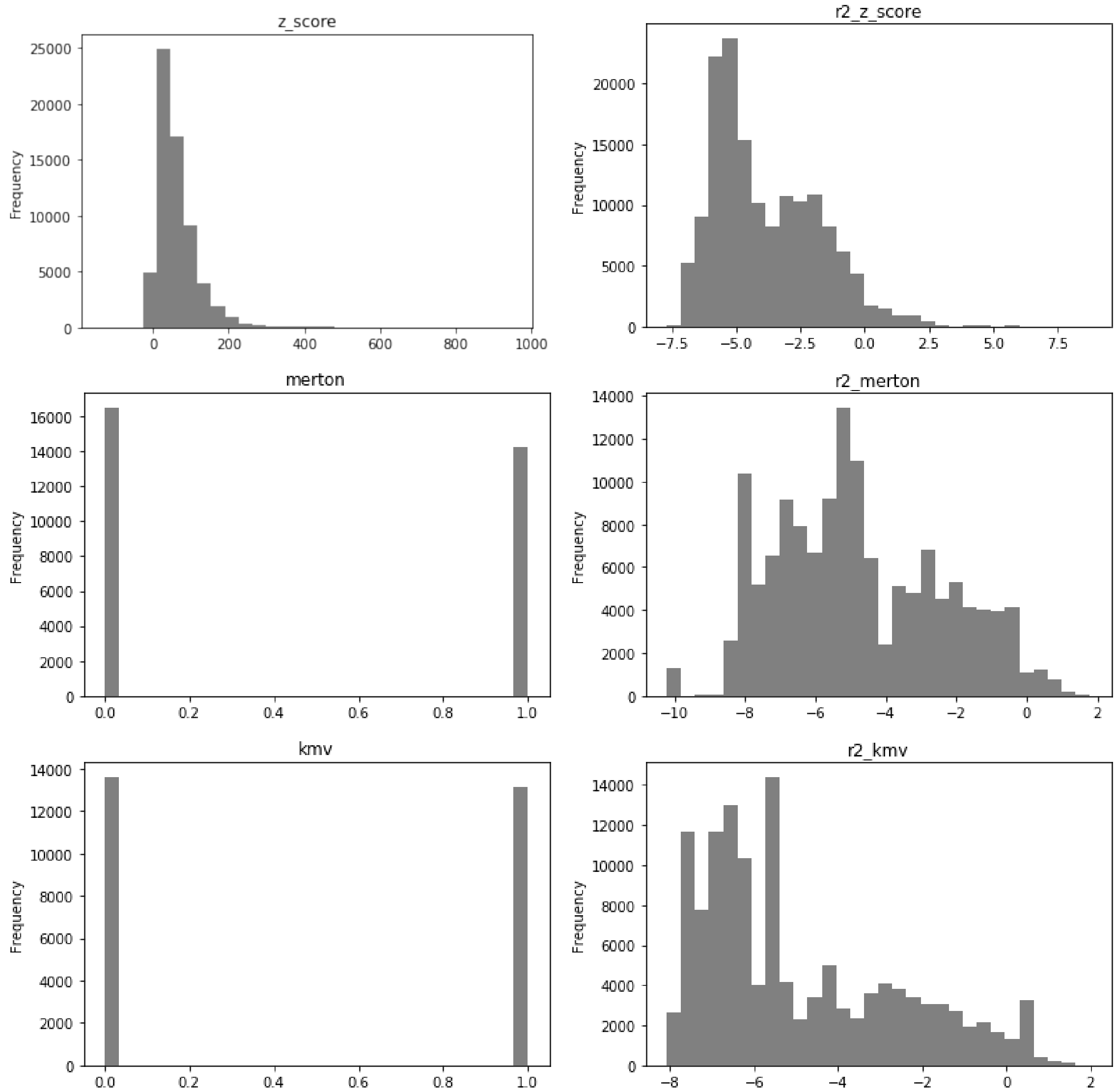


**Fig. 1** Distribution of values for default risk (left-hand side) and system and systemic risk (right-hand side) dependent variables. Source: authors' elaboration.

in relation to assets and loans, Lerner, HHI and CR5 indexes, etc.), risk related variables (risk exposure, bailout probabilities based on Merton and KMV measures, etc.) and macroeconomic and microeconomic variables (GDP growth and MC), besides the geographical region of the bank and the income group of the country where the bank is located. Table 1 presents descriptive statistics for both dependent and independent variables.

Table 1: Descriptive statistics of variables

| | count | mean | std. dev. | min. | max. |
|---|---|---|---|---|---|
| **Default risk dependent variables** | | | | | |
| z_score | 63,690 | 3.74 | 1.05 | -10.16 | 9.73 |
| merton | 30,883 | 0.46 | 0.50 | 0.00 | 1.00 |
| kmv | 26,851 | 0.49 | 0.50 | 0.00 | 1.00 |
| **Systemic risk dependent variables** | | | | | |
| r2_merton | 138,489 | -4.79 | 2.36 | -10.22 | 1.79 |
| r2_kmv | 130,248 | -4.94 | 2.33 | -8.09 | 1.99 |
| r2_z_score | 150,616 | -3.88 | 2.08 | -7.72 | 8.76 |

Continued on next page

Table 1: Descriptive statistics of variables

| | count | mean | std. dev. | min. | max. |
|---|---|---|---|---|---|
| **Accounting** | | | | | |
| LONG_TERM_DEPOSITS | 44,612 | 10,444.96 | 59,427.88 | 0.00 | 1,486,700.00 |
| CASH_DEPOSITS | 45,047 | 7,484.41 | 50,508.04 | 0.00 | 1,305,600.00 |
| SAVINGS_DEPOSITS | 37,957 | 5,604.50 | 33,140.53 | 0.00 | 1,271,100.00 |
| TOTAL_LIABILITIES | 62,614 | 53,975.12 | 241,513.78 | 0.00 | 3,587,200.00 |
| EQUITY | 62,621 | 3,638.05 | 15,762.70 | -11,814.38 | 290,433.31 |
| TOTAL_DEPOSITS | 61,909 | 30,671.61 | 136,574.45 | 0.00 | 2,666,600.00 |
| TOTAL_LOANS | 57,560 | 27,272.33 | 107,509.84 | 0.00 | 1,913,300.00 |
| LOG_TOTAL_ASSETS | 62,561 | 8.12 | 2.26 | -4.98 | 15.14 |
| LEVERAGE_RATIO | 62,583 | 10.82 | 59.92 | -8,695.71 | 7,419.11 |
| NON_PERFORMING_LOANS | 37,230 | 98,555.15 | 777,826.91 | 0.00 | 27,479,333.00 |
| NET_INCOME | 67,676 | 154.38 | 2,086.75 | -37,968.62 | 75,453.00 |
| NON_INTEREST_INCOME | 24,360 | 348.83 | 1,333.91 | -30,731.20 | 21,061.00 |
| INTEREST_INCOME | 77,242 | 354.98 | 1,932.37 | -748.00 | 53,221.20 |
| NON_INTEREST_EXPENSES | 77,020 | 238.85 | 1,348.18 | -4,751.90 | 179,174.30 |
| INTEREST_EXPENSES | 76,528 | 232.29 | 1,552.02 | -907.20 | 49,475.30 |
| NET_INTEREST_INCOME | 77,097 | 213.89 | 1,001.81 | -13,974.50 | 32,945.70 |
| **Market** | | | | | |
| MARKET_CAP | 59,794 | 4,331.76 | 17,312.55 | 0.00 | 338,916.50 |
| MARKET_TO_BOOK_RATIO | 59,283 | 0.05 | 1.81 | 0.00 | 375.00 |
| HIST_STOCK_RETURN | 102,506 | 104.32 | 1,029.80 | 0.00 | 45,000.05 |
| STOCK_RETURN_STD_DEV | 65,369 | 0.02 | 0.01 | 0.00 | 0.14 |
| MEAN_INDEX_RETURN | 145,700 | 0.00 | 0.00 | -0.02 | 0.01 |
| INDEX_RETURN_STD_DEV | 145,700 | 0.00 | 0.00 | 0.00 | 0.00 |
| RISK_PREMIA | 99,874 | -0.02 | 0.32 | -11.12 | 11.13 |
| **Performance** | | | | | |
| ROA | 64,102 | 0.00 | 0.61 | -136.75 | 70.61 |
| ROE | 64,134 | 0.02 | 1.38 | -230.53 | 148.21 |
| ROA_TOTAL_ASSETS | 65,758 | 10,9856.78 | 489,601.33 | 0.02 | 7,093,400.00 |
| ROE_TOTAL_ASSETS | 65,784 | 6,934.17 | 30,379.58 | -23,522.89 | 571,103.66 |
| EQUITY_TOTAL_ASSETS | 65,747 | 0.10 | 0.24 | -19.90 | 1.00 |
| ROA_STD_DEV | 107,267 | 0.03 | 0.58 | 0.00 | 21.09 |
| **Competition and concentration** | | | | | |
| MARKET_SHARE | 61,517 | 0.06 | 0.16 | 0.00 | 1.00 |
| MARKET_SHARE_ASSETS | 62,566 | 0.06 | 0.16 | 0.00 | 1.00 |
| MARKET_SHARE_LOANS | 57,546 | 0.06 | 0.16 | 0.00 | 1.00 |
| HHI | 135,728 | 0.18 | 0.19 | 0.05 | 1.00 |
| CR5 | 135,728 | 0.72 | 0.15 | 0.32 | 1.00 |
| LERNER | 18,823 | -162.04 | 366.01 | -17,495.54 | 22,465.43 |
| LOG_N_BANKS | 155,775 | 5.28 | 2.20 | 0.00 | 7.17 |
| **Risk** | | | | | |
| RISK_EXPOSURE | 35,216 | 15.67 | 157.55 | 0.00 | 15,283.22 |
| TOTAL_ASSETS_RISK_EXPOSURE | 35,889 | 9.57 | 91.56 | 0.00 | 7,643.90 |
| BAILOUT_PROB_MERTON | 30,882 | 0.02 | 0.09 | 0.00 | 1.00 |
| BAILOUT_PROB_KMV | 26,850 | 0.02 | 0.08 | 0.00 | 1.00 |
| BAILOUT_PROB_MERTON_EX_I | 30,882 | 0.01 | 0.04 | 0.00 | 0.25 |
| BAILOUT_PROB_KMV_EX_I | 26,850 | 0.01 | 0.03 | 0.00 | 0.25 |
| **Macroeconomic and microeconomic** | | | | | |
| GDP_GROWTH | 147,264 | 0.65 | 0.60 | -2.88 | 6.13 |
| T_BILL_RETURN | 155,775 | 0.01 | 0.00 | 0.01 | 0.02 |
| MARGINAL_COST | 130,114 | 1.52 | 1.54 | -27.75 | 34.97 |

Note: Prepared by the authors.
std. dev. = standard deviation; min. = minimum value; max. = maximum value.
Monetary values expressed in dollars.
BAILOUT_PROB_<METHOD> variables employed only when the dependent variable was calculated using a method other than the one employed to calculate these independent variables.

Figure 2 shows the breakdown of observations by region (a) and income group (b) of the countries in which each bank in our sample has headquarters, the two only categorical variables. Most of the observations come from North America, followed by banks from Europe & Central Asia and East Asia & and the Pacific. As a consequence, most of the observations belong to banks in high-income countries. There are only 134 observations from low-income countries.

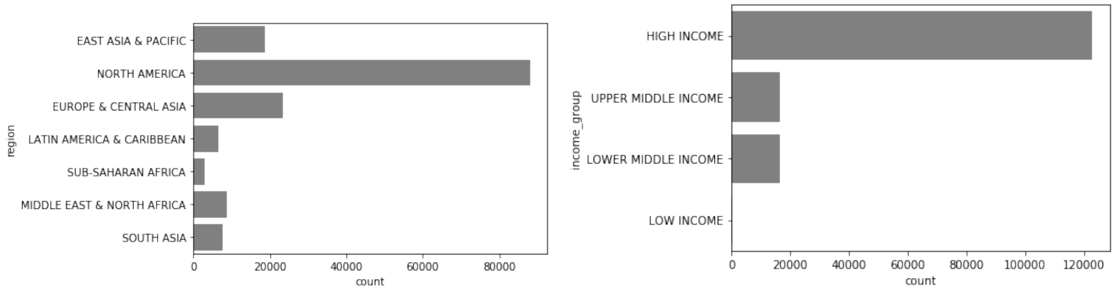


**Fig. 2** Breakdown of observations by (a) region and (b) income group. Source: authors' elaboration.

Preprocessing of the variables involved min-max scaling, encoding categorical values as numeric (one hot encoded as dummy variables), and imputing missing values with the k-nearest neighbors algorithm (with k = 2, euclidean distance as the distance metric for searching neighbors and weighting all neighbors of a given observation uniformly). Since we have observations ranging from 2000 to 2016, randomly dividing the observations into training and test samples would risk using observations from the future to predict an observation from the past. Observations up to 2012 were then used as the training sample (approximately 75% of the observations, according to the dependent variable and algorithm) and observations from 2012 to 2016 were employed as the test sample (approximately 25% of the observations, according to the dependent variable and algorithm as well).

We then employed a LASSO regression for pre-selecting variables to be used in the prediction of default and systemic risks, employing both Random Forest and XGBoost methods, explained in the following subsections. While these models are not state-of-the-art when applied in a standalone fashion for default probability prediction, the primary objective of our study is to apply the explainable artificial intelligence framework (by means of SHAP values described in subsection 2.5) to contribute to this problem domain.

The best parameters for each implementation model were chosen with a time series cross-validation approach, to take into account the temporal nature of the observations mentioned before. We aim to maximize the $R^2$ measure since our purpose here is to enhance the explaining power of the models. The values used in the cross-validation for each of the hyperparameters are shown in table 2.

Table 2: Hiperparameters values used in cross-validation

| Hiperparameter | Model | Values |
|---|---|---|
| alpha | Lasso | 100 values from 0.00001 to 0.5 |
| max_depth | Random Forest | 10, 100, 200 |
| max_features | Random Forest | auto, sqrt and log2 |
| min_samples_split | Random Forest | 3, 5, 30 |
| min_samples_leaf | Random Forest | 5, 10, 80 |
| n_estimators | Random Forest | 100, 1000, 2000 |
| colsample_bylevel | XGBoost | 0.6, 0.7 |
| colsample_bytree | XGBoost | 0.6, 0.7 |
| gamma | XGBoost | 0.01, 1 |
| learning_rate | XGBoost | 0.0001, 1 |
| max_delta_step | XGBoost | 0.1, 10 |
| max_depth | XGBoost | 6, 15 |
| min_child_weight | XGBoost | 10, 500 |
| n_estimators | XGBoost | 10, 100 |
| reg_alpha | XGBoost | 0.1, 100 |
| reg_lambda | XGBoost | 0.1, 100 |
| subsample | XGBoost | 0.4, 0.7 |

Note: Prepared by the authors.

Hiperparameters are named as in the scikit-learn and xbgoost Python modules.
alpha described in section 3.1.
Random forest hiperparamaters described in table 5.
XGBoost hiperparamaters described in table 7.

SHAP values for each of the dependent variables (z_score, kmv, merton, r2_z_score, r2_kmv, and r2_merton) were then calculated for each method (Random Forest and XGBoost) and their importance was analyzed. Results for the best default risk and systemic risk dependent variable and method are presented in section 4.

We employed the scikit-learn implementation for the random forest model (Pedregosa et al, 2011), the xgboost module implementation for XBGoost (Chen and Guestrin, 2016), and the shaps implementation for shap values (Lundberg and Lee, 2017).

Figure 3 summarizes the methodology employed in our study:

1. Calculate default and systemic risk measures according to Z-score, KMV, and Merton models as dependent variables;
2. Filter independent variables (16 accounting, 7 competition and concentration, 6 performance, 3 macro, and microeconomic, 7 market-related, and 6 risk-related features) using Lasso regression;
3. Forecast default or systemic risk according to RF and XGBoost models;
4. Explain the variables' importance with SHAP values for the two best models (one for default and one for systemic risk).

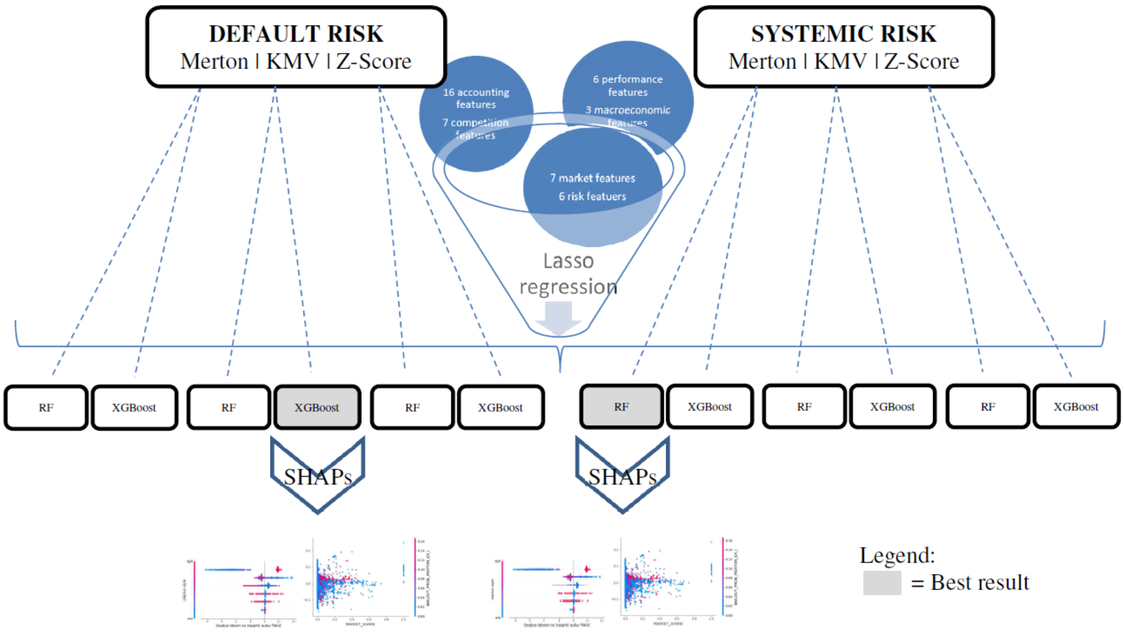


**Fig. 3** Methodology summary: Lasso regression filters independent variables that forecast, with random forest and XGBoost models, default and systemic risk calculated with Z-score, KMV, and Merton models. The best-performing models for each kind of risk were analyzed with SHAP values. Source: authors' elaboration.

## 3.1 Least Absolute Shrinkage and Selection Operator (LASSO)

The Least Absolute Shrinkage and Selection Operator (LASSO) for Tibshirani (1996) is a regularization method in which a penalty term is added to the likelihood function optimized in linear regression. The unconstrained OLS estimates $\hat{\boldsymbol{\beta}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T y$ can be vulnerable to high variance, which in turn can affect inference negatively. Thus, a penalty term for the magnitude of the coefficients can control the variance. Similarly to ridge regression, in which the penalty term is the $\ell_2$ norm for the $\boldsymbol{\beta}$ parameters, in LASSO the penalty is the $\ell_1$ norm. The main difference is that the LASSO can yield a set of sparse solutions for the betas, making LASSO an embedded feature selection method, as the algorithm training process is done simultaneously with the feature selection.

The coefficients of the LASSO regression are the solutions to the following constrained optimization problem:

$$\boldsymbol{\beta}(\boldsymbol{\alpha}) = argmin_{\boldsymbol{\beta}} \left\{ \frac{1}{N}(y - \boldsymbol{x}\boldsymbol{\beta})^{\mathrm{T}}(y - \boldsymbol{x}\boldsymbol{\beta}) + \alpha||\boldsymbol{\beta}||_1 \right\} \tag{18}$$

where $||.||_1$ is $\ell_1$ norm operator and $\alpha$ is a free regularization parameter that controls the degree of shrinkage of the betas. Therefore, sufficiently large values for $\alpha$ will effectively force some betas to be zero, producing a sparse solution for the LASSO estimator. In general, the optimal $\alpha$ that minimizes the out-of-sample error is found by manually tuning through cross-validation.

The values for the hyperparameters selected by the cross-validation procedure are shown in tables 3 and 4.

Table 3: Fitted lasso hyperparameter for each dependent original variable

| Dependent variable | z_score | merton | kmv |
|---|---|---|---|
| alpha | 0.353542 | 0.00001 | 0.00001 |

Note: Prepared by the authors. Hiperparameter named as in the scikit-learn Python modules. alpha is the regularization parameter in equation 18.

Table 4: Fitted lasso hyperparameter for each dependent transformed variable

| Dependent variable | r2_z_score | r2_merton | r2_kmv |
|---|---|---|---|
| alpha | 0.030313 | 0.030313 | 0.00001 |

Note: Prepared by the authors. Hiperparameter as in table 3.

### 3.2 Random Forest

Random forest (Breiman, 2001) is an ensemble-based model composed of bootstrap aggregating of several decision trees. In each tree, the observations present in each node are recursively partitioned into two mutually exclusive subsets such that the subsets generated have maximum purity (homogeneity), in the sense that an ideal decision rule would be able to completely separate two different classes without misclassifications (Breiman et al, 2017; Genuer and Poggi, 2020). The complexity of a decision tree can be controlled by the number of nodes and the minimum size of partitions of sub-nodes. As the name suggests, random forest creates several decision trees and groups them into one "forest" of trees, taking bootstrap samples with size $m$ from the set of explanatory variables (features) when partitioning the tree at each of its nodes. The final decision vote will be given by the majority vote for classification problems (and the average value for regression problems). Thus, by grouping several decision trees, random forest reduces the overall level of variance of decision trees while maintaining a low bias for data observed in-sample (Genuer and Poggi, 2020).

As proposed by Breiman (2001), the random forest also yields a measure of feature importance for each explanatory variable, calculated as a weighted average between all nodes of the decrease in the respective node impurity, with the node probability being estimated by the proportion of bootstrap samples that reach that specific node; the intuition is that a feature tends to be more relevant to explain the target variable the more that it can contribute to building decision rules that reduce the node's impurity more strongly. Algebraically, the estimated importance of the i-th variable at the t-th tree is defined as:

$$\mathcal{R}_{i_t} = \frac{\sum\limits_{j \in \mathbb{J}} \mathcal{N}_j}{\sum\limits_{l=1}^{\Omega} \mathcal{N}_l} \tag{19}$$

where $\mathbb{J}$ is the set of nodes that splits on the i-th feature, $\Omega$ is the number of nodes in the decision tree, and $\mathcal{N}_j$ is the importance of the j-th node, given as:

$$\mathcal{N}_j = w_j \xi_j - \sum_{k \in j^{\#}} w_k \xi_k \tag{20}$$

where $w_j$ is the weighted number of samples that reaches the j-th node, $\xi_j$ is the value of the impurity function at the j-th node, and $j^{\#}$ is the set of all child nodes split from the j-th node.

Finally, the estimated importance for the i-th variable in the random forest is simply the average importance across all $T$ decision trees:

$$\mathcal{N}_{RF_i} = \sum_{t=1}^{T} \frac{\mathcal{R}_{i_t}}{T} \tag{21}$$

In this paper, we estimated the random forest with binary decision trees (each node has two child nodes), composed by $T = 2000$ trees (n_estimators in the sci-kit learn implementation) and with $m = 100\%$ of the observations (max_samples in the sci-kit learn implementation), using the square error as the purity function to define the splitting rules, following Breiman (2002)'s recommendations.

The values for the hyperparameters selected by the grid search procedure are shown in tables 5 and 6.

Table 5: Fitted random forest hyperparameters for each default risk dependent variable

| Hyperparameter | z_score | merton | kmv |
|---|---|---|---|
| max_depth | 10 | 100 | 100 |
| max_features | log2 | auto | auto |
| min_samples_leaf | 5 | 10 | 10 |
| min_samples_split | 3 | 5 | 3 |
| n_estimators | 2000 | 2000 | 2000 |

Note: Prepared by the authors.
Hiperparameters are named as in the scikit-learn Python module.
max_depth is the maximum depth of the trees.
max_features is the number of features used to determine the best split at each step ("auto" employs the total number of features available, "sqrt" is the square root of the total number of features and "log2" e the log in base 2 of the total number of features).
min_samples_split is the minimum number of observations to split an internal node.
min_samples_nodes is the minimum number of observations for a node to be a leaf.
n_estimators is the number of trees in the forest.

Table 6: Fitted random forest hyperparameters for each systemic risk dependent variable

| Hyperparameter | r2_z_score | r2_merton | r2_kmv |
|---|---|---|---|
| max_depth | 100 | 100 | 100 |

Table 6: Fitted random forest hyperparameters for each systemic risk dependent variable

| Hyperparameter | r2_z_score | r2_merton | r2_kmv |
|---|---|---|---|
| max_features | log2 | auto | auto |
| min_samples_leaf | 5 | 10 | 5 |
| min_samples_split | 30 | 5 | 5 |
| n_estimators | 100 | 100 | 1000 |

Note: Prepared by the authors. Hiperparameters as in table 5.

### 3.3 XGBoost

The XGBoost algorithm used in this work is an efficient implementation of the Gradient Boosted Tree (or Gradient Tree Boosting or Gradient Boosted Regression Trees) algorithm, also known as Gradient Boosting Machine (GBM) (Chen and Guestrin, 2016). It is built on the decision trees algorithm described earlier (section 3.2) and is known for its speed and accuracy, although tuning its hyperparameters can be non-trivial (Müller and Guido, 2017).

Boosting attempts to combine several weak predictors into a strong one by training them sequentially, with each one attempting to correct its predecessor. The most common methods for this are AdaBoost and Gradient Boosting. AdaBoost trains the predictors sequentially, changing the weights of each predictor according to its accuracy for classification problems (or any chosen measure, including those applicable to regression problems). The more accurate the predictor, the higher its weight in the final ensemble. Gradient Boosting, in turn, fits each new predictor to the residual errors of its predecessor, trying to minimize a given loss function. Mathematically, the algorithm tries, at each step t, to minimize the regularized objective given by equation 22:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x_i})) + \Omega(f_t) \tag{22}$$

where $l$ is a convex and differentiable loss function, $\mathbf{x_i}$ is the vector of all features for observation $i$, $y_i$ is the actual value for the target or dependent variable for observation $i$, $\hat{y}_i^{(t-1)}$ is the prediction given by the previous model for instance $i$, $f_t$ is the model built on the residuals, and $\Omega$ is a regularization term, also known as the learning rate, which penalizes complexity and prevents over-fitting. According to Chen and Guestrin (2016), this "cannot be optimized using traditional optimization methods in Euclidean space" and thus, a second-order Taylor approximation can be employed, as in equation 23:

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n} [l(y_i, \hat{y}^{(t-1)}) + g_i f_t(\mathbf{x_i})) + \frac{1}{2} h_i f_t^2(\mathbf{x_i})] + \Omega(f_t) \tag{23}$$

where $\hat{y}^{(t-1)}$ is the mean prediction of the target variable in step t, $g_i$ and $h_i$ are the first and second order gradients on the loss function, respectively, given by equations 24 and 25:

$$g_i = \frac{\partial l(y_i, \hat{y}^{(t-1)})}{\partial \hat{y}^{(t-1)}} \tag{24}$$

$$h_i = \frac{\partial^2 l(y_i, \hat{y}^{(t-1)})}{\partial (\hat{y}^{(t-1)})^2} \tag{25}$$

The loss function can take many forms, usually defined as the root mean squared error (RMSE) for regression problems and the logistic log loss for classification problems such as the task in this paper. The logistic log loss function is given by equation 26:

$$\mathcal{L}^{(t)} = -\sum_{i=1}^{n} y_i \ln(p) + (1 - y_i) \ln(1 - p) \tag{26}$$

where $y_i$ is the actual class of instance i and $p$ is its associated probability (or pseudo-probability) score, given by the sigmoid/logistic equation (equation **??**). The final prediction is the sum of the predictions from each tree in the model. The values for the hyperparameters selected by the cross-validation procedure in our study are shown in tables 7 and 8.

Table 7: Fitted XGB hyperparameters for each original dependent variable

| Hyperparameter | z_score | merton | kmv |
|---|---|---|---|
| colsample_bylevel | 0.6 | 0.7 | 0.7 |
| colsample_bynode | 1 | 1 | 1 |
| colsample_bytree | 0.6 | 0.7 | 0.7 |
| gamma | 0.01 | 1 | 1 |
| learning_rate | 1 | 1 | 1 |
| max_delta_step | 0.1 | 0.1 | 0.1 |
| max_depth | 15 | 15 | 6 |
| min_child_weight | 10 | 10 | 10 |
| n_estimators | 100 | 10 | 10 |
| reg_alpha | 0.1 | 0.1 | 0.1 |
| reg_lambda | 100 | 100 | 100 |
| subsample | 0.7 | 0.7 | 0.7 |

Note: Prepared by the authors.
Hiperparameters are named as in the xgboost Python module.
colsample_bylevel, colsample_bynode and colsample_bytree
are the subsample ratio of columns for each level, split and tree, respectively.
gamma is the minimum reduction in loss to make another partition on a leaf node.
learning_rate is the boosting learning rate.
max_delta_step is the maximum delta to estimate each tree weigth.
max_depth is the maximum depth of the trees for base learners.
min_child_weight is the minimum sum of instance weights in a child node.
n_estimators is the number of gradient boosted trees.
reg_alpha is the L1 regularization on weights.
reg_lambda is the L2 regularization on weights.
subsample is the subsample ratio of instances in training.

Table 8: Fitted XGB hyperparameters for each smoothed dependent variable

| Hyperparameter | r2_z_score | r2_merton | r2_kmv |
|---|---|---|---|
| colsample_bylevel | 0.6 | 0.7 | 0.7 |
| colsample_bynode | 1 | 1 | 1 |
| colsample_bytree | 0.7 | 0.6 | 0.7 |
| gamma | 0.01 | 1 | 1 |
| learning_rate | 1 | 1 | 1 |
| max_delta_step | 0.1 | 0.1 | 10 |
| max_depth | 6 | 6 | 15 |
| min_child_weight | 10 | 10 | 500 |
| n_estimators | 100 | 100 | 100 |
| reg_alpha | 0.1 | 0.1 | 100 |
| reg_lambda | 100 | 100 | 100 |
| subsample | 0.4 | 0.7 | 0.7 |

Note: Prepared by the authors. Hiperparameters as in table 7.

### 3.4 Shapley Additive Explanation (SHAP)

Finally, to provide interpretability for the results yielded by the machine learning models, we applied Shapley Additive Explanation (henceforth SHAP), an Explainable Artificial Intelligence method introduced by Lundberg and Lee (2017) that aims at explaining individual machine learning model predictions, inspired by Shapley (1953)'s work on cooperative game theory. Besides, SHAP is a model-agnostic method, implying that it can be applied regardless of the chosen prediction model, in contrast with model-specific methods for variable importance, such as the importance metric of decision trees and random forest algorithms. As discussed in Lundberg and Lee (2017), SHAP unifies a wide class of additive feature attribution techniques used for machine learning model explanations, such as Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al, 2016), which approximates linear interpretable models near a given prediction; and Shapley sampling values (Štrumbelj and Kononenko, 2014), which provide estimates for feature importance in linear models under the presence of multicollinearity, by approximating the effect of removing each feature from the learner as a weighted average of differences between the predictions of a model trained with and without the respective feature. In this sense, while being computationally expensive, SHAP values assign importance values for each feature for a particular prediction, thus allowing to decompose of the impact of each variable in the predicted outcome compared to the average prediction for the sampled observations.

The main motivation of Shapley (1953)'s article was to show under which circumstances the cooperation between a set of players leads to a larger aggregate payoff over the scenario in which all players act individually, and how the surplus payoff is distributed between the players. Also, the players are assumed to have different "weights" (relevance) in the coalition. Adapting this rationale to the context of machine learning and explainable artificial intelligence, the set of explanatory variables (features) act like the "players" that have different contributions to the model's prediction, which is a result of the decision function, who plays the role of the "coalition". In this sense, as pointed out by Strumbelj and Kononenko (2010), Shapley (1953)'s problem can be adapted to obtain how each feature contributes to the model's decision for each new observation, which in turn can be regarded as an evaluation method of local model-agnostic variable importance.

In a general case of a dependent variable $y = f(\boldsymbol{x})$ being explained by a vector $\boldsymbol{x}$ composed by $p$ explanatory variables, by means of a decision function $f(.)$. We first define $\Pi$ as a permutation of the indexes $\{1, 2, ..., p\}$ which represents an ordering of the $p$ features. Furthermore, we define $\pi(\Pi, i)$ as the set of the indexes associated with features that come before the $i - th$ variable in permutation $\Pi$ – in other words, for permutation $\Pi = \{3, 1, 4, 2\}$, $\pi(\Pi, 4) = \{3, 1\}$ and $\pi(\Pi, 3) = \emptyset$. Finally, we define $f(x_{\#})$ as the predicted value of decision function $f(.)$ applied to a particular observation of interest $x_{\#} \in \mathbb{R}^p$.

We now define the SHAP value as:

$$\varphi(x_{\#}, i) = \frac{1}{p!} \sum_{\Pi} \Delta^{i|\pi(\Pi,i)}(x_{\#}) \tag{27}$$

with

$$\begin{aligned}\Delta^{i|\pi(\Pi,i)}(x_{\#}) =& \mathbb{E}_{\boldsymbol{x}}\{f(\boldsymbol{x})|x_{\pi(\Pi,i)_1} = x_{\pi(\Pi,i)_{1_\#}}, \ldots, x_{\pi(\Pi,i)_{|\pi(\Pi,i)|}} = x_{\pi(\Pi,i)_{|\pi(\Pi,i)|_\#}}, x_i = x_{i_\#}\} - \\ & \mathbb{E}_{\boldsymbol{x}}\{f(\boldsymbol{x})|x_{\pi(\Pi,i)_1} = x_{\pi(\Pi,i)_{1_\#}}, \ldots, x_{\pi(\Pi,i)_{|\pi(\Pi,i)|}} = x_{\pi(\Pi,i)_{|\pi(\Pi,i)|_\#}}\}.\end{aligned} \tag{28}$$

where $p!$ is the number of possible permutations (orderings) between the $p$ explanatory variables, $\mathbb{E}$ is the expected value operator, $|.|$ is the cardinality operator, and $\Delta^{i|\pi(\Pi,i)}(x_{\#})$ is the difference between the expected prediction made by $f(.)$ for observation $x_{\#}$ using the features with indexes $i \cup \pi(\Pi, i)$ and the expected prediction made by $f(.)$ for observation $x_{\#}$ using only the features with indexes $\pi(\Pi, i)$.

Therefore, in summary, the SHAP value associated with each feature $x_i$ to a specific observation $x_{\#}$ is the average of the relative impact of $x_i$ on the model's prediction across all possible cases of feature permutations.

For a dataset with a large number of features $p$, the computation of the SHAP value can be unfeasible, in which cases alternative formulations based on consistent estimates using permutation samples and Monte Carlo methods can be applied to reduce processing time, as seen in, for instance, Štrumbelj and Kononenko (2014) and Lundberg and Lee (2017).

Similarly to the Shapley values from Shapley (1953), SHAP values have the properties of 1) symmetry (SHAP values of interchangeable variables are equal); 2) additivity (the sum of SHAP values calculated for two models is equal to the SHAP value of the sum of those two models); 3) dummy feature (SHAP value of irrelevant features is equal to zero); and 4) local accuracy (the sum of the SHAP values of all features is equal to the model's prediction for that specific observation).

## 4 Results and discussion

### 4.1 Performance of the models

Tables 9 and 10 present $R^2$, RMSE (root mean square error), and MAE (mean absolute error) for the three default risk-dependent variables (one for each of the models described in section 2.1) and for the three systemic risk dependent variables (one for each model described in section 2.2), respectively, according to either of the two machine learning models (random forest and XGBoost) described in section 3.

For the default risk dependent variables, the best performance was achieved by the XGBoost algorithm with the dependent variable calculated with the KMV model, when $R^2$ reached almost 60%, i.e almost 60% of the bank's default risk variability is explained by the features selected by the LASSO regression, and both RMSE and MAE were the smallest of the set.

Table 9: Evaluation measures for each default risk dependent variable and model

| Dependent variable | Model | $R^2$ | RMSE | MAE |
|---|---|---|---|---|
| z_score | Random Forest | 0.069642 | 71.997645 | 34.393490 |
| z_score | XGB | 0.059789 | 72.377919 | 35.207742 |
| merton | Random Forest | 0.506396 | 0.349224 | 0.299259 |
| merton | XGB | 0.449082 | 0.368942 | 0.263609 |
| kmv | Random Forest | 0.412020 | 0.382528 | 0.352650 |
| kmv | XGB | 0.592897 | 0.318298 | 0.236718 |

Note: Prepared by the authors.

For the systemic risk variables, the best performance was achieved when the dependent variable was calculated with the Merton model and Random Forest was employed as the forecasting method, when $R^2$ reached almost 71.5% and RMSE and MAE were the smallest of the set.

Table 10: Evaluation measures for each systemic risk dependent variable and model

| Dependent variable | Model | $R^2$ | RMSE | MAE |
|---|---|---|---|---|
| r2_z_score | Random Forest | 0.429070 | 1.452417 | 1.188878 |
| r2_z_score | XGB | 0.474432 | 1.393524 | 1.156054 |
| r2_merton | Random Forest | 0.714231 | 1.376529 | 1.114268 |
| r2_merton | XGB | 0.540651 | 1.745216 | 1.373592 |
| r2_kmv | Random Forest | 0.342054 | 1.583431 | 1.207218 |
| r2_kmv | XGB | 0.461598 | 1.432377 | 1.164294 |

Note: Prepared by the authors.

The variables selected by the LASSO regression for the dependent variable calculated with the KMV method, the best result among the default risk variables, are shown in figure 4. Variables with zero coefficients were discarded. Results for the models reported in table 9 include only these

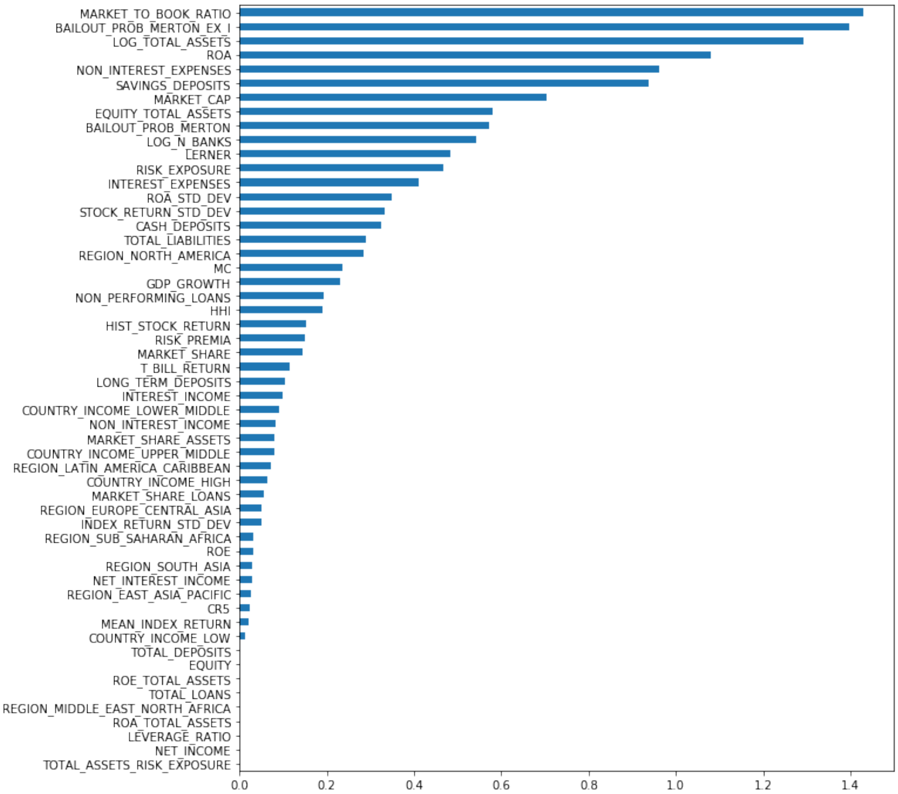


**Fig. 4** Selected variables by the Lasso regression for the default risk dependent variable calculated by the KMV method. Source: author's elaboration.

variables. The LASSO regression could filter only a few variables related to accounting, performance, and geographical location, variables that are highly correlated with others that were kept.

With the systemic risk variables, the Lasso regression was able to filter more variables, keeping only six of them, as shown in figure 5 for the $r2\_merton$ dependent variable, the best one among the systemic risk variables. As with the default risk ones, results reported in table 10 include only these variables.

## 4.2 Importance of variables

### *4.2.1 Default risk*

For predicting the default risk calculated with the KMV method, the most important variables, calculated by the SHAP method described in section 3.4, are BAILOUT_PROB_MERTON_EX_I, MARKET_SHARE_ASSETS, MARKET_TO_BOOK_RATIO, TOTAL_LIABILITIES, BAILOUT_PROB_MERTON, and LOG_N_BANKS, as shown in the global variable importance graphs of figure 6. These results are consistent with the Merton model, not shown. For the Z-score (not shown), geographical location categorical variables and LOG_N_BANKS are the most important. Since the KMV and Merton models use basically as inputs for accounting and market variables, their results were expected to be similar. For the Z-score, only accounting variables are used and the location of the bank, together with the number of banks in the country, play a major role in predicting default risk.

Variables BAILOUT_PROB_MERTON and BAILOUT_PROB_MERTON_EX_I are proxies for the perception of the probability of liquidity injection by central banks in the financial system, calculated with the Merton method, as described in section 2.3.3. These variables are relatively similar, but BAILOUT_PROB_MERTON_EX_I excludes bank $i$ from the analysis used to calculate it, while BAILOUT_PROB_MERTON does not. It is worth noting that these variables were only employed when the dependent variable was calculated with the KMV model (and

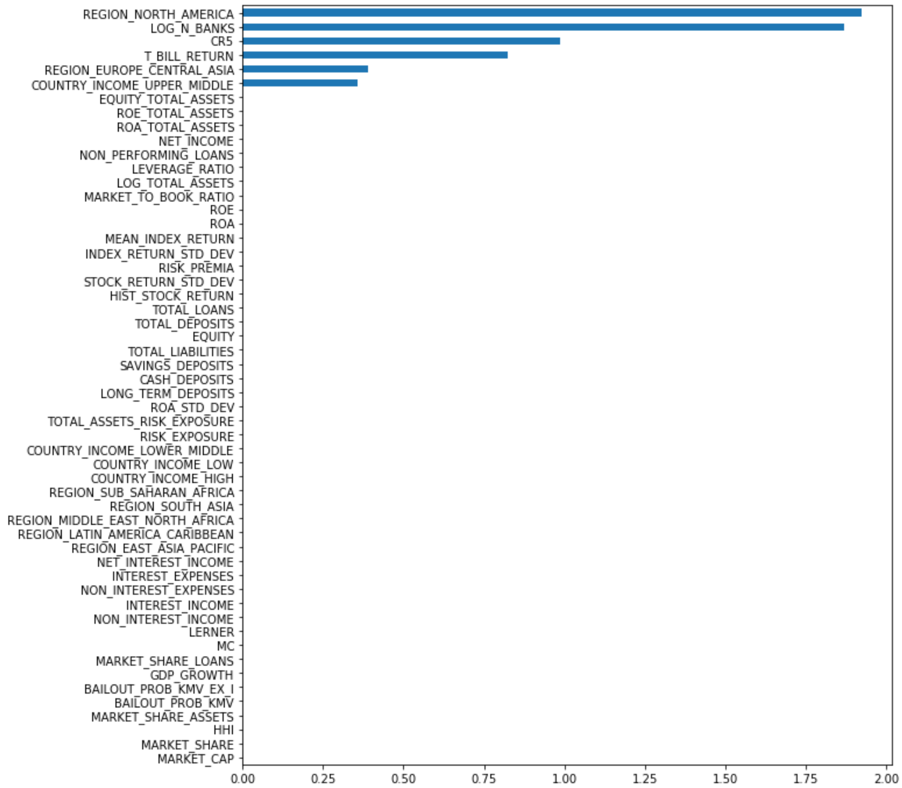


**Fig. 5** Selected variables by the Lasso regression for the systemic risk dependent variable calculated by the Merton method. Source: author's elaboration.

BAILOUT_PROB_KMV and BAILOUT_PROB_KMV_EX_I independent variables were only employed when the dependent variable was calculated with the Merton model). The higher their value, the higher a bank's default probability predicted by the algorithm, as expected, i.e, expectations for bailout probabilities increase the risk of the bank's default, even though in more concentrated financial markets, as shown in the partial dependence plots of the figure 7 (plots (a) and (e)). These plots show, for the value of the independent variable of each observation, the corresponding SHAP value. These findings are almost in line with what is reported by Schiozer et al (2018), for example.

MARKET_SHARE_ASSETS is the market share of a bank in asset terms. As expected, higher values of this variable are associated with a lower KMV prediction or a lower bank's default probability, as shown in figure 7, panel (b). MARKET_TO_BOOK_RATIO, a bank's market capitalization in relation to its book value, shows a slight positive relationship with the predicted KMV measure, with higher-valued market capitalization banks as the ones that make the algorithm predict higher default probabilities more often, as shown in figure 7, panel (c). These results show that banks with relatively low market share but higher market capitalization in relation to their book values are the ones for which the algorithm predicts higher default risk values. These findings corroborate results reported by Beck et al (2013), for example.

Finally, TOTAL_LIABILITIES also play an important role in the predictions, with higher values of these variables pulling predictions to the positive camp, as expected and shown in figure 7, panel (d). That is, the algorithm tends to predict higher KMV scores for banks with higher total liabilities. Bussmann et al (2021) also find that accounting and performance variables (total assets to total liabilities, EBITDA to interest coverage ratio, shareholders funds plus non-current liabilities and profits before taxes plus interests paid, current ratio and others in their case) are also important. They analyze the probability of default of companies for which loans were granted, though.

It is worth noting that these variables are not exactly the most important variables according to the Lasso possibly due the nonlinearities described in section 2.4 and shown in figure 7.

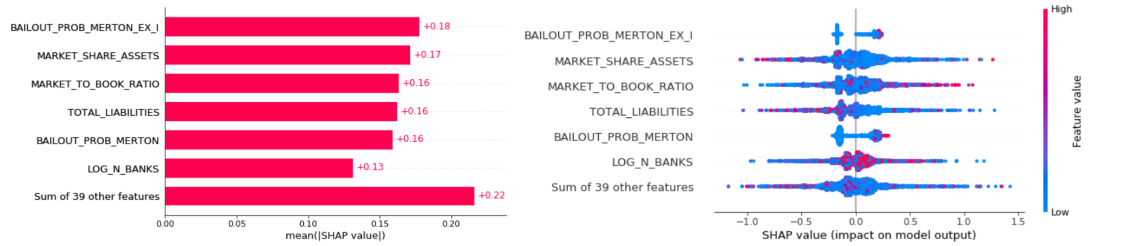


**Fig. 6** Global variable importance graphs for the dependent variable calculated with the KMV model and XGBoost as the forecasting method; (a) average SHAP values and (b) all individual SHAP values. Source: author's elaboration.

Figure 7 also presents the interaction of each of the six most important variables with other selected variables. Higher values of BAILOUT_PROB_MERTON_EX_I, for instance, are associated with higher values of MARKET_SHARE and with higher values of default probability (panel a). Higher values of MARKET TO BOOK RATIO, when associated with low values of TOTAL LIABILITIES, tend to pull predictions to the higher values of default probabilities (panel f).

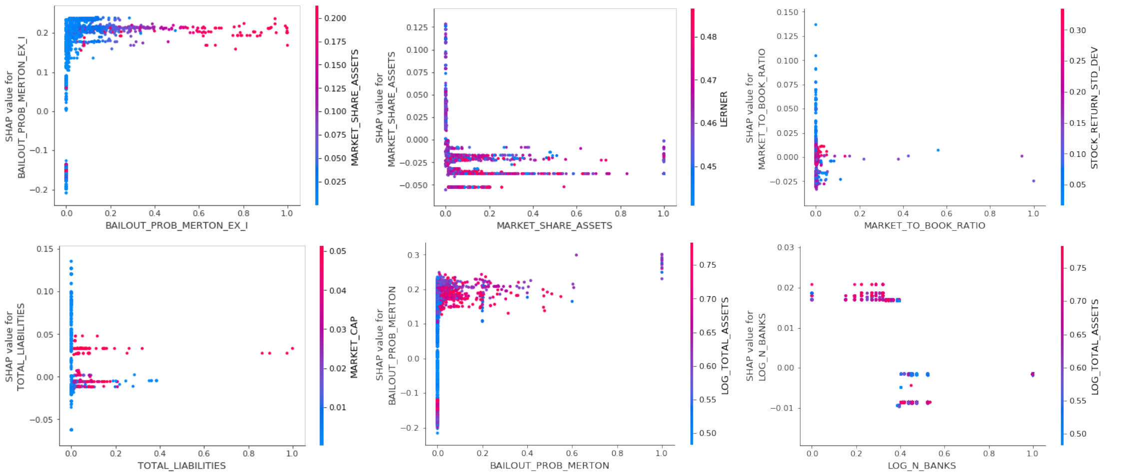


**Fig. 7** Partial dependence plots with interactions for the KMV model with XGB as the forecasting method (a) BAILOUT_PROB_MERTON_EX_I (b) MARKET_SHARE_ASSETS (c) MARKET_TO_BOOK_RATIO (d) TOTAL_LIABILITIES (e) BAILOUT_PROB_MERTON (f) LOG_N_BANKS. Source: author's elaboration.

Finally, it is also possible to dig into the most important variable for specific observations' predictions, as shown in the local explainability plots of figure 8. We present two observations where the predictions are respectively low (left panel) and high (right panel). The left panel shows an observation for which the algorithm predicts a low probability of default. The bailout probabilities are the main variables pulling the prediction below average for this bank. MARKET_CAP and MARKET_SHARE_ASSETS, on the other hand, drive the prediction higher, but they are not enough to offset the effects of the two first variables. This is probably a big conservative bank with a relatively high market share and capitalization but a low market-to-book ratio. The sum of the contributions of each variable (in this case, negative), produces a final prediction below the average.

For the bank in the right panel, bailout perception is also the main driver, but they are now high, pulling predictions above the average. Among the most important variables, only MARKET_SHARE_ASSETS and TOTAL_LIABILITIES contribute to a lower value for the prediction of the default probability. This bank seems to be a relatively small bank, with a relatively small market cap.

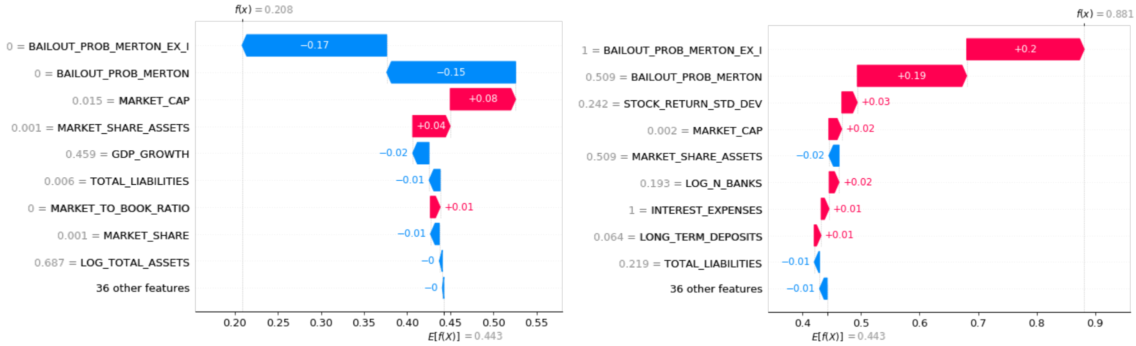


**Fig. 8** Local explainability plots for the KMV model calculated with the XGB algorithm (a) for a low prediction observation and for (b) a high prediction observation. Source: authors' elaboration.

### *4.2.2 Systemic risk*

For the systemic risk dependent variable calculated with the Merton model, the most important variables are LOG_N_BANKS, T_BILL_RETURN, CR5, COUNTRY_INCOME_UPPER_MIDDLE, REGION_EUROPE_CENTRAL_ASIA, REGION_NORTH_AMERICA as shown in figure 9. These results are consistent with the results from the KMV and Z-score models, not shown.

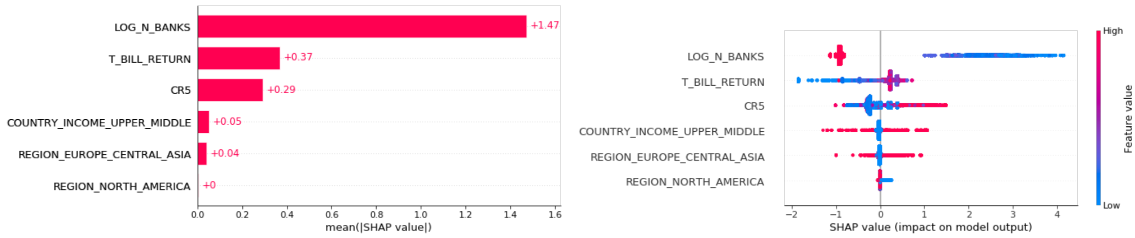


**Fig. 9** RF model for Merton systemic risk measure (a) mean SHAP values and (b) individual SHAP values. Source: author's elaboration.

In systemic risk forecasting, competition plays a more important role. Higher numbers of banks (LOG_N_BANKS) in a country pull the predictions of the algorithms towards lower values for the Merton model, as shown in figure 10, panel (a). It shows that a possibly less concentrated market contributes to lower predictions for systemic risk. The CR5 measure corroborates this perception, since higher values of this variable, indicating more concentrated markets, pull the predictions for the Merton model to higher values, and consequently to higher systemic risk, as shown in figure 10, panel (c). These findings are in line with Anginer et al (2014) reports. They find a strong negative correlation between competition and systemic risk.

The macroeconomic variable T_BILL_RETURN, a proxy for the general level of interest rates in the world economy, shows that a more restricted environment, represented by higher interest rates, tends to pull systemic risk predictions by the RF algorithm toward a higher level, as shown in figure 10, panel (b). This result is expected since higher interest rates increase the cost of funding to banks.

Finally, the geographical location of a bank also plays a crucial role in the predictions of systemic risk. The location of a bank in North America tends to pull predictions into the lower systemic risk zone, as shown in figure 10, panel (f). Individual banks may face relatively high default risk, especially small banks, but the system as a whole is considered solid and, hence, the measures of systemic risk are low. Europe's and Central Asia location tends to pull systemic risk in higher value zones (figure 10, panel (e)), probably due to the period of analysis, starting in 2012, the year of the European sovereign debt crisis. The perception of systemic risk in Europe is higher than in other regions (most of the observations come from developed countries' banks, as shown in figure 2).

Figure 11 presents local explainability plots for two observations where the predictions are respectively low (left panel) and high (right panel). The bank from the left panel is in a competitive

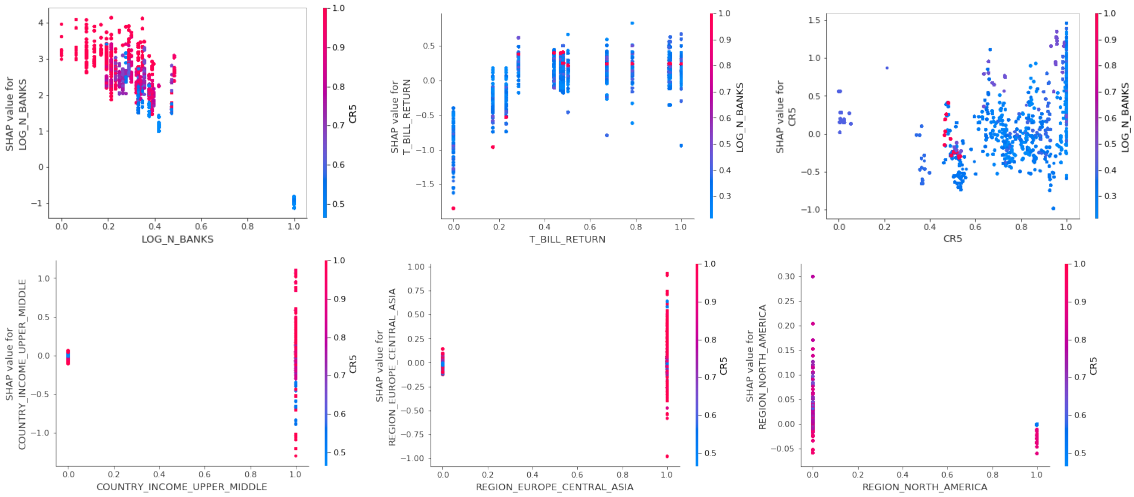


**Fig. 10** Partial dependence for the RF model for Merton systemic risk with interactions (a) LOG_N_BANKS (b) T_BILL_RETURN (c) CR5 (d) COUNTRY_INCOME_UPPER_MIDDLE (e) REGION_EUROPE_CENTRAL_ASIA (f) REGION_NORTH_AMERICA. Source: authors' elaboration.

environment, and both LOG_N_BANKS and CR5 play the leading role in predicting systemic risk, which is low in accordance with Anginer et al (2014), for example. The bank from the right panel is located in a concentrated market and the important variables are the same, but now pulling the systemic risk prediction to a higher value. In this case, moral hazard may play an important role, with the bank counting on an implicit guarantee from the government. It can then assume riskier operations, increasing the perception of systemic risk in the country.

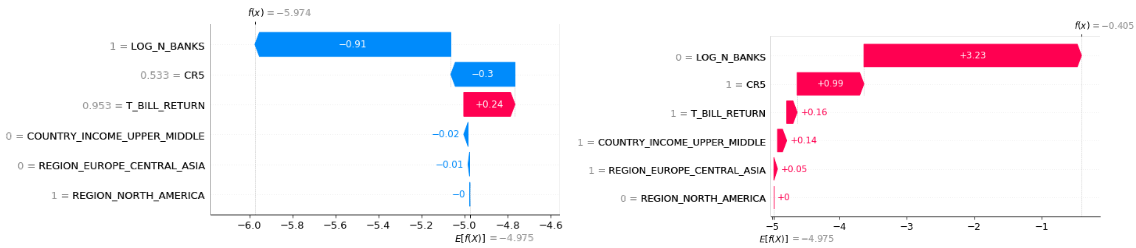


**Fig. 11** Local explainability plots for the systemic risk Merton model calculated with the RF algorithm (a) for a low prediction observation and for (b) a high prediction observation. Source: authors' elaboration.

## 5 Final remarks

In this paper, we explained banking default risk and systemic risk predictions with SHAP values, an agnostic explainable artificial intelligence (XAI) method based on game theory that shows variable importance based on the contribution of each variable to each prediction.

Our data comprise 55,775 observations from 2,325 world banks. For each observation, there are 45 variables with accounting, market, performance, competition and concentration, risk, and macroeconomic and microeconomic measures for each bank and quarter (from 2000 to 2016). Dependent variables for both default and systemic risks were calculated with the Merton, KMV, and Z-score models, totaling six possibilities. We first filtered independent variables through Lasso regressions and then applied both random forest and XGBoost to predict each of the six dependent variables.

The best results according to the $R^2$ measure were achieved when default risk was calculated with the KMV model and predicted with the XGBoost algorithm and when systemic risk was calculated with the Merton model predicted with the random forest algorithm. These two sets of predictions were then subjected to the SHAP values methodology. According to this methodology, we found that the most important variables for the default risk prediction are the probability of a bailout calculated with the Merton model, the market share in terms of assets, the market-to-book ratio, total liabilities, and the number of banks in the market (a measure of concentration). These results are in line with results from Schiozer et al (2018), Beck et al (2013), and Bussmann et al (2021), for example.

For the systemic risk predictions, the most important variables are the number of banks in the country, the level of interest rates, the market share of the top 5 largest banks, and the region of the bank (in North America and Europe). The competition and concentration variable's importance corroborate the work of Anginer et al (2014), for example.

These results may be of interest to regulators and supervisors that can employ machine learning not only to predict default and systemic risk but also to have insights into what are the main drivers for these probabilities.

Future research could employ other machine learning methods, such as deep neural networks and other state-of-the-art algorithms, or even clustering algorithms to perform aggregate analyses. Other XAI methods could also be employed for comparison and robustness of results. If available, other variables could also be added to the sample.

**Disclaimer 1:** The authors declare that they do not have a financial interest or personal relationships that could influence the work carried out in this paper.

**Disclaimer 2:** The views expressed in this work are of entire responsibility of the authors and do not necessarily reflect those of their respective affiliated institutions nor those of its members.

## References

Acharya VV (2009) A theory of systemic risk and design of prudential bank regulation. Journal of Financial Stability 5(3):224–255, DOI 10.1016/j.jfs.2009.02.001

Akins B, Li L, Ng J, Rusticus TO (2016) Bank Competition and Financial Stability: Evidence from the Financial Crisis. Journal of Financial and Quantitative Analysis 51(1):1–28, DOI 10.1017/S0022109016000090

Altman EI (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance 23(4):589–609

Amidu M, Wolfe S (2013) Does bank competition and diversification lead to greater stability? Evidence from emerging markets. Review of Development Finance 3(3):152–166, DOI 10.1016/j.rdf.2013.08.002, URL http://dx.doi.org/10.1016/j.rdf.2013.08.002

Anginer D, Demirguc-Kunt A (2014) Has the global banking system become more fragile over time? Journal of Financial Stability 13:202–213, DOI 10.1016/j.jfs.2014.02.003, URL http://dx.doi.org/10.1016/j.jfs.2014.02.003

Anginer D, Demirguc-Kunt A, Zhu M (2014) How does competition affect bank systemic risk? Journal of Financial Intermediation 23(1):1–26, DOI 10.1016/j.jfi.2013.11.001, URL http://dx.doi.org/10.1016/j.jfi.2013.11.001

Beck T, De Jonghe O, Schepens G (2013) Bank competition and stability: Cross-country heterogeneity. Journal of Financial Intermediation 22(2), DOI 10.1016/j.jfi.2012.07.001

Beutel J, List S, von Schweinitz G (2019) Does machine learning help us predict banking crises? Journal of Financial Stability 45:100,693

Bharath ST, Shumway T (2008) Forecasting Default with the Merton Distance to Default Model. The Review of Financial Studies 21(3):1339–1369, DOI 10.1093/rfs/hhn044

Black F, Scholes M (1973) The pricing of options and corporate liabilities. The Journal of Political Economy URL http://www.jstor.org/stable/1831029

Breiman L (2001) Random forests. Machine learning 45(1):5–32

Breiman L (2002) Manual on setting up, using, and understanding random forests v3. 1. Statistics Department University of California Berkeley, CA, USA 1(58):3–42

Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. Routledge

Burns K, Moosa IA (2015) Enhancing the forecasting power of exchange rate models by introducing nonlinearity: Does it work? Economic Modelling 50:27–39, DOI 10.1016/j.econmod.2015.06.003

Bussmann N, Giudici P, Marinelli D, Papenbrock J (2021) Explainable machine learning in credit risk management. Computational Economics 57(1):203–216

Carvalho D, Ferreira MA, Matos P (2015) Lending relationships and the effect of bank distress: Evidence from the 2007–2009 financial crisis. Journal of Financial and Quantitative Analysis 50(6):1165–1197

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, KDD '16, p 785–794, DOI 10.1145/2939672.2939785, URL https://doi.org/10.1145/2939672.2939785

Cortes C, Vapnik V (1995) Support-vector networks. Machine learning 20(3):273–297

Croxson K, Bracke P, Jung C (2019) Explaining why the computer says 'no'. FCA 5:31

Dahiya S, Saunders A, Srinivasan A (2003) Financial distress and bank lending relationships. The Journal of Finance 58(1):375–399

Fama EF, French KR (1992) The cross-section of expected stock returns. the Journal of Finance 47(2):427–465

Fama EF, French KR (2015) A five-factor asset pricing model. Journal of Financial Economics 116(1):1 – 22, DOI https://doi.org/10.1016/j.jfineco.2014.10.010, URL http://www.sciencedirect.com/science/article/pii/S0304405X14002323

Feng G, Polson NG, Xu J (2018) Deep learning in characteristics-sorted factor models. arXiv preprint arXiv:180501104

Fisher A, Rudin C, Dominici F (2019) All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research 20(177):1–81

Fu XM, Lin YR, Molyneux P (2014) Bank competition and financial stability in Asia Pacific. Journal of Banking & Finance 38:64–77, DOI 10.1016/j.jbankfin.2013.09.012, URL http://dx.doi.org/10.1016/j.jbankfin.2013.09.012

Gan L, Wang H, Yang Z (2020) Machine learning solutions to challenges in finance: An application to the pricing of financial products. Technological Forecasting and Social Change 153:119,928

Genuer R, Poggi JM (2020) Random forests. In: Random Forests with R, Springer, pp 33–55

Gropp R, Hakenes H, Schnabel I (2011) Competition, risk-shifting, and public bail-out policies. Review of Financial Studies 24(6):2084–2120, DOI 10.1093/rfs/hhq114

Gruszczyński M (2020) Modeling financial distress and bankruptcy. In: Financial Microeconometrics, Springer, pp 77–119

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. The Review of Financial Studies 33(5):2223–2273

Hasan I, Marton K (2003) Development and efficiency of the banking sector in a transitional economy: Hungarian experience. Journal of Banking and Finance 27(12):2249–2271, DOI 10.1016/S0378-4266(02)00328-X

Hassani H, Huang X, Silva E (2018) Digitalisation and big data mining in banking. Big Data and Cognitive Computing 2(3):18

Höwer D (2016) The role of bank relationships when firms are financially distressed. Journal of Banking & Finance 65:59–75

Hsu MW, Lessmann S, Sung MC, Ma T, Johnson JE (2016) Bridging the divide in financial market forecasting: machine learners vs. financial economists. Expert Systems with Applications 61:215–234

Iwanicz-Drozdowska M, Ptak-Chmielewska A (2019) Prediction of banks distress–regional differences and macroeconomic conditions. Acta Universitatis Lodziensis Folia Oeconomica 6(345):73–57

Kozak S, Nagel S, Santosh S (2020) Shrinking the cross-section. Journal of Financial Economics 135(2):271–292

Leo M, Sharma S, Maddulety K (2019) Machine learning in banking risk management: A literature review. Risks 7(1):29

Lown CS, Osler CL, Strahan PE, Sufi A (2000) The Changing Landscape of the Financial Services Industry : What Lies Ahead ? FRBNY Economic Policy Review (October):39–55, DOI Article

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30:4765–4774

Mama HB (2017) Innovative efficiency and stock returns: Should we care about nonlinearity? Finance Research Letters

Merton RC (1974) On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The Journal of Finance 29:449–70

Milne A (2014) Distance to default and the financial crisis. Journal of Financial Stability 12:26–36, DOI 10.1016/j.jfs.2013.05.005, URL http://dx.doi.org/10.1016/j.jfs.2013.05.005

Müller AC, Guido S (2017) Introduction to Machine Learning with Python: A Guide for Data Scientists. DOI 10.1017/CBO9781107415324.004, arXiv:1011.1669v3

Newey WK, West KD (1987) A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Ecometrica 55(3):703–708, DOI 10.2307/1913610

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825–2830

Peng Y, Nagata MH (2020) An empirical overview of nonlinearity and overfitting in machine learning using covid-19 data. Chaos, Solitons & Fractals p 110055

Ribeiro MT, Singh S, Guestrin C (2016) " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144

Schiozer R, Mourad F, Vilarins RS (2018) Bank risk, bank bailouts and sovereign capacity during a financial crisis: a cross-country analysis. Journal of Credit Risk DOI https://doi.org/10.21314/jcr.2018.246

Shapley LS (1953) A value for n-person games. Contributions to the Theory of Games 2(28):307–317

Strumbelj E, Kononenko I (2010) An efficient explanation of individual classifications using game theory. The Journal of Machine Learning Research 11:1–18

Štrumbelj E, Kononenko I (2014) Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems 41(3):647–665

Tabak BM, Fazio DM, Cajueiro DO (2012) The relationship between banking market competition and risk-taking: Do size and capitalization matter? Journal of Banking & Finance 36(12):3366–3381, DOI 10.1016/j.jbankfin.2012.07.022, URL http://linkinghub.elsevier.com/retrieve/pii/S0378426612002026

Tabak BM, Fazio DM, Cajueiro DO (2013) Systemically important banks and financial stability: The case of Latin America. Journal of Banking & Finance 37:3855–3866, DOI 10.1016/j.jbankfin.2013.06.003

Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological) pp 267–288

Wang CW, Chiu WC, Peña JI (2017) Effect of rollover risk on default risk: Evidence from bank financing. International Review of Financial Analysis 54:130–143, DOI 10.1016/j.irfa.2016.09.009, URL http://dx.doi.org/10.1016/j.irfa.2016.09.009

Xiaomao X, Xudong Z, Yuanfang W (2019) A comparison of feature selection methodology for solving classification problems in finance. In: Journal of Physics: Conference Series, IOP Publishing, vol 1284, p 012026