The role of ESG in predicting bank financial distress: cross-country evidence[*]

Alberto Citterio[**]

Department of Economics, University of Insubria, Via Monte Generoso 71, Varese 21100, Italy

Abstract

Extant literature suggests that corporate social responsibility (CSR) has a potential impact on performance and reputation of companies. However, with reference to banks, the literature provides only limited evidence of the relationship between non-financial performance and riskiness. As a result, to the best of our knowledge, there are not studies that consider CSR as a predictor of bank's financial distress. We investigate this issue by analysing a sample of 362 commercial banks headquartered in the US and EU-28 members states, over the period from 2012 to 2019. We first assess a methodology to evaluate the association of bank financial distress and ESG-score, which we use as a proxy for CSR, both in its aggregate specification and for its each sub-component. We find a significant risk-reducing effect that is primarily explained by the social (S) component. After evaluating such relationship, we develop a model to detect financial distress that includes ESG-score as one of the predictors. Our findings show that the model can reach an Area Under the ROC Curve (AUC) higher than the 90%, suggesting an excellent ability to detect financial distress. In particular, we find that non-financial performance has a promising predictive ability, since it is even higher than other traditional accounting variables.

Keywords: Financial distress, Bank riskiness, Prediction models, CSR, ESG

JEL codes: G21, G33, M14, C53

1. Introduction

Default forecasting is a key issue in the managerial decision-making for financial institutions and policy makers (Caprio and Klingebiel, 2003). A business failure or, more in general, a financial distress, is usually anticipated by several warning symptoms. The understanding of such anticipating factors enables supervisor and managers to intervene promptly before the financial situation of an institution further deteriorates. When problematic banks are identified too late, substantial costs are associated with their resolution (Honohan and Klingebiel, 2000). In addition, because of the strong interconnectedness between banks, systemic bank failures have impact not only on bank's shareholders, but they can cause national and international contagion (Benston and Kaufman, 1995), which can affect financial market stability and the economic growth. If the detection of problems occurs in an early stage, regulatory action can be taken either to prevent a bank from failing or to minimize the cost to the authorities and thus to taxpayers (Thomson, 1991), ensuring a better allocation of governments and authorities' resources. Therefore, appropriate early warning models thus represent a useful tool both to identify and interpret such symptoms and to predict if a business will suffer financial distress through mathematical and statistical methods.

[*] This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[**] Corresponding author.

E-mail address: a.citterio2@uninsubria.it

During the last fifty years, and especially after the 2008 global financial crisis, an increasing number of early warning system models have been developed to detect and prevent severe bank financial distresses.
The analysis of these studies (among others, Adnan Aziz and Dar, 2006; Balcaen and Ooghe, 2006; Bellovary et al., 2007; Citterio, 2020) shows that prediction models have almost always been entirely based on accounting variables as independent explanatory variables. However, this characteristic entails significant limits, since accounting reports usually depict a firm's past performance and may not be informative for predicting the future (Zavgren, 2006). In addition, accounting information could reflect the recorded book value of an asset and not the "true" one (Agarwal and Taffler, 2008). Because of the above, several authors suggest the opportunity to include additional variables, such as macroeconomic indicators (e.g., Flannery, 1998; Jagtiani and Lemiuex, 2001) and non-financial information (e.g., Simpson and Gleason, 1999; Berger et al., 2016; Fernando et al., 2019) to improve the predictive accuracy of models. While macroeconomic variables and market structural information have been tested in several studies (e.g., Arena, 2008; Männasoo and Mayes, 2009; Chiaramonte and Casu, 2017), there is only limited evidence of the influence of non-financial performance on a bank's probability to be in financial distress. This gap is particularly relevant, since sustainable behaviors have become one of the most pressing issues for society, policy makers and stakeholders for several decades now (e.g., Engle, 2007; Friede et al., 2015). The need to meet sustainable goal has been primarily supported by international bodies like United Nations, which has recently supported the adoption of the United Sustainable Development Agenda 2030 and the Paris Agreement (2015) signed by 193 countries. The plan includes a shared commitment to achieve 169 environmental, economic, social, and institutional targets by 2030. Within the banking sector, the loss in trust and credibility suffered during the 2008 global financial crisis induced banks to adopt social responsibility actions and increase their transparency and compliance to sustainability standards and guidelines (Cornett et al., 2016; Miralles-Quirós et al., 2019), also limited by a more stringent regulatory approach. In this context, the consensus about the ability of the banking sector to promote sustainable behaviors is rapidly grown. In particular, while the banks' direct contribution in gaining environmental and social goals may be modest, financial intermediaries could play a pivotal role on the sustainability of other industries through the lending channel (Scholtens and Van't Klooster, 2019; Beck et al., 2010): banks can select investment projects and orient funds according to the non-financial results of the target companies, with the effect to promote sustainable approaches among their borrowers (Thompson & Cowton, 2004; García-Sánchez and García-Meca, 2017; Kemp-Benedict, 2018). Moreover, the role of corporate social responsibility now represents not only a matter of ethical behavior but also a problem of prudential supervision. Several regulatory initiatives (e.g., OECD Guidelines for Multinational Enterprises; Directive 2014/05/EU; Action Plan on Sustainable Finance) have been indeed promoted to strengthen the transparency of sustainability disclosure, to develop responsible approaches, and to foster banks to identify tools and approaches to measure and monitor risks connected to environmental, social, and governance dimensions.
Despite the growing interest in this issue among media, regulators and academics, linkage between CSR and default risk of banks has been not yet fully understood. To fill this gap, we first examine this relationship by investigating their correlation. Once we identify and understand the impact of bank's sustainable initiatives on bank's riskiness, we test the goodness of ESG performance as an early warning signal in a bank' financial distress prediction model. We use a highly representative dataset composed of an unbalanced panel of 362 banks headquartered in EU-28 and USA, covering the period 2012-2019. To evaluate non-financial performance, we collect the Thomson Reuters ESG-scores, which enable us to analyse the relationship both at an aggregate level and for each sub-component.
The paper contributes to the literature in several ways. First, we seek to assess whether non-financial performance has a significant impact on banks' risks. In addition, we conduct a panel data analysis to evaluate the specific contribution of each pillar within the ESG score (environmental, social, and governance) individually. This allows us to better understand the specific drivers of the risk-reduction. Second, to the best of our knowledge, it is the first study to present a prediction model which include Corporate social performance (CSP) as a predictor variable. Thus, we can identify the contribution of sustainable initiative in

discriminating between healthy and financially distressed banks. In this context, we also provide a contribution in the methodological approach by means of a comparison between traditional statistical technique (Linear Discriminant analysis and Logistic regressions), machine learning approaches (Decision trees and Support vector machine), and ensemble methods (Random forest and XgBoost). This allows us to evaluate and discuss the widely held view that, on average, ensemble classifier could outperform individual techniques (e.g., Kumar and Ravi, 2007; Demyanyk and Hasan, 2010).

Our analysis shows that the overall CSR performance has a significant effect in reducing riskiness. Furthermore, we demonstrate that each pillar has a positive risk-reducing effect: it should be note that, while social performance has a significant positive effect, environmental and governance pillars have a weaker positive contribution. The second part of the analysis shows that ESG scores has a promising ability to predict financial distress in combination with accounting and macroeconomic variables. These findings are of particular interest to both academics, bank management, and financial regulators as they contribute to the debate on the relevance of non-financial performance in promoting bank stability.

The remainder of this paper is organized as follows. Section 2 presents a brief overview of main empirical contributions in bankruptcy prediction modelling and examines the potential role of non-financial performance as an early warning signal. Consequently, we develop our research hypothesis. Section 3 provides a description of the sample, the dataset, and the methodology applied. Section 4 presents and discusses the results. Finally, Section 5 summarises the main insights and concludes.

2. Literature review

This paper relates to two different strands of literature: first, it relates to studies that predict failure or financial distress of banking institutions; and second, it contributes to the literature evaluating the mechanism according to which non-financial performance could influence banks performance and stability.

A large stream of research focuses on the determinants and predictability of bank failures. The aim of such studies is to develop an early warning system model that can identify distressed banks prior to failure or a severe financial distress in order to allow managers and/or supervisors to take the necessary preventive or remedial actions.

Literature on bankruptcy prediction historically originated from the Uniform Financial Institutions Rating System (UFIRS), also known as CAMEL rating, introduced by the US regulator in 1979 (Reidhill and O'Keefe, 1997). CAMEL rating is an internal supervisory tool that assesses banks condition measuring Capital adequacy, Asset quality, Management quality, Earnings, and Liquidity through balance sheet indicators. In 1997 an indicator of "Sensitivity to market risk" was adopted as a sixth component.

A wide range of studies applied the CAMEL framework to the analysis of bank failures, focusing on a variety of accounting variables. These studies identify a set of recurring bank failure predictors: financial distress probability increases for banks with low capitalization (e.g., Andersen, 2008, Betz et al., 2014), low asset quality (e.g., Berger et al., 2016; Chiaramonte et al., 2016), high concentration of business or commercial real estate loans (e.g., Cole and White, 2012; Jordan et al., 2010), cost inefficiency (e.g., Mayes and Stremmel, 2014), low profitability (e.g., Poghosyan and Čihak, 2011; Serrano-Cina and Gutierrez-Nieto, 2012), low liquidity (e.g., Arena, 2008; Cipollini and Fiordelisi, 2012), high reliance on non-core deposit funding (e.g., Hong et al., 2014; Altunbas et al., 2015) and other measures of poor performance.

Hence, most papers analysing individual bank failures or distress events essentially focused on accounting variable (Balcaen and Ooghe, 2006), although over time studies suggested to introduce non-traditional or off-balance activities to improve the model accuracy (e.g., Flannery, 1998; Simpson and Gleason, 1999; Campbell et al., 2008; Berger et al., 2016).

Consequently, a number of papers, led by the study of Flannery (1998), included stock market information as signal of potential adverse market perception of the health of the observed institutions (e.g., Bongini et al., 2002; Distinguin, 2006; Avkiran and Cai, 2012). However, these approaches are only applicable for publicly listed banks. In addition, market-based information tends to have a short horizon (Betz et al., 2014) and it strongly relies on the market efficiency assumption, according to which markets are always able to reflect all available information (Fama, 1970). Furthermore, Campbell et al. (2011) demonstrate that market information has only little forecasting power after controlling for other variables.

Moreover, recent empirical analysis has shown that macroeconomic indicators have often predated banking crises (e.g., Evens et al., 2000; Halling and Hayden, 2006; Arena, 2008; Aubuchon and Wheelock, 2010). There has been a growing awareness of the relevance of the macroeconomic context and the health of the financial system to the performance of individual banks indicators, with macroeconomic indicators employed to assess the cyclicality of the economic system and thus to legitimate certain inefficiencies exhibited by firms. Aubuchon and Wheelock (2010) analyse the existence of regional patterns in bank failure in the United States during the period 2007-2010. The results show that bank failures were higher in states with the largest reduction in personal income and gross state product and the largest increases in unemployment rates. According to their study, banks health is highly vulnerable to local economic shocks. Similarly, Arena (2008) argues that CAMEL variables, though they significantly affect probability of failure, are not sufficient to explain cross-country differences. Studying East Asian and Latin American case, the author highlights that banking system structure and macroeconomic variables contribute to explain the likelihood of failure.

Turning to the second strand of literature to which this study is related, the distress of the banking system during the 2008 financial crisis has sparked a debate about the need for a new sustainable behavior (Weber & Remer, 2011, Hurley et al., 2014). However, despite the increasing attention on such aspect, there is still not a homogeneous definition of CSR (Sheehy, 2015). Among practitioners, the non-financial performance of a firm is commonly measured by the so-called ESG factors, i.e., environmental, social and governance factors (Friede et al., 2015).

Starting with the study of Simpson and Gleason (1999), several studies have paid particular attention to the role of corporate governance dimension as a non-financial predictor of bank's financial distress (e.g., Switzer and Wang, 2013; Berger et al., 2016; Switzer et al., 2018). Switzer and Wang (2013) show that banks with larger and more independent boards are associated with lower credit risk level. Similarly, Berger et al. (2016) demonstrate that higher shareholding of lower-level managers and non-CEO higher level managers is positively associated with default risk. The more recent evidence is provided by Switzer et al. (2018), which find that institutional ownership and board independence reduce the risk of being in financial distress, while insider ownership, CEO duality and board size have the opposite effect. Overall, despite the heterogeneity of such variables, there exists some evidences concerning the role of corporate governance in explaining the probability of failure. These results are also confirmed in the wider range of studies on non-financial firms[1].

Despite the number of contributions on corporate governance in banking, almost no study has empirically investigated how environmental and social factors might influence risk of default. Because of the above, we base our hypothesis of a positive contribution of non-financial performance in reducing the risk of financial distress borrowing from the studies which evaluate the relationship between Corporate social performance (CSP) and Corporate financial performance (CFP) or reputation. There is indeed a growing number of papers that have empirically investigated how Corporate social performance influences Corporate financial performance. Though empirical research has found a weak relation between CSR and company performance (Soana, 2011) and, sometimes, not significant at all (Chih et al., 2010, Saxena and Kohli, 2012), some interesting conclusions emerge. Literature suggests that ESG performance could influence financial performance both in a direct and indirect channel. Directly, non-financial performance can increase financial returns or market value (Margolis and Walsh, 2001; Friede et al., 2015) or reduce the cost of capital (Bassen et al. 2006, Ciciretti et al., 2014). The relationship is influenced indirectly through an increase in reputation (Branco and Rodrigues, 2006): firms that serve the implicit claims of stakeholder could enhance company reputation (Makni et al., 2009). A higher reputation allows to improve bank's image in the market (Carreras et al., 2013), recruit and retain the best talents (Glavas and Kelley, 2014), and attract more creditworthy

[1] For a comprehensive review of the previous papers the reader can refer to Fernando et al. (2019)

borrowers (Linthicum et al., 2019). All these benefits can contribute in the medium term to reach higher profit, better asset quality and, accordingly, higher financial conditions (Cheng et al., 2014). An increase in reputation thus could entail a significant reduction in riskiness (Krueger et al., 2010). It is also worth considering that the hypothesised inconclusiveness of similar works (e.g., Margolis et al., 2009; Weber, 2014) strongly depends on important aspects that make it difficult to compare existing studies. For example, Jo et al., (2015) show that the differential recognition of environmental problems by executives around the world affects the impact of non-financial performance. In detail, customers in Europe and North America tend to react more positively to environmental management than those in the Asia Pacific region. Similarly, García-Sánchez and García-Meca (2017), testing a sample of 159 bank from 9 countries for the period 2004-2010, show that bank's commitment to CSR practices enhance earnings in countries with higher levels of investor protection and bank regulation, suggesting that more socially responsible banks are more valuable in a stricter regulatory environment.

Another aspect that should be taken into consideration is the choice of the time window: several papers that control for the 2008 financial crisis highlight that downturn periods might affect the validity of the above-mentioned relationship. Esteban-Sanchez et al. (2017) show that the recent global financial crisis had a negative interaction effect on the CSP-CFP relationship. Likewise, Forcadell and Aracil (2017) find that reputation for sustainability strategies do not improved returns during the period 2008-2013. In addition, Miralles-Quirós et al. (2019) highlight that financial stakeholder start to give more value to ESG performance after the financial crisis.

Finally, studies have used several different CSP measure, which strongly limited the comparability among studies. While some authors use information from annual reports (e.g., Carnevale and Mazzucca, 2014; Jo et al., 2015; Nobanee and Ellili, 2016) or surveys (e.g., Saxena and Kohli, 2012), other authors strongly rely on agency ratings (e.g., Soana, 2011; Cornett et al., 2014; Dell'Atti et al., 2017, Esteban-Sanchez et al., 2017). In addition, while some authors focus on comprehensive ESG scores, other works evaluate the impact on financial profitability of each sub-component or only for some specific dimensions. The identified discordant effects of the three dimensions on reputation and performance (Dell'Atti et al., 2017; Esteban-Sanchez et al., 2017) justify our choice to evaluate the relationship both at an aggregate and disaggregate level.

This concise review of the literature demonstrates that the study on the relationship between CSP and individual bank performance and stability is still developing. Starting from the existing literature, we take a further step and extend the underlying assumptions to investigate whether non-financial performance contribute to reduce bank's probability of financial distress and whether they should be considered in the construction of an early warning system.

3 Data and methodology

3.1 Data

The sample of this study consists of all banks headquartered in the USA and in the EU-28 member states, during the period 2012-2019. The decision to not cover the period prior to 2012 aims to neutralize the potential influence in our analysis of the 2008 financial crisis on banks' performance and probability of default. As shown above, papers that control for the global financial crisis period (e.g., Esteban-Sanchez et al., 2017; Forcadell and Aracil, 2017; Mirelles-Quirós et al., 2019) show that the positive effect of sustainable initiatives on banks performance were reduced or wiped out. It is reasonable to infer that the same effect could occur in the relation between CSR and probability of default.

We first collect financial data from consolidated yearly financial statements obtained from Orbis Bank Focus database. We use yearly data instead of quarterly data to minimize the influence of seasonal effects on bank financial indicators. We then obtain macroeconomic data from World Bank database to control for possible country-effects. Financial and macroeconomic data have been then matched with ESG data collected from Thomson Reuters' Eikon, which has been extensively used in previous CSR studies (Dell'Atti et al., 2017;

Gangi et al., 2018; Neitzert and Petras, 2020, among others). We limited the analysis to all banks for which we collect financial and ESG information for at least two consecutive years within the reference period. Due to data limitation, the final sample consists of 362 commercial banks headquartered in 19 countries (for an average representativity in terms of total asset of 88% and 85% of US and EU banking sector respectively), for a total of 1611 firm-year observation.

3.2 Variables definition

3.2.1 Dependent variables: probability of default

The literature on bank failures shows a highly fragmentation, especially with respect to the definition of bank failure[2]. In particular, financial distress could have different degrees of severity: mild financial distress may just be temporary, while serious financial distress corresponds to business failure or bankruptcy (Sun et al., 2014). Consequently, authors define in different ways the failure of a business in their studies.
The juridical definition of bankruptcy, although it allows to classify banks with an objective criterion (Charitou et al., 2004), is characterized by an extremely low frequency rate, which it is even more rare within the banking sector. We thus decide to estimate financial distress focusing on accounting-based risk measures. We approximate bank risk using the Z-score, which has been widely used in previous studies (Boyd and Graham, 1986; Boyd et al., 2006; Leave and Levine, 2009; Delis et al., 2012; De Young and Torna, 2013). Despite the extensive use of this measure, there is a lack of consensus on a standard way to construct time-varying z-score (Li and Malone 2016). We therefore compute this measure following different approaches. The first approach (labelled Z-score$_1$) is defined as:

$$Z\text{-}score_1 = \frac{ROA + (Equity/Total\ Assets)}{\sigma(ROA)}$$

To test the robustness of our results, we then computed Z-score using other two approaches (hereafter, respectively, Z-score$_2$ and Z-score$_3$):

$$Z\text{-}score_2 = \frac{mean(ROA) + mean(Equity/Total\ Assets)}{\sigma(ROA)}$$

$$Z\text{-}score_3 = \frac{mean(ROA) + mean(Equity/Total\ Assets)}{\max\limits_{T_{-5}<t<T}(ROA_t) - \min\limits_{T_{-5}<t<T}(ROA_t)}$$

We calculate the standard deviation of ROA and each average for rolling time windows of five years, as computed in similar studies (e.g., Neitzert and Petras, 2020). Z-score measures the distance from insolvency connecting the level of the capital with returns (Roy, 1952), and it can be interpreted as the quantity of variability of returns the capital could absorb before the bank become insolvent. Z-score is then inversely proportional of the probability of insolvency and higher value indicate that the bank is more stable. To reduce the skewness of the z-score we use the natural logarithm, which is normally distributed (hereafter we consider each Z-score as the natural logarithm of the Z-scores). We then avoid the problem of negative z-scores adding for each approach a specific constant *k* to increase the minimum value to 0.001. We therefore compute:

$$\min\,(Z - score_i) + k > 0.001$$

Finally, we also confirm our results using the Merton Distance-to Default (DD) measure, where default occurs when the market value of a firm's assets falls below the face value of its liabilities. The DD measure, which boast wide range of applications (e.g., Vassalou and Xing, 2004; Duffie et al., 2007, Bharath and Shumway, 2008; Anginer and Demirguc-Kunt, 2004), has been demonstrated to outperform other market-

[2] For a comprehensive review of the previous papers the reader can refer to Balcaen and Ooghe (2006) and Citterio (2020).

based indicators in terms of prediction of bank default (Gropp et al., 2006). In the context of the DD model, the default risk is calculated as the number of standard deviations by which the market value of a bank assets needs to fall to reach the default point. It implies that higher level of DD means a healthier bank. We express the DD for each bank in the sample at the end of each year *t* as:

$$DD_t = \frac{\ln\,(V_{A,t}/X_t) + (r_f - 0.5\sigma_{A,t}^2)T}{\sigma_{A,t}\sqrt{T}}$$

where $V_{A,t}$ is the market value of assets at the end of the fiscal year *t,* X_t is the book value of the bank's liabilities, r_f is the risk-free rate, which is approximated – as made in Vallascas and Hagendorff (2013) - to the yield on one-year U.S. treasury bills for U.S. banks and the Euribor rate with a maturity of 12 month for European banks, $\sigma_{A,t}$ is the annualized asset volatility, and *T* is the horizon over which we predict the probability of failure and it is set equal to 1. The computation of $V_{A,t}$ and $\sigma_{A,t}$, which are not directly observable, relies on an iterative process based on the Black and Scholes (1973) pricing model. In detail, the model requires the resolution of a system of nonlinear equation, where the market value of the bank is expressed as a function of the asset value:

$$V_{E,t} = V_{A,t}N\left(d_{1,t}\right) - X_t e^{-r_f T} N(d_{2,t})$$

$$\sigma_{E,t} = \left(\frac{V_{A,t}}{V_{E,t}}\right) N(d_{1,t})\sigma_{A,t}$$

where

$$d_{1,t} = \frac{\ln\left(V_{A,t}/X_t\right) + \left(r_f + 0.5\sigma_{A,t}^2\right)T}{\sigma_{A,t}\sqrt{T}}$$

$$d_{2,t} = d_{1,t} - \sigma_{A,t}\sqrt{T}$$

The resolution of the system, which is based on the computation of a Newton search algorithm that we have performed using the Matlab software, requires a starting value for the annualised asset volatility, that is defined - by default - as the volatility of equity multiplied by the ratio of the market value of equity to the sum of the market value of equity and the book value of total liabilities, i.e.:

$$\sigma_{A,t} = \sigma_{E,t}\, V_{E,t}/\left(V_{E,t} + X_t\right)$$

The DD measure may be calculated only for listed banks. Due to this limitation, we perform the distance to default for a subset of the main sample composed by 338 listed banks, for an unbalanced panel of 1536 firm-year observations. Data were collected from Thomson Reuters' Eikon database. A synthetic summary of dependent variables is included in table 1.

3.2.2. Independent variables: ESG scores

In line with previous CSR studies, we approximate the environmental, social, and corporate governance performance using the Thomson Reuters ESG-score (e.g., Dell'Atti et al., 2017; Esteban-Sanchez et al., 2017; Gangi et al., 2018; Miralles-Quirós et al., 2019). Thomson Reuters performs a weighted analysis of 178 performance indicators built over approximately 400 different data points collected from publicly available reports (e.g., company websites, annual reports, non-financial reports), covering over 6000 companies since 2002. The underlying measures are then grouped into 10 categories, which are combined to reflect the company's ESG performance (Figure 1). The comprehensive ESG-score is therefore a weighted average of Environmental scores (33%), Social scores (35.5%) and Governance scores (30.5%). Thomson Reuters provides an annual 0-100 score both at aggregate level as well as for each sub-component, which will allow us to consider separately the overall effect of the comprehensive score and the influence of each component in our regression model.

Figure 1: Thomson Reuters' ESG-score composition

Source: author's re-elaboration of Thomson Reuters (2017)

3.2.3 Control variables

We have included two categories of indicators in order to consider multiple variable to predict bank financial deterioration. Firstly, we have included indicators from banks' income statements and balance sheet, which provides information about the symptoms of incipient crisis (Sinkey, 1975). In line with the extant literature, we have focused on the traditional CAMEL indicators that capture credit, operational and liquidity risk and that have found to be highly correlated with distress (Tam and Kiang, 1992; Bongini et al., 2001; Poghosyan and Cihak, 2011). Capital adequacy (C) is proxied by the Equity-to-assets ratio (ETA), which has been identified in the literature as a good predictor of failure (e.g., Pille and Paradi, 2002; Andersen, 2008). ETA has the advantage to be available across all the sample and to be not influenced by manipulation and arbitrariness that conversely affect risk-weighted measures. Capital serves as a buffer for unexpected financial losses and it is expected to be negatively correlated with probability of default and consequently positively correlated with Z-score and DD measure. The paper uses two indicators to measure asset quality (A). First, we identify the ratio of total loans to total assets (TLTA), which provides an identification of the overall involvement in traditional lending activities and is expected to increase the risk of bank failure. A second variable capturing a different aspect of the asset quality is the level of nonperforming loans (NPL), defined by the amount of nonperforming loans to gross loans. A higher level of the NPL ratio indicates lower quality of the loan portfolio since non-performing assets are likely to become losses in a subsequent period. In line with other studies, we proxied the Management efficiency (M) through the cost-to-income ratio (CIR). Since the CIR reflects the cost efficiency, it is expected a positive relationship with the probability of default. Next, we include the Return on average assets (ROA) as a proxy for bank earnings (E). ROA is the most frequently used measure of profitability and we expect a negative sign for the relation with financial distress since an increase in profitability should reduce the probability of failure. Liquidity (L) is proxied by the customer deposit to total asset ratio (CDTA), which identifies the portion of assets that is financed by retail funding. Given that retail deposits tend to be a more stable source than the interbank market or securities funding, especially in periods of crisis (Shleifer and Vishny, 2010; Betz et al., 2014; Altunbas et al., 2015), we expect a negative relationship with probability of default. In addition to the abovementioned accounting indicators, we include a variable that capture the degree of income diversification. Following Stiroh (2004), we incorporate the Non-interest income to total operating revenues ratio (NIOR), which is expected to be negatively related with probability of failure because diversification may lead to risk reduction.

Table 1: Description of Variables and data sources

Variables	Description	Source
Dependent variables:		
Z-score$_1$	The natural logarithm of the sum of the current value of return on average assets and the current value of equity to total assets over the five-year standard deviation of return on average assets	BankFocus
Z-score$_2$	The natural logarithm of the sum of the five-year moving average of return on average assets and the five-year moving average of equity to total assets over the five-year standard deviation of return on averages assets	BankFocus
Z-score$_3$	The natural logarithm of the sum of the five-year moving average of return on average assets and the five-year moving average of equity to total assets over the range between maximum and minimum return on average assets over previous 5 years	BankFocus
DD measure	The sum of the current market value of assets over the book value of liabilities and the difference between the risk-free rate and 0.5 multiplied by the squared asset volatility over the asset volatility	Thomson Reuters
Independent variables:		
ESG	Measure of the overall corporate social responsibility	Thomson Reuters
ENV	Measure of bank's environmental performance	Thomson Reuters
SOC	Measure of bank's social performance	Thomson Reuters
GOV	Measure of bank's governance practices	Thomson Reuters
Control variables (Accounting variables)		
ETA	The ratio of total equity to total assets	BankFocus
TLTA	The ratio of Total loans to total assets	BankFocus
NPL	The ratio of Non-performing loans to gross loans	BankFocus
CIR	The ratio of Operating expenses to operating income	BankFocus
ROA	The ratio of Net income before taxes to average assets	BankFocus
CDTA	The ratio of customer deposits to total assets	BankFocus
NIOR	The ratio of Non-interest income to net operating revenue	BankFocus
Control variables (Macroeconomic variables)		
GDP	Annual growth rate of real gross domestic product	World Bank
INF	Inflation rate	World Bank
HHI	Sum of the squared market share of each bank competing in a country (in terms of total assets)	BankFocus

Secondly, we include macroeconomic variables and a measure of market concentration. Several authors indeed highlight the opportunity to include macroeconomic indicators (e.g., Flannery, 1998; Jagtian and Lemieux, 2001) to control for macroeconomic imbalances. We include the annual GDP growth (GDP) and the annual inflation rate (INF), provided by the World Bank Database. Literature suggests a negative relationship between probability of distress and GDP (e.g., Arena, 2008; Mãnnasoo and Mayes, 2009) and a positive sign for the inflation rate (e.g., Chiaramonte et al., 2016). Lastly, we compute the Herfindahl-

Hirschman index (HHI) as a measure of banking system concentration. The HHI is defined as the sum of the squared market share of each bank competing in a country. We collect the share market (in term of total assets) for each bank from BankFocus. The resulting score range from 0 to 1, which we rescale from 0 to 100 for homogeneity with the other indicators. There is no consensus about the link between HHI to probability of default. While some authors support the "concentration-fragility view" (Boyd and De Nicolo, 2005; Poghosyan and Čihak, 2011), according to which banks located in more concentrated banking sector are found to be more vulnerable, others defend the "concentration-stability" theory (Allen and Gale, 2000; Chiaramonte and Casu, 2017), which suggest a negative relationship between market concentration and probability of default.
To reduce the potential influence of outliers and single erroneous data points, data are winsorised at the 1^{st} and 99^{th} percentile. Following similar studies (e.g., Neitzert and Petras, 2020), we exclude macro-economic data and ESG scores from the winsorisation process, because they are subject to several check by their respective data providers. Table 1 contains a description of all variables used.

3.3 Descriptive statistics

Table 2 shows the descriptive statistics for our risk measures, dependent variables, and control variables. The mean Z-score$_1$ is 3.97, and the mean DD measure is 5.63, while the standard deviations are 1.02 and 2.69, respectively. Mean value of our dependent variables are in line with similar research on bank default risks (for the Z-score see Chiaramonte et al., 2016; Liu et al., 2013; Beck et al., 2013. For the DD model see Vallascas and Hagendorf, 2013; Bhagat et al., 2015).

Table 2: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent variables:					
Z-score$_1$	1,611	3.97	1.02	-6.90	6.84
Z-score$_2$	1,611	3.86	1.12	-2.38	6.85
Z-score$_3$	1,611	2.85	1.12	-3.45	5.86
DD	1,536	5.63	2.69	-1.24	27.44
Independent variables:					
ESG	1,611	42.87	19.73	3.96	95.01
ENV	1,611	25.04	33.38	0.00	97.44
SOC	1,611	43.80	21.30	2.42	97.32
GOV	1,611	49.04	21.64	2.21	97.16
Control variables (Accounting variables)					
ETA (%)	1,611	10.67	5.41	-3.30	73.42
TLTA (%)	1,611	62.20	20.18	0.07	111.13
NPL (%)	1,611	4.12	7.55	0.00	95.04
CIR (%)	1,611	61.63	27.64	-525.33	203.69
ROA (%)	1,611	0.91	1.33	-7.15	16.54
CDTA (%)	1,611	63.89	20.29	0.09	94.59
NIOR (%)	1,611	35.85	24.84	-20.28	319.51
Control variables (Macroeconomic variables)					
GDP (%)	1,611	2.05	1.83	-9.13	25.16
INF (%)	1,611	1.58	1.01	-1.73	5.65
HHI	1,611	5.95	3.85	2.45	38.80

This table shows the descriptive statistics (mean, maximum, minimum and standard deviation) of dependent variables, independent variables and control variables. The statistics are based on the non-winsorized data.

Regarding the summary statistics of ESG measures, we observe that, on average, the governance pillar score (GOV) is higher with respect to other two components of the comprehensive ESG measure. On average, the banks in the sample have scored 25.04 points out of 100 for environmental performance, 43.80 for social performance, and 49.04 for corporate governance performance. This result is in line with the expectations, since strict regulatory initiatives have been put into place, especially after the 2008 global financial crisis, to improve the governance of the financial sector, whose impact on bank performance and on risk-taking have been extensively studied in the literature (e.g., Leaven and Levine, 2009, Beltratti and Stulz, 2012; Switzer et al., 2018), although with mixed results

Finally, Table 3 provides the matrix of correlation across variables. We decide to introduce ESG performance individually in our model due the high correlation among them. The remaining magnitudes are conversely in general low.

3.4 Methodology

The methodology described in this section consists of two parts. First, we evaluate the relation between default risk and corporate social responsibility performance in order to assess the goodness of ESG scores as early-warning signals. Second, we present and compare the prediction performances of different models.

1. Panel regression model

In order to provide a preliminary empirical evidence of the relation between bank failure and level of ESG performance we estimate a series of unbalanced panel data regressions, defined as follows:

$$DRM_{i,t} = \beta_0 + \beta_1 CSR_{i,t-1} + \beta_2 BS_{i,t-1} + \beta_3 MV_{j,t-1} + \mu_i + \varepsilon_{i,t-1}$$

where $DRM_{i,t}$ are the default risk measures at time *t* as defined in section 4.2.1, CSR measures the sustainability performances at time *t*-1 that are approximated by the ESG-score and the three pillars that we add separately one at a time, $BS_{i,t-1}$ represents the bank specific control variables, $MV_{i,t-1}$ describes the macroeconomic variables, μ_i is the unobserved time-invariant individual effect and $\varepsilon_{i,t-1}$ is the error term. The indices *i, j, t* represents, respectively, the bank, the country, and the fiscal year. We use one year lagged independent variables to reduce the risk of endogeneity issues caused by potential reverse causality or simultaneity bias.

We firstly run the Breusch-Pagan test to assess the existence of individual unobserved heterogeneity and the appropriateness of the pooled ordinary least square (OLS) model. The results suggest the existence of individual-specific effects, which indicate that the pooled OLS would conduct to biased estimates. We then run the Hausman test to verify whether the unobserved individual effects are correlated with time varying regressors. The Hausman test confirms the abovementioned correlation, which highlight that random effect would lead to inconsistent estimation. We therefore always apply fixed effect model to test our hypothesis.

2. Financial distress prediction and machine learning approach

The second purpose of this analysis is to integrate non-financial performance indicators in a proper early warning system and evaluate their gain contribution in the models. Most of the studies which propose a bank default prediction model use as a dependent variable a dummy variable that takes the value of one when bank *i* experiences a financial distress in period *t* and zero otherwise. We therefore start from the definition of Z-score_1 to create such a 0-1 dependent variable. Since it does not exist a commonly accepted threshold that distinguishes between healthy and distressed banks, we systematically run each model several time to test the predictive ability using different thresholds. More specifically, we start considering financial distressed the observations falling below the 5^{th} percentile of the empirical probability distribution of the Z-score. We then run the model using the 10^{th} percentile and continue until we reach the 95^{th} percentile. In each session, if the observation falls in the respective percentile, it takes value one and zero otherwise.

Table 3: Correlations

	Z-score$_1$	Z-score$_2$	Z-score$_3$	DD	ESG	ENV	SOC	GOV	EQTA	TLTA	NPL	CIR	ROA	CDTA	NIOR	GDP	INF	HHI
Z-score$_1$	1.0000																	
Z-score$_2$	0.9774	1.0000																
Z-score$_3$	0.9741	0.9975	1.0000															
DD	0.3590	0.3486	0.3457	1.0000														
ESG	-0.1813	-0.1907	-0.1947	-0.2703	1.0000													
ENV	-0.2503	-0.2614	-0.2648	-0.3322	0.8492	1.0000												
SOC	-0.2205	-0.2239	-0.2279	-0.2580	0.9231	0.8206	1.0000											
GOV	-0.0028	-0.0154	-0.0181	-0.1271	0.7643	0.4382	0.4948	1.0000										
EQTA	0.1348	0.1602	0.1631	0.3006	-0.3093	-0.3756	-0.3071	-0.1429	1.0000									
TLTA	0.0722	0.0507	0.0556	0.0514	-0.3520	-0.4004	-0.3624	-0.1746	-0.0148	1.0000								
NPL	-0.5015	-0.5398	-0.5373	-0.3540	0.1924	0.2785	0.2060	0.0332	-0.0689	0.1153	1.0000							
CIR	-0.1890	-0.1816	-0.1831	-0.0127	0.0541	0.0844	0.1103	-0.0528	0.0242	-0.2920	0.0847	1.0000						
ROA	0.2177	0.2393	0.2390	0.2423	-0.1527	-0.2321	-0.1603	-0.0303	0.6986	-0.0713	-0.2277	-0.1973	1.0000					
CDTA	0.2881	0.2797	0.2848	0.2232	-0.4549	-0.5465	-0.4655	-0.1994	0.1265	0.4576	-0.2306	-0.1555	0.0323	1.0000				
NIOR	-0.1814	-0.1479	-0.1527	-0.0652	0.3231	0.3598	0.3535	0.1376	0.0967	-0.6663	0.0001	0.3241	0.1215	-0.4863	1.0000			
GDP	0.2091	0.2333	0.2361	0.0046	-0.1442	-0.2030	-0.1295	-0.0609	0.1641	0.0391	-0.1609	0.0124	0.1952	0.2226	-0.0873	1.0000		
INF	-0.0155	0.0193	0.0215	0.1580	-0.0912	-0.1244	-0.0830	-0.0454	0.0843	-0.0155	-0.2570	-0.0104	0.1119	0.0556	-0.0262	-0.0931	1.0000	
HHI	-0.2317	-0.2781	-0.2806	-0.2174	0.2436	0.3242	0.1897	0.1585	-0.1950	-0.0264	0.3830	-0.1032	-0.1679	-0.2722	0.0574	-0.1655	-0.1649	1.0000

This table provides the correlation matrix of our dependent, independent and control variables.

The underlying model of the predictions that we will test is therefore described as follows:

$$D_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where D_i = 1 if the observation is below the n^{th} percentile of the empirical distribution of the Z-score$_1$ and D_i = 0 otherwise; X_i represents the vector of one-year lagged characteristics (see Table 2 for the full list of variables); and ε_i is the normally distributed error term with zero mean.

The model is estimated using different approaches. Despite the widespread opinion that – on average – ensemble classifier could outperform the individual techniques (e.g., Kumar and Ravi, 2007; Demyanyk and Hasan, 2010) since they allows to integrate several predictions, there is no unanimous consensus on which model outperform the others (e.g. Kimmel et al., 2016). In particular, while intelligent techniques usually require less assumptions and they allow to approximate nonlinear functions, the determination of parameters is often complex and arbitrary. It should also be considered that the configuration and elaboration of technically sophisticated methods are often very time consuming and the interpretation of the contribution of individual variables is sometimes quite complex. For these reasons, we decide to estimate our model using statistical methods, AI methods and ensemble methods to compare the predictive results of different techniques and to guarantee the robustness of our results. Specifically, we choose models widely used in literature: Logit (e.g. Bongini et al., 2002; Canbas et al., 2005; Betz et al., 2014; Momparler et al., 2016; Ekinci and Erdal, 2017; Poghosyan and Čihak, 2011; Cole and White, 2012) and Linear discriminant analysis (e.g Serrano-Cinca and Gutiérrez-Nieto, 2012; Cox and Wang, 2014; Le and Viviani, 2018) as statistical methods, Classification trees (e.g Ekinci and Erdal, 2017; Bräuning et al., 2019) and Support vector machine (e.g Boyacioglu et al.,2009; Ecer, 2013; Papadimitriou et al., 2013; Gogas et al., 2018; Jing and Fang, 2018) as AI methods and Random forests (e.g Iturriaga and Sanz, 2015; Carmona et al., 2019; Shrivastava et al., 2020) and XgBoost (e.g Carmona et al., 2019) as ensemble methods. In our applications the initial sample of banks is split into two subsamples: 70% of the sample is used as a training set for identification purposes and 30% of the sample is used to validate the model.

The prediction model is then used to predict financial distress as follows:

$$\widehat{D}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

where $\widehat{D}_i$ represents the expected probability of being in financial distress given the set of X characteristics. The classification of a bank into one of the two categories strongly depends on the level of the cut-off point in terms of probability of failure. In general, the standard approach assigns to an observation a value zero if the expected probability of being in financial distress is lower than 0.5 and a value of one otherwise. However, this approach could be not efficient because that cut-off may not be the cut-off that maximize the overall accuracy of the model. One way to address this issue is to use indices that minimize the level of Type I and Type II errors, which are calculated using the predictive classification table, also known as Confusion Matrix or Error Matrix

		PREDICTED	
		DISTRESSED	NON-DISTRESSED
ACTUAL	DISTRESSED	**TP**	**FN**
	NON-DISTRESSED	**FP**	**TN**

where:

- TP (True Positive): Number of distressed banks correctly predicted as distressed.
- FP (False Positive): Number of healthy banks incorrectly predicted as distressed.

- TN (True Negative): Number of healthy banks correctly predicted as healthy.
- FN (False Negative): Number of distressed banks incorrectly predicted as healthy.

Type I error identifies the percentage of healthy banks classified as distressed (i.e., false alarm) and it can be expressed as:

$$\text{Type I} = \frac{FP}{FP + TN}$$

Type II error occurs when the model fails to identify a failed bank (i.e., missed failure) and it is computed as:

$$\text{Type II} = \frac{FN}{TP + FN}$$

It should be considered that, in general, changing the threshold entail a reduction in one of the two type of error and a simultaneous increase in the second one. More in detail, a higher cut-off entails an increase in Type II error, since a higher number of banks would be categorized as healthy. Conversely, a lower cut-off results in a higher number of banks on the list of distressed banks, which tend to increase the Type I error. From an economic perspective, while false positive lead to additional bank examination costs for the mis-classified healthy banks, missing failures typically imply higher resolution costs or delayed resolutions. Even if literature supports the idea that false negatives are more costly than false positives (Persons, 1999; Poghosyan and Čihak, 2011; Cole and White, 2012; Le and Viviani, 2018) we decide to determine the optimal cut-off since both errors entail additional costs for regulators. We therefore use the Youden index (YI), proposed by Youden (1950), where the two errors are equally weighted. The resulting optimal cut-off corresponds to the threshold that maximizes the distance to the identity line. YI is computed as follows:

$$\text{YI} = \max(\text{sensitivies} + \text{specificities} - 1)$$

where sensitivity (Se) identifies the True Positive rate and specificity (Sp) correspond to the True Negative rate and they are computed, respectively, as follows:

$$\text{Se} = \frac{TP}{TP + FN}$$

$$\text{Sp} = \frac{TN}{TN + FP}$$

We then compare the YI of each model to evaluate the corresponding predictive ability. In addition, we use the Area Under the ROC Curve (AUC) to visualize the performance of our binary classification problem. The ROC curve represents the set of the corresponding specificity rate (x-axis) and sensitivity rate (y-axis) for each cut point between the 0 and 1. The AUC thus represents a measure of the ability of the model to correctly classifies the observations in their respective category. Higher the AUC, the better the model is at distinguishing between classes. The AUC varies between 0.5 and 1 where 1 represents perfect prediction capacity and 0.5 means that the classifier is no better than random guessing.

4. Main Results

This section presents the results, focusing on two key issues: what is the contribution of sustainable performance and other control variables on bank financial distress and to what extent do indicators predict bank vulnerabilities.

Table 4 summarises the estimated coefficient for panel data regression models using fixed effect that we construct to examine whether bank default risk in period *t* is influenced by the bank's CSR activities in period *t*-

1. Model (1-3) show the results for the different specification of Z-score, while Model 4 show coefficients for the regression that consider the Merton distance-to-default (DD) as dependent variable[3]. We find a positive and highly significant coefficient on ESG-score (significant at 1% level in each Z-score approach and at 10% level for the DD measure), confirming the initial hypothesis that on average banks with better non-financial performance are less likely to file bankruptcy. The result of our analysis seems coherent with prior investigations that find both a positive impact of sustainable initiatives on bank performance and bank's reputation. Models show that probability of default is also related to other bank characteristics. The results for the control factors are in general significant in the expected directions. Overall, with respect to bank-specific variables, we show that capital adequacy, asset quality, earnings and liquidity are correlated at least once with risk measures. More in detail, models confirm that higher level of capitalization and higher level of ROA are strongly associated with lower risk, while an increase on NPLs in the loan portfolio entail an increase of default risk.

Table 4: Multivariate FE panel regression of risk measures on the ESG comprehensive score

Variable	Z-score$_1$	Z-score$_2$	Z-score$_3$	DD
ESG (-1)	0.0149***	0.0170***	0.0167***	0.0171*
	(0.0026)	(0.0027)	(0.0026)	(0.0089)
ETA (-1)	0.0397***	0.0645***	0.0675***	0.0092
	(0.0142)	(0.0148)	(0.0147)	(0.049)
NPL (-1)	0.0113	-0.0164**	-0.0165**	-0.0034
	(0.0071)	(0.0074)	(0.0073)	(0.0246)
TLTA (-1)	0.002	0.0016	0.0017	-0.0162
	(0.004)	(0.0042)	(0.0042)	(0.0143)
CIR (-1)	-0.0002	-0.0003	-0.0004	-0.0002
	(0.0011)	(0.0012)	(0.0012)	(0.006)
ROA (-1)	0.0796**	0.0327	0.039	0.2570**
	(0.0361)	(0.0375)	(0.0373)	(0.1279)
NIOR (-1)	0.0008	0.0001	0.000	0.0027
	(0.0019)	(0.0019)	(0.0019)	(0.0074)
CDTA (-1)	-0.0004	0.004	0.0046	0.0219**
	(0.0031)	(0.0033)	(0.0032)	(0.011)
GDP (-1)	0.0271**	0.0192*	0.0229**	-0.3240***
	(0.0112)	(0.0117)	(0.0116)	(0.041)
INF (-1)	-0.1489***	-0.1447***	-0.1372***	-0.0279
	(0.0171)	(0.0178)	(0.0177)	(0.0583)
HHI (-1)	-0.0146	0.0181	0.0169	0.1117*
	(0.0182)	(0.019)	(0.0189)	(0.0621)
Constant	2.9528***	2.2371***	1.1671***	4.1530***
	(0.371)	(0.3861)	(0.3835)	(1.3212)
Observations	1,611	1,611	1,611	1,536
R-Squared	0.1285	0.1293	0.1315	0.0562
F-Test	16.5939	16.7145	17.038	6.4316

This table presents the results of multivariate FE panel regressions. The dependent variables are the bank risk measures z-score and Merton distance-to-default (DD), as described in section 3.2.1. ESG comprehensive score is our target variable. As control variables we include CAMEL covariates (ETA, NPL, TLTA, CIR, ROA, CDTA), a measure of diversification (NIOR) and macroeconomic variables (GDP, INF, HHI), as defined in section 3.2.3. All the explanatory variables are one year lagged. Data are winsorized at the 1% of each tail. Statistical significance is denoted at the 10% (*), 5% (**), and 1% (***) significance level. Standard errors are reported in parentheses. The last rows include R^2 and F test statistics.

[3] The R^2 of models in Table 4 range from the 0.0562 for the panel regression where we use the DD measure to the 0.1315 for the model specification with Z-score$_3$. It is worth considering that the level of the R^2 is in line with the proportion of the variance explained by models in similar studies (Sassen et al., 2016; Lin and Dong, 2018; Neitzert and Petras, 2020, among others).

In addition, the analysis shows a weak positive correlation between the level of customer deposits on total assets and our risk measures, in line with the expectations.
The CAMEL covariates that are find never significant in our model specifications are TLTA, CIR and NIOR. Although the insignificance of the level of total loans over total assets is in contrast with most of the literature (among others, DeYoung 2003; Arena, 2008; Altunbas et al., 2015), it should be noted that other authors (Cole and White, 2012; Berger et al., 2016) highlight that the amount of loan is not relevant in comparison with bank default, while the effective aspect which matter is the composition of the loan portfolio and, in particular, the weight of Construction and Development Loans and commercial real estate, which are found strongly associated with failure. With respect to the cost-to-income ratio, although the coefficients always suggest a positive relation with Z-score and DD measure, the estimates are always not significant. The latter results are anyway in line with previous studies (e.g., Poghosyan and Čihak, 2011; Betz et al., 2014; Chiaramonte et al., 2016), which indicate that low cost does not necessarily entail lower probability of experiencing distress. Surprisingly, we do not find any correlation between diversification level and default risk. Even though the relative coefficients are positive in each model specification and in line with the expectations, they are always not statistically significant.

Table 5: Multivariate FE panel regression of risk measures on the Environmental pillar score (ENV)

Variable	Z-score_1	Z-score_2	Z-score_3	DD
ENV (-1)	0.0072***	0.0068***	0.0068***	-0.0013
	(0.0021)	(0.0022)	(0.0022)	(0.0075)
ETA (-1)	0.0368**	0.0615***	0.0644***	0.007
	(0.0143)	(0.015)	(0.0149)	(0.049)
NPL (-1)	0.0094	-0.0189**	-0.0190**	-0.0082
	(0.0071)	(0.0075)	(0.0074)	(0.0246)
TLTA (-1)	0.0039	0.0039	0.0039	-0.0122
	(0.0041)	(0.0042)	(0.0042)	(0.0142)
CIR (-1)	-0.0002	-0.0003	-0.0004	-0.0003
	(0.0011)	(0.0012)	(0.0012)	(0.006)
ROA (-1)	0.0882**	0.0414	0.0476	0.2590**
	(0.0364)	(0.0381)	(0.0378)	(0.1281)
NIOR (-1)	0.0013	0.0005	0.0004	0.0032
	(0.0019)	(0.002)	(0.002)	(0.0074)
CDTA (-1)	0.0002	0.0047	0.0053	0.0225**
	(0.0032)	(0.0033)	(0.0033)	(0.011)
GDP (-1)	0.0360***	0.0291**	0.0326***	-0.3152***
	(0.0112)	(0.0117)	(0.0117)	(0.0408)
INF (-1)	-0.1442***	-0.1396***	-0.1322***	-0.0265
	(0.0173)	(0.018)	(0.0179)	(0.0584)
HHI (-1)	-0.0225	0.0095	0.0084	0.1045*
	(0.0184)	(0.0192)	(0.0191)	(0.0621)
Constant	3.2852***	2.6472***	1.5674***	4.6859***
	(0.3676)	(0.384)	(0.3812)	(1.3036)
Observations	1,611	1,611	1,611	1,536
R-Squared	0.1127	0.1073	0.1102	0.0533
F-Test	14.292	13.5229	13.937	6.0764

This table presents the results of multivariate FE panel regressions. The dependent variables are the bank risk measures z-score and Merton distance-to-default (DD), as described in section 3.2.1. Environmental pillar score is our target variable. As control variables we include CAMEL covariates (ETA, NPL, TLTA, CIR, ROA, CDTA), a measure of diversification (NIOR) and macroeconomic variables (GDP, INF, HHI), as defined in section 3.2.3. All the explanatory variables are one year lagged. Data are winsorized at the 1% of each tail. Statistical significance is denoted at the 10% (*), 5% (**), and 1% (***) significance level. Standard errors are reported in parentheses. The last rows include R^2 and F test statistics.

With respect to macroeconomic factors, Table 4 confirms that GDP growth and inflation rate are relevant determinant of bank failure. With respect to Models (1-3), we find that higher rate GDP growth and lower inflation rate are associated with more stable macroeconomic environment, which entails a reduction in bank distress probability. In line with the literature, macroeconomic environment plays a significant role in explaining individual bank distress (Betz et al., 2014; Chiaramonte and Casu, 2017 among others), which confirms the opportunity to implement and integrate macro prudential regulation. However, with respect to Model 4, GDP growth is negatively associated with DD measure. This unexpected result could be explained by the different time windows considered in the construction of dependent variables: while Z-scores are built on a 5 years-time horizon, DD measure reflects market values in time *t.* Our untabulated results show that DD measure in time *t* is strictly and positively correlated with GDP growth in time *t,* in line with the expectations. Additionally, the results suggest a weak negative correlation between market concentration and probability of financial distress, which entail a possible confirm of the "concentration-stability" view.

Table 6: Multivariate FE panel regression of risk measures on the Social pillar score (SOC)

Variable	Z-score$_1$	Z-score$_2$	Z-score$_3$	DD
SOC (-1)	0.0139***	0.0149***	0.0146***	0.0190**
	(0.0023)	(0.0024)	(0.0024)	(0.0078)
ETA (-1)	0.0371***	0.0617***	0.0647***	0.0065
	(0.0142)	(0.0148)	(0.0147)	(0.0489)
NPL (-1)	0.093	-0.0156**	-0.0157**	-0.0013
	(0.0071)	(0.0074)	(0.0073)	(0.0247)
TLTA (-1)	0.0032	0.0031	0.0031	-0.0151
	(0.004)	(0.0042)	(0.0042)	(0.0142)
CIR (-1)	-0.0003	-0.0004	-0.0005	-0.0006
	(0.0011)	(0.0012)	(0.0012)	(0.006)
ROA (-1)	0.0815**	0.035	0.0412	0.2568**
	(0.036)	(0.0376)	(0.0373)	(0.1278)
NIOR (-1)	0.0006	-0.0001	-0.0003	0.0023
	(0.0019)	(0.0019)	(0.0019)	(0.0074)
CDTA (-1)	-0.0014	0.0031	0.0036	0.0202*
	(0.0031)	(0.0033)	(0.0032)	(0.011)
GDP (-1)	0.0249**	0.0174	0.0212*	-0.3285***
	(0.0112)	(0.0117)	(0.0116)	(0.0411)
INF (-1)	-0.1460***	-0.1414***	-0.1340***	-0.0258
	(0.0171)	(0.0178)	(0.0177)	(0.0582)
HHI (-1)	-0.0151	0.0172	0.016	0.1125*
	(0.0182)	(0.019)	(0.0189)	(0.0621)
Constant	3.0023***	2.3261***	1.2551***	4.1554***
	(0.368)	(0.3838)	(0.3812)	(1.3102)
Observations	1,611	1,611	1,611	1,536
R-Squared	0.1308	0.1281	0.1302	0.058
F-Test	16.9347	16.5369	16.8538	6.6402

This table presents the results of multivariate FE panel regressions. The dependent variables are the bank risk measures z-score and Merton distance-to-default (DD), as described in section 3.2.1. Social pillar score is our target variable. As control variables we include CAMEL covariates (ETA, NPL, TLTA, CIR, ROA, CDTA), a measure of diversification (NIOR) and macroeconomic variables (GDP, INF, HHI), as defined in section 3.2.3. All the explanatory variables are one year lagged. Data are winsorized at the 1% of each tail. Statistical significance is denoted at the 10% (*), 5% (**), and 1% (***) significance level. Standard errors are reported in parentheses. The last rows include R^2 and F test statistics.

To evaluate the contribution of each ESG factor on bank risk, we regressed our risk measures on the three pillar scores ENV, SOC, GOV instead of using the aggregated ESG score. Table 5, 6, and 7 summarises the results of panel regressions of each of the three pillars, using the same dependent variables and control variables used for the comprehensive score in Table 4. The models show that on average all the sub-components are positively and significantly related with z-score and DD measure and thus negatively related with bank default risk. However, the coefficients reveal that there is no homogeneity in the association of the regression of risk measures over all the specification of the ESG score. In particular, we find that social performances are strongly and significantly correlated with risk measures in each different approach, as shown in Table 5 (more specifically, it is significant at 1% level in each Z-score approach and at 5% level for the DD measure). Furthermore, the coefficients associated with SOC pillar is the largest among the sub-components. The high relevance of social performances could be explained by the potential impact on reputation. As highlighted by Dell'Atti et al. (2017), banks that behave ethically and provide superior services are considered to be valuable by customers. Therefore, as stated by Krueger et al. (2010), an increase in reputation entails a significant reduction in riskiness. On the other hand, ENV and GOV dimensions have a similar impact on bank default risk. Even though they are significantly correlated at 1% with Z-scores, the relation with DD measure is not statistically significant. The lower magnitude is coherent with results obtained in similar studies: a possible explanation could be found in the work by Brogi and Lagasio (2018), which state that the improvements in the management performance may take longer to deploy their effects. The weak contribution of environmental performance could be justified by the still relative weak focus on the environmental impact of banking activity - although it is gradually increasing-, which we have identified examining the mean value of each subcomponent in the descriptive statistics (see Table 2). Nevertheless, the positive and significant relation with risk measures may be considered as robustness check of what we find in models described in Table 4.

Table 7: Multivariate FE panel regression of risk measures on the Governance pillar score (GOV)

Variable	Z-score$_1$	Z-score$_2$	Z-score$_3$	DD
GOV (-1)	0.0050***	0.0065***	0.0063***	0.0044
	(0.0015)	(0.0016)	(0.0016)	(0.0052)
ETA (-1)	0.0397***	0.0649***	0.0678***	0.0084
	(0.0144)	(0.0149)	(0.0148)	(0.0490)
NPL (-1)	0.0089	-0.0191**	-0.0191***	-0.0068
	(0.0071)	(0.0074)	(0.0074)	(0.0246)
TLTA (-1)	0.0027	0.0022	0.0022	-0.0141
	(0.0041)	(0.0043)	(0.0042)	(0.0143)
CIR (-1)	-0.0002	-0.0002	-0.0003	-0.0001
	(0.0011)	(0.0012)	(0.0012)	(0.0060)
ROA (-1)	0.0783**	0.0307	0.0370	0.2578**
	(0.0364)	(0.0379)	(0.0377)	(0.1281)
NIOR (-1)	0.0009	0.0002	0.0000	0.0031
	(0.0019)	(0.0020)	(0.0020)	(0.0074)
CDTA (-1)	0.0005	0.0051	0.0056*	0.0228**
	(0.0032)	(0.0033)	(0.0033)	(0.0110)
GDP (-1)	0.0327***	0.0253**	0.0289**	-0.3167***
	(0.0112)	(0.0117)	(0.0116)	(0.0408)
INF (-1)	-0.1483***	-0.1445***	-0.1370***	-0.0274
	(0.0173)	(0.0180)	(0.0179)	(0.0584)
HHI (-1)	-0.0180	0.0148	0.0136	0.1066*
	(0.0184)	(0.0192)	(0.0190)	(0.0622)
Constant	3.2567***	2.5552***	1.4799***	4.5023***
	(0.3691)	(0.3843)	(0.3816)	(1.3102)
Observations	1,611	1,611	1,611	1,536
R-Squared	0.1122	0.1125	0.1150	0.0538
F-Test	14.2298	14.2610	14.6259	6.1412

This table presents the results of multivariate FE panel regressions. The dependent variables are the bank risk measures z-score and Merton distance-to-default (DD), as described in section 3.2.1. Governance pillar score is our target variable. As control variables we include CAMEL covariates (ETA, NPL, TLTA, CIR, ROA, CDTA), a measure of diversification (NIOR) and macroeconomic variables (GDP, INF, HHI), as defined in section 3.2.3. All the explanatory variables are one year lagged. Data are winsorized at the 1% of each tail. Statistical significance is denoted at the 10% (*), 5% (**), and 1% (***) significance level. Standard errors are reported in parentheses. The last rows include R^2 and F test statistics.

We are also interested in whether our model can be used as a forecasting tool by regulators and whether ESG score may increase the predictive accuracy of such tools. To answer these questions, we test and compare the predictive accuracy of models presented in Section 3.4, using as predictors the control variables tested in the previous analysis and the ESG score in its comprehensive specification. For each model, we split the sample into 2 subsets: a Training set (70% of data) and a Test set (30% of data). Overall, we use a total of 1127 observations to train and cross-validate the model and the remaining 484 observations to test the ability to predict unknown data (out-of-sample forecasting). Among the several performance measures for supervised algorithms[4], we decide to compare accuracy, Type I and Type II error rates and the AUC.

Figure 2 shows the out-of-sample predictive performance of models based upon the estimation procedure presented in Section 3.4, which recursively test each model using an increasing percentile of the Z-score. Specifically, in Figure 2, the six models are compared computing the Youden Index, which identify for each specification the best threshold that minimize the sum of Type I and Type II errors. In detail, higher value of the Youden Index implies better capacity of the model to reduce the incorrect classifications.

Figure 2: Models comparison by Youden index

Youden index for different Z-score percentile. Source: Authors' estimations

Figure 2 shows the that, on average, models follow similar patterns. However, the analysis supports the opinion that ensemble models outperform single classifiers. Ensemble models, and in particular Random forests, outperform other algorithms in each estimation procedure. Conversely, among single classifiers, there is not a clear difference between models: nevertheless, classification trees seem to have the lowest performance.

Furthermore, the figure indicates that the predictive ability of models decreases as the level of the percentile increases, i.e., a higher percentage of banks are identified as distressed. This pattern probably implies that the z-score is particularly suitable for identify banks in particularly severe financial distress. This result is in line

[4] The reader can refer to Caruana and Niculescu-Mizil (2006) and Ferri et al. (2009) for a comprehensive review of performance measured used in previous papers.

with the finding in Chiaramonte et al. (2016), which highlights that the highest percentage of failure is found in the tenth deciles of the probability distribution of the Z-score. Table 8 thus report the confusion matrix and performance measures of our models in the fifth percentile specification. The parameters refer to the out-of-sample forecasting, where 26 distress episodes have been detected. The diagonal elements in the classification table summarize the number of observations which are correctly classified. The off-diagonal elements present the misclassification, including Type I and Type II errors.

Table 8: Out-of-sample performance of prediction methods

Method	Confusion matrix		Sensitivity (%)	Specificity (%)	OPA (%)	AUC (%)
Logit	22	4	84.6 %	89.3	89.1	92.0
	49	409				
LDA	25	1	96.2	81.0	81.8	92.4
	87	371				
SVM	23	3	88.5	85.2	85.3	91.4
	68	390				
DT	21	5	80.8	90.0	89.5	92.0
	46	412				
RF	22	4	84.6	90.8	90.5	91.3
	42	416				
Xgboost	21	5	80.8	94.3	93.6	92.7
	26	432				

Table gives the confusion matrix and accuracy measures for the prediction models when using the fifth percentile of the empirical probability distribution of the z-score. For each model we use the cut-off identify by the Youden Index. Sensitivity is the fraction of those predicted distressed that are actually distressed. Specificity is the fraction of those that are predicted non-distressed that are actually non-distressed. Overall Prediction Accuracy (OPA) is the fraction of those predicted correctly. AUC: Area under the Receiver Operation Characteristic curve

From the table, it can be seen that the overall accuracy of our models ranges from 81.8% to 93.6%. According to the experimental results, we conclude that prediction performance of ensemble methods is slightly better than single classifiers. Again, there are no substantial difference between statistical and AI methods. It should be noted that LDA, which has the worst overall accuracy, shows the lowest level of Type II error, which is considered by far the most significant issue, at the cost of a higher Type I error. However, sensitivity and specificity level strongly depend on the cut-off identified by the Youden Index. The Area under the ROC curve allows to evaluate the performance of a classification problem at various threshold settings, independently from the cut-off selected. Table 8 shows that models have similar performance, since AUC range from 91.4 and 92.7. Overall, every model seems to exhibit an excellent accuracy. Our results confirm that our model can easily be applied to a large number of banks, even those that do not fail, and has a very promising ability in distinguishing healthy from distressed banks.

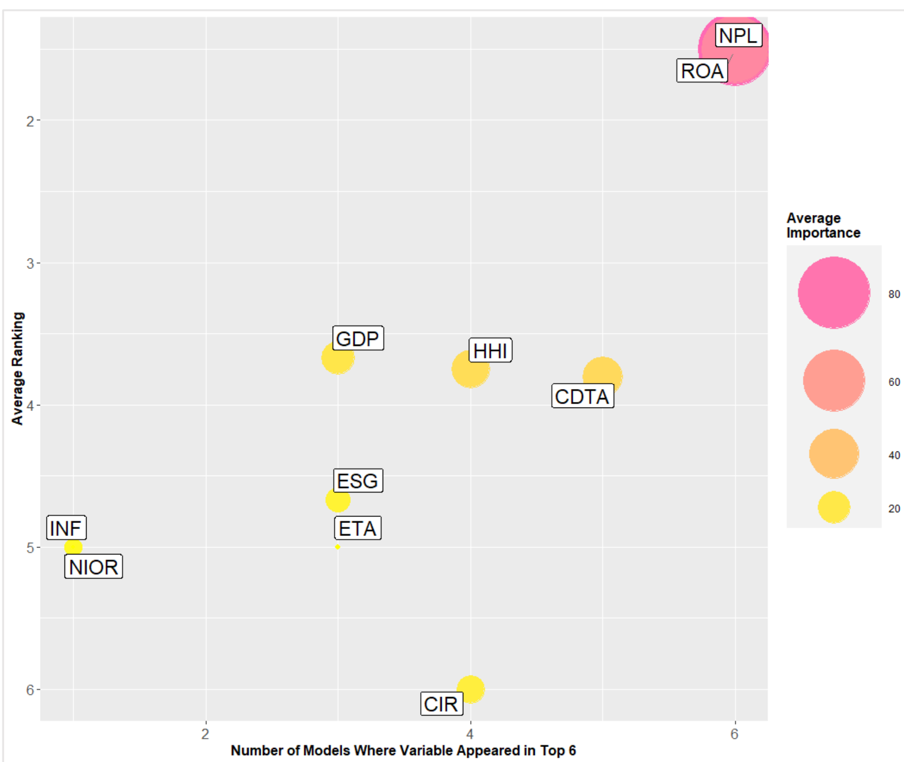
Finally, we show the improvement in predictive power of each variable which we include in the model. In order to determine the relative contribution factor of the variables, we apply for each model a different compatible method. In detail, we apply the following procedures:

- Logit: the absolute value of the *t*-statistic for each model parameter.
- LDA: the absolute value of the standardized coefficients of the linear discriminant analysis.
- SVM: we use the variable selection algorithms, described in Guyon and Elisseeff (2003), which consists in the optimization of the objective function of variable selection. The objective function consists of two terms that compete with each other: the goodness-of-fit and the number of variables.
- DT: the percentage of training set samples that fall into all the terminal nodes after the split

- RF: the Gini index, which is a widely used tools that provides a metrics of how close a model or variable is to the ideal prediction. Such index highlights the contribution of each variable to the homogeneity of the nodes and leaves.
- Xgboost: the "Gain" contribution of each feature to the model. It represents the average gain across all splits of each considered tree. High value denotes important feature to predict the response variable (Chen and Benesty, 2016).

After the computation of each procedure, we normalized the output of each variable importance method on a 0-100 scale for comparison needs. Figure 3 illustrates the overall variable importance across the models considered in the analysis. The y-axis, which represents the variable's average ranking across models, is reversed so that the top variables are shown in the upper right corner. Although this may not provide a comprehensive description of the contributions, it can show general trends and provide a useful basis for interpretation. Not surprisingly, NPL ratio and the level of earnings are the most important factors across all the models. Such factors are indeed always found as the main predictors of a severe financial crisis. Among the others, macroeconomic factors display a relatively high importance value. With respect of ESG-score, the analysis demonstrates that, on average, it occupies the sixth position among the eleven selected variables. Despite its relative predictive capacity, ESG-score has a higher importance than more traditional variables, such as ETA, the level of diversification (NIOR), the management efficiency (CIR) and TLTA, which it has been never found among the top 6 variables.

Figure 3: Variable importance across models



The figure captures three different dimensions: the x-axis identifies the number of models where the feature appeared in top 6; the y-axis shows the feature's average ranking across models; the dimension of the spots highlights the normalized importance value across models. We reverse the y-axis to increase the figure's interpretability.

5. Conclusions

Despite the increasing attention on non-financial performance in recent years among media, investors, and regulators, the existing literature has provided only limited evidence on the relationship between Corporate social performance and bank riskiness. To fill this gap, this study seeks to address two research questions. On the one hand, we deepen the understanding of the relationship between non-financial performance and bank risk. On the other hand, we develop a bank's financial distress prediction model to evaluate its predictive ability and assess the goodness of the ESG-score as a predictor variable. For these purposes, we use an unbalance dataset of 362 commercial banks headquartered in the USA and EU-28 member states covering the period 2012-2019.

First, to assess the relationship between CSP and bank risk we apply two well-known risk measures: an accounting measure (Z-score, which we apply in three different approaches for robustness test) and a market-based measure (Merton distance-to-default). We then perform fixed effects regressions using a large set of control variables and the Thomson Reuters ESG-score, which we introduce as a measure for non-financial performance, both at an aggregate level as well as disaggregated in its three sub-components. We find that higher levels of the aggregated ESG score correspond to a significant reduction for all the risk measures. The decomposition of the ESG-score in the environmental, social and governance dimension shows that each individual components has a risk-reducing effect, which it confirms our initial findings. However, the analysis reveals that the three sub-components affect the overall risk with different intensities: whereas social performance are always statistically significant, governance and environmental dimensions have instead a weaker risk-reducing effect. Our empirical analysis supports the idea that ethical /sustainable company behavior is considered to be valuable by stakeholders. Conversely, while improvements in management dimension may deploy its effect in larger time horizon, the relation between environmental performance and risk may depend on the importance that stakeholders of a given industry assign to such issues (Pérez and Rodriguez del Bosque, 2014; Dell'Atti et al., 2017).

On this basis, we define our second research question. We thus assess the prediction ability of ESG-score by developing a bank's financial distress prediction model. The choice to compare different statistical and AI methods has a twofold objective: such analysis allows us to strength the validity of our findings and verify the popular opinion that, on average, ensemble classifiers could outperform individual techniques. Empirical results give some evidence that support the abovementioned thesis. Even more relevant, results of absolute usefulness and AUC highlight that our predictive model has an excellent predictive ability in each specification. Finally, the analysis of variable importance suggests that the level of non-performing loans and levels of earnings have by far the largest predictive capacity. Among the other variables, macroeconomic factors play a crucial role in identifying severe financial distress. In addition, the analysis show that ESG-score may increase the predictive ability of models, even if with a lower magnitude. Nevertheless, non-financial performance is found to have higher predictive ability than others traditional variables, such as management efficiency, level of diversification, and incidence of total loans on total assets.

Our study has relevant implications for policy makers as well for managers and investors. This study shows that the proper identification and achievement of sustainable goals do not only have potential positive effect on accounting and market performance, but it also contributes to reduce bank riskiness. From a managerial perspective, banks should continue to implement ESG practices in their business plan, as requested also by new directions set by policy makers internationally (e.g., EBA Action Plan on Sustainable Finance, 2019). With respect to banks regulators, such results suggest the necessity to continue to improve the norms requiring better sustainability disclosure in favour of stakeholders and investors, higher diffusion of best-practices, and the identification of proper tools and approaches to measure and monitor risks connected to environmental, social and governance dimension. Furthermore, the analysis confirms that ESG-score may have enough predictive ability to be included in more advanced prediction models.

Nonetheless, the interpretation of our results should encounter some limitations. Firstly, though Thomson Reuters ESG score have found widely applications in previous studies, we are aware that non-financial ratings still suffer from lack of standardized rules and formal auditing processes that add a subjective nature of ratings (LaBella et al., 2019) that could lead to different evaluations of the same banks (Soana, 2011). Secondly, we based our analysis on an unbalanced panels of commercial banks headquartered in well developed markets. As highlighted in previous studies (e.g., Jo et al., 2015; Miralles-Quirós et al., 2019), the effect of sustainable performance may change across country and markets. It may therefore be difficult to generalize our results on wider samples. Finally, endogeneity issue is only partially addressed using lagged independent variables. In addition, the effect of non-financial performance on bank riskiness could change in larger time windows. Further investigations may be conducted to deeper analyse and understand the abovementioned issues.

References

Adnan Aziz, M., & Dar, H. A. (2006). Predicting corporate bankruptcy: where we stand? *Corporate Governance: The International Journal of Business in Society*, 6(1), 18-33.

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, *32*(8), 1541-1551.

Allen, F., & Gale, D. (2000). Comparing financial systems. Cambridge, MA: MIT Press.

Altunbas, Y., Manganelli, S., & Marques-Ibanez, D. (2011). Bank risk during the financial crisis: do business models matter? *ECB Working Paper Series*, No. 1394.

Andersen, H. (2008). Failure prediction of Norwegian banks: A logit approach. *Norges Bank Working Paper,* No. 2008/2.

Anginer, D., & Demirguc-Kunt, A. (2014). Bank Capital and Systemic Stability. In *Policy Research Working Papers*. The World Bank.

Arena, M. (2008). Bank failures and bank fundamentals: A comparative analysis of Latin America and East Asia during the nineties using bank-level data. *Journal of Banking & Finance*, *32*(2), 299-310.

Aubuchon, C. P., & Wheelock, D. C. (2010) The geographic distribution and characteristics of U.S. bank failures, 2007–2010: do bank failures still reflect local economic conditions? *Federal Reserve Bank of St. Louis Review, 92 (5), 395-415*.416

Avkiran, N. K., & Cai, L. C. (2012). Predicting bank financial distress prior to crises. In *New Zealand Finance Colloquium.*

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, *38*(1), 63-93.

Bassen, A., Meyer, K., & Schlange, J. (2006). The influence of corporate responsibility on the cost of capital. *SSRN Network Papers.*

Beck, T., De Jonghe, O., & Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation*, 22(2), 218–244.

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2010). Financial institutions and markets across countries and over time: The updated financial development and structure database. *The World Bank Economic Review*, *24*(1), 77-92.

Bellovary, J., Giacomino, D., & Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, 33, 1-42.

Beltratti, A., & Stulz, R. M. (2012). The credit crisis around the globe: Why did some banks perform better? *Journal of Financial Economics*, 105(1), 1–17.

Benston, G. J., & Kaufman, G. G. (1995). Is the Banking and Payments System Fragile? In Coping with Financial Fragility and Systemic Risk (pp. 15–46). Springer US.

Berger, A. N., Imbierowicz, B., & Rauch, C. (2016). The roles of corporate governance in bank failures during the recent financial crisis. *Journal of Money, Credit and Banking*, *48*(4), 729-770.

Betz, F., Oprică, S., Peltonen, T. A., & Sarlin, P. (2014). Predicting distress in European banks. *Journal of Banking & Finance*, *45*, 225-241.

Bhagat, S., Bolton, B., & Lu, J. (2015). Size, leverage, and risk-taking of financial institutions. *Journal of Banking & Finance*, 59, 520–537.

Bharath, S. T., & Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *Review of Financial Studies*, 21(3), 1339–1369

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3), 637-654.

Bongini, P., Laeven, L., & Majnoni, G. (2002). How good is the market at assessing bank fragility? A horse race between different indicators. *Journal of Banking & Finance*, 26(5), 1011–1028.

Boyacioglu, M. A., Kara, Y., & Baykan, Ö. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications*, *36*(2), 3355-3366.

Boyd, J. H., & De Nicolò, G. (2005). The Theory of Bank Risk Taking and Competition Revisited. *The Journal of Finance*, 60(3), 1329–1343.

Boyd, J. H., & Graham, S. L. (1986). Risk, Regulation, and Bank Holding Company Expansion Into Nonbanking. *Quarterly Review*, 10(2).

Branco, M. C., & Rodrigues, L. L. (2006). Corporate social responsibility and resource-based perspectives. *Journal of business Ethics*, *69*(2), 111-132.

Bräuning, M., Malikkidou, D., Scricco, G., & Scalone, S. (2019). A new approach to Early Warning Systems for small European banks. *ECB Working Paper Series*, No. 2348.

Brogi, M., & Lagasio, V. (2019). Environmental, social, and governance and company profitability: Are financial intermediaries different?. *Corporate Social Responsibility and Environmental Management*, *26*(3), 576-587.

Campbell, J. Y., Hilscher, J. D., & Szilagyi, J. (2011). Predicting financial distress and the performance of distressed stocks. *Journal of Investment Management,* 9(2), 14-34.

Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, *166*(2), 528-546.

Caprio,G. , & Daniela K. (1996). Bank Insolvencies: Cross-Country Experience. *Policy Research Working Paper,* No. 1620.

Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, *61*, 304-323.

Carnevale, C., & Mazzuca, M. (2014). Sustainability report and bank valuation: evidence from E uropean stock markets. *Business Ethics: A European Review*, *23*(1), 69-90.

Carreras, E., Alloza, Á., & Carreras, A. (2013). *Corporate reputation* (Vol. 1). LID Editorial.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Proceedings of the 23rd International Conference on Machine Learning - ICML '06. the 23rd international conference.

Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European accounting review*, *13*(3), 465-497.

Chen, T., & Benesty, M. (2016). XGBoost: eXtreme gradient boosting. R package version 0.4-3. https://cran.r project.org/web/packages/XGBoost/vignettes/XGBoost.pdf.

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic management journal*, *35*(1), 1-23.

Chiaramonte, L., & Casu, B. (2017). Capital and liquidity ratios and financial distress. Evidence from the European banking industry. *The British Accounting Review*, *49*(2), 138-161.

Chiaramonte, L., Liu, H., Poli, F., & Zhou, M. (2016). How accurately can Z-score predict bank failure?. *Financial Markets, Institutions & Instruments*, *25*(5), 333-360.

Chih, H. L., Chih, H. H., & Chen, T. Y. (2010). On the determinants of corporate social responsibility: International evidence on the financial industry. *Journal of Business Ethics*, *93*(1), 115-135.

Ciciretti, R., Kobeissi, N., & Zhu, Y. (2014). Corporate social responsibility and financial performance: an analysis of bank community responsibility. *International Journal of Banking, Accounting and Finance*, *5*(4), 342-373.

Cipollini, A., & Fiordelisi, F. (2012). Economic value, competition and financial distress in the European banking system. *Journal of Banking & Finance*, *36*(11), 3101-3109.

Citterio, A. (2020). Bank Failures: Review and Comparison of Prediction Models. *SSRN Network Papers*.

Cole, R. A., & White, L. J. (2012). Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research*, *42*(1-2), 5-29.

Cornett, M. M., Erhemjamts, O., & Tehranian, H. (2014). Corporate social responsibility and its impact on financial performance: Investigation of US commercial banks. *Working paper*.

Cornett, M. M., Erhemjamts, O., & Tehranian, H. (2016). Greed or good deeds: An examination of the relation between corporate social responsibility and the financial performance of US commercial banks around the financial crisis. *Journal of Banking & Finance*, *70*, 137-159.

Cox, R. A., & Wang, G. W. Y. (2014). Predicting the US bank failure: A discriminant analysis. *Economic Analysis and Policy*, *44*(2), 202-211.

Delis, M. D., & Staikouras, P. K. (2011). Supervisory Effectiveness and Bank Risk. *Review of Finance*, 15(3), 511–543.

Dell'Atti, S., Trotta, A., Iannuzzi, A. P., & Demaria, F. (2017). Corporate social responsibility engagement as a determinant of bank reputation: An empirical analysis. *Corporate Social Responsibility and Environmental Management*, *24*(6), 589-605.

Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. *Omega*, *38*(5), 315-324.

DeYoung, R. (2003). De novo bank exit. *Journal of Money, Credit and Banking*, 711-728.

DeYoung, R., & Torna, G. (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, *22*(3), 397-421.

Distinguin, I., Rous, P., & Tarazi, A. (2006). Market discipline and the use of stock market data to predict bank financial distress. *Journal of Financial Services Research*, *30*(2), 151-176.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635–665.

Ecer, F. (2013). Comparing the bank failure prediction performance of neural networks and support vector machines: The Turkish case. *Economic research-Ekonomska istraživanja*, *26*(3), 81-98.

Ekinci, A., & Erdal, H. İ. (2017). Forecasting bank failure: Base learners, ensembles and hybrid ensembles. *Computational Economics*, *49*(4), 677-686.

Esteban-Sanchez, P., de la Cuesta-Gonzalez, M., & Paredes-Gazquez, J. D. (2017). Corporate social performance and its relation with corporate financial performance: International evidence in the banking industry. *Journal of cleaner production*, *162*, 1102-1110.

Evans, O., Leone, A. M., Gill, M., & Hilbers, P. (2000). Macroprudential indicators of financial systems soundness, *IMF Occasional paper,* No. 192.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

Fernando, J. M. R., Li, L., & Hou, Y. (2019). Corporate governance and default prediction: a reality test. *Applied Economics*, *51*(24), 2669-2686.

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1), 27–38.

Flannery, M. J. (1998). Using market information in prudential bank supervision: A review of the US empirical evidence. *Journal of Money, Credit and Banking*, 273-305.

Forcadell, F. J., & Aracil, E. (2017). European banks' reputation for corporate social responsibility. *Corporate Social Responsibility and Environmental Management*, *24*(1), 1-14.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210-233.

Gangi, F., Meles, A., D'Angelo, E., & Daniele, L. M. (2019). Sustainable development and corporate governance in the financial system: Are environmentally friendly banks less risky?. *Corporate Social Responsibility and Environmental Management*, *26*(3), 529-547.

García-Sánchez, I. M., & García-Meca, E. (2017). CSR engagement and earnings quality in banks. The moderating role of institutional factors. *Corporate Social Responsibility and Environmental Management*, *24*(2), 145-158.

Glavas, A., & Kelley, K. (2014). The effects of perceived corporate social responsibility on employee attitudes. *Business Ethics Quarterly*, *24*(2), 165-202.

Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, *34*(3), 440-455.

Gropp, R., Vesala, J., & Vulpes, G. (2006). Equity and Bond Market Signals as Leading Indicators of Bank Fragility. *Journal of Money, Credit and Banking, 38*(2), 399-428.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(7-8), 1157–1182.

Halling, M., & Hayden, E. (2006). Bank failure prediction: a two-step survival time approach. *SSRN Network Papers.*

Hong, H., Huang, J. Z., & Wu, D. (2014). The information content of Basel III liquidity risk measures. *Journal of Financial Stability*, *15*, 91-111.

Honohan, P., Klingebiel, D. (2000). Controlling fiscal costs of banking crises. *Policy Research Working Paper,* No. 2441.

Hurley, R., Gong, X., & Waqar, A. (2014). Understanding the loss of trust in large banks. *International Journal of Bank Marketing*, 32(5), 348–366.

Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with applications*, *42*(6), 2857-2869.

Jagtiani, J., & Lemieux, C. (2001). Market discipline prior to bank failure. *Journal of Economics and Business*, *53*(2-3), 313-324.

Jing, Z., & Fang, Y. (2018). Predicting US bank failures: A comparison of logit and data mining models. *Journal of Forecasting*, *37*(2), 235-256.

Jo, H., Kim, H., & Park, K. (2015). Corporate environmental responsibility and firm performance in the financial services sector. *Journal of business ethics*, *131*(2), 257-284.

Jordan, D. J., Rice, D., Sanchez, J., Walker, C., & Wort, D. H. (2010). Predicting bank failures: Evidence from 2007 to 2010. *SSRN Network Papers.*

Kemp-Benedict, E. (2018). Investing in a green transition. *Ecological Economics*, *153*, 218-236.

Kimmel, R. K., Thornton, J. H., Jr., & Bennett, S. E. (2016). Can statistics-based early warning systems detect problem banks before markets? *The North American Journal of Economics and Financ*e, 37, 190–216.

Krueger, T. M., Wrolstad, M. A., & Van Dalsem, S. (2010). Contemporaneous relationship between corporate reputation and return. *Managerial Finance*, 36(6), 482–490.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research*, *180*(1), 1-28.

LaBella, M. J., Russell, J., & Novikov, D. (2019). The devil is in the details: the divergence in ESG data and implications for responsible investing. *Research paper QS Investor.*

Laeven, L., & Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93(2), 259–275.

Le, H. H., & Viviani, J. L. (2018). Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance*, *44*, 16-25.

Li, X., & Malone, C. B. (2016). Measuring Bank Risk: An Exploration of Z-Score. *SSRN Network Papers.*

Lin, K. C., & Dong, X. (2018). Corporate social responsibility engagement of financially distressed firms and their bankruptcy likelihood. *Advances in Accounting*, 43, 32–45.

Linthicum, C., Reitenga, A. L., & Sanchez, J. M. (2010). Social responsibility and corporate reputation: The case of the Arthur Andersen Enron audit failure. *Journal of Accounting and Public Policy*, *29*(2), 160-176.

Liu, H., Molyneux, P., & Wilson, J. O. S. (2012). Competion and stability in european banking: a regional analysis. *The Manchester School*, 81(2), 176–201.

Makni, R., Francoeur, C., & Bellavance, F. (2009). Causality between corporate social performance and financial performance: Evidence from Canadian firms. *Journal of Business Ethic*s, 89(3), 409–422.

Männasoo, K., & Mayes, D. G. (2009). Explaining bank distress in Eastern European transition economies. *Journal of Banking & Finance*, *33*(2), 244-253.

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative science quarterly*, *48*(2), 268-305.

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance. *SSRN Network Papers.*

Mayes, D. G., & Stremmel, H. (2014). The effectiveness of capital adequacy measures in predicting bank distress: SUERF Study 2014/1. *Brussels: Larcier*.

Miralles-Quirós, M. M., Miralles-Quirós, J. L., & Redondo-Hernández, J. (2019). The impact of environmental, social, and governance performance on stock prices: Evidence from the banking industry. *Corporate Social Responsibility and Environmental Management*, *26*(6), 1446-1456.

Momparler, A., Carmona, P., & Climent, F. (2016). Banking failure prediction: a boosting classification tree approach. *Spanish Journal of Finance and Accounting/Revista Española De Financiación Y Contabilidad*, *45*(1), 63-91.

Neitzert, F., & Petras, M. (2019). Corporate Social Responsibility and Bank Risk. *SSRN Network Papers.*

Nobanee, H., & Ellili, N. (2016). Corporate sustainability disclosure in annual reports: Evidence from UAE banks: Islamic versus conventional. *Renewable and Sustainable Energy Reviews*, *55*, 1336-1341.

Papadimitriou, T., Gogas, P., Plakandaras, V., & Mourmouris, J. C. (2013). Forecasting the insolvency of US banks using support vector machines (SVMs) based on local learning feature selection. *International Journal of Computational Economics and Econometrics*, *3*(1-2), 83-90.

Pérez, A., & Rodríguez del Bosque, I. (2014). Customer CSR expectations in the banking industry. *International Journal of Bank Marketing*, 32(3), 223–244.

Persons, O. (1999). Using financial information to differentiate failed vs. surviving finance companies in Thailand: an implication for emerging economies. *Multinational Finance Journal*, *3*(2), 127-145.

Pille, P., & Paradi, J. C. (2002). Financial performance analysis of Ontario (Canada) Credit Unions: An application of DEA in the regulatory environment. *European Journal of Operational Research*, 139(2), 339–350.

Poghosyan, T., & Čihak, M. (2011). Determinants of bank distress in Europe: Evidence from a new data set. *Journal of Financial Services Research*, *40*(3), 163-184.

Reidhill, J and O'Keefe, J (1997). Off-site surveillance systems, in Federal Deposit Insurance Corporation History of the Eighties – Lessons for the Future, *FDIC*, 477-520.

Roy, A. D. (1952). Safety First and the Holding of Assets. *Econometrica*, 20(3), 431.

Sassen, R., Hinze, A.-K., & Hardeck, I. (2016). Impact of ESG factors on firm risk in Europe. *Journal of Business Economics*, 86(8), 867–904.

Saxena, M., & Kohli, A. S. (2012). Impact of Corporate Social Responsibility on Corporate Sustainability: A Study of the Indian Banking Industry. *IUP Journal of Corporate Governance*, *11*(4).

Scholtens, B., & van't Klooster, S. (2019). Sustainability and bank risk. *Palgrave Communications*, *5*(1), 1-8.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2013). Partial least square discriminant analysis for bankruptcy prediction. *Decision support systems*, *54*(3), 1245-1255.

Sheehy, B. (2014). Defining CSR: Problems and Solutions. *Journal of Business Ethics*, 131(3), 625–648.

Shleifer, A., & Vishny, R. W. (2010). Unstable banking. *Journal of Financial Economics*, 97(3), 306–318.

Shrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting. *Cogent Economics & Finance*, *8*(1), 1729569.

Simpson, W. G., & Gleason, A. E. (1999). Board structure, ownership, and financial distress in banking firms. *International Review of Economics & Finance*, *8*(3), 281-292.

Sinkey, J. F. (1975). A multivariate statistical analysis of the characteristics of problem banks. *The Journal of Finance*, *30*(1), 21-36.

Soana, M. G. (2011). The relationship between corporate social performance and corporate financial performance in the banking sector. *Journal of Business Ethics*, *104*(1), 133-148.

Stiroh, K. J. (2004). Do Community Banks Benefit from Diversification? *Journal of Financial Services Research*, 25(2/3), 135–160.

Sun, J., Li, H., Huang, Q. H., & He, K. Y. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, *57*, 41-56.

Switzer, L. N., & Wang, J. (2013). Default risk and corporate governance in financial vs. non-financial firms. *Risk and Decision Analysis*, *4*(4), 243-253.

Switzer, L. N., Tu, Q., & Wang, J. (2018). Corporate governance and default risk in financial firms over the post-financial crisis period: International evidence. *Journal of International Financial Markets, Institutions and Money*, *52*, 196-210.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science*, *38*(7), 926-947.

Thompson, P., & Cowton, C. J. (2004). Bringing the environment into bank lending: implications for environmental reporting. *The British Accounting Review*, *36*(2), 197-218.

Thomson, J. B. (1991). Predicting bank failures in the 1980s. *Federal Reserve Bank of Cleveland Economic Review*, *27*(1), 9-20.

Vallascas, F., & Hagendorff, J. (2013). CEO Bonus Compensation and Bank Default Risk: Evidence from the U.S. and Europe. Financial Markets, Institutions & Instruments, 22(2) Vassalou, M., & Xing, Y. (2004). Default Risk in Equity Returns. *The Journal of Finance*, 59(2), 831–868.

Weber, O. (2014). The financial sector's impact on sustainable development. *Journal of Sustainable Finance & Investment*, 4(1), 1-8.

Weber, O., & Remer, S. (2011). *Social Banks and the Future of Sustainable Finance*. Londo: Routledge.

Youden, W. J. (1950). Index for rating diagnostic tests. Cancer, 3(1), 32–35.

Zavgren, C. V. (1985). Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance & Accounting*, 12(1), 19–45.