GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 2308 – May 2023

Are Basel III requirements up to the task? Evidence from bankruptcy prediction models

Pierre Durand, Gaëtan Le Quang, Arnold Vialfont

Abstract:

Using a database comprising US bank balance sheet variables covering the 2000-2018 period and the list of failed banks as provided by the FDIC, we run various models to exhibit the main determinants of bank default. Among these models, Logistic Regression, Random Forest, Histogram-based Gradient Boosting Classification and Gradient Boosting Classification perform the best. Relying on various machine learning interpretation tools, we manage to provide evidence that 1) capital is a stronger predictor of default than liquidity, 2) Basel III capital requirements are set at a too low level. More precisely, having a look at the impact of the interaction between capital ratios (the risk-weighted ratio and the simple leverage ratio) and the liquidity ratio (liquid assets over total assets) on the probability of default, we show that the influence of capital on this latter completely outweighs that of liquidity, which is in fact very limited. From a prudential perspective, this questions the recent stress put on liquidity regulation. Concerning capital requirements, we provide evidence that setting the risk-weighted ratio at 15% and the simple leverage ratio at 10% would significantly decrease the probability of default without hampering banks'activities. Overall, these results therefore call for strengthening capital requirements while at the same time releasing the regulatory pressure put on liquidity.

Keywords:
Basel III; capital requirements ; liquidity regulation ; bankruptcy prediction models ; statistical learning ; classification.

JEL codes:
C44, G21, G28

Working Paper

Are Basel III requirements up to the task? Evidence from bankruptcy prediction models

Pierre Durand[a], Gaëtan Le Quang[b], Arnold Vialfont[c]

May 2023

Abstract

Using a database comprising US bank balance sheet variables covering the 2000-2018 period and the list of failed banks as provided by the FDIC, we run various models to exhibit the main determinants of bank default. Among these models, Logistic Regression, Random Forest, Histogram-based Gradient Boosting Classification and Gradient Boosting Classification perform the best. Relying on various machine learning interpretation tools, we manage to provide evidence that 1) capital is a stronger predictor of default than liquidity, 2) Basel III capital requirements are set at a too low level. More precisely, having a look at the impact of the interaction between capital ratios (the risk-weighted ratio and the simple leverage ratio) and the liquidity ratio (liquid assets over total assets) on the probability of default, we show that the influence of capital on this latter completely outweighs that of liquidity, which is in fact very limited. From a prudential perspective, this questions the recent stress put on liquidity regulation. Concerning capital requirements, we provide evidence that setting the risk-weighted ratio at 15% and the simple leverage ratio at 10% would significantly decrease the probability of default without hampering banks' activities. Overall, these results therefore call for strengthening capital requirements while at the same time releasing the regulatory pressure put on liquidity.

Keywords: Basel III; capital requirements ; liquidity regulation ; bankruptcy prediction models ; statistical learning ; classification.
JEL classification : C44, G21, G28

[a]Université Paris Est Créteil, ERUDITE, 94010 Créteil Cedex, France.

[b]Univ Lyon, Université Lumière Lyon 2, GATE UMR 5824, F-69130 Ecully, France.

[c]Université Paris Est Créteil, ERUDITE, 94010 Créteil Cedex, France.

1 Introduction

Banking regulation has been profoundly reshaped since the 2007-08 crisis. Capital requirements have indeed been tightened, liquidity ratios added to complement them and bail-in standards implemented to prevent costly bailouts from occurring. While all these instruments are related to one or several features of the 2007-08 crisis – the tightening of capital requirements being the response to banks' massive under-capitalization *prior* to the crisis, the liquidity ratios being the response to the notorious liquidity spirals (Brunnermeier and Pedersen, 2009) that materialized at the end of the 2000s, the bail-in standards answering the need to protect taxpayers from very costly bailouts – no consensus exists concerning the optimal design of these rules and the impact of their joint-implementation. Indeed, there is no consensus concerning the optimal level of capital requirements (Dagher et al., 2020), no more consensus as to whether liquidity and capital standards are actually complements or substitutes (Clerc et al., 2022), and there is a growing fear that the current tendency to multiply the number and the complexity of rules prevents banking regulation from efficiently reaching its goal (Herring, 2018; Haldane and Madouros, 2012).

This paper aims at answering these questions in order to help better assess the current state of banking regulation. To do so, we build bankruptcy prediction models which are applied to a database comprising US bank balance sheet variables covering the 2000-2018 period. The objective is twofold. First, from a methodological perspective, we aim at determining which model among seven performs best at predicting bank default. Second, from a regulatory perspective, the purpose is to disentangle which variables, among a wide range of balance sheet variables, impact the most the probability of default. More precisely, we resort to seven different models: Logit, Random Forests (RF), K-Nearest Neighbors (KNN), Gradient Boosting Classification (GBC), Histogram-based Gradient Boosting Classification (HGBC), Linear Support Vector Classification (Linear SVC) and Multi Layer Perceptron (MLP). Logit, RF, GBC and HGBC perform the best, while Linear SVC, KNN and MLP are lagging behind. The performance of the various models being established, we then resort to several machine learning interpretation tools to cast light on the regulatory questions mentioned above. We provide evidence that:

- Capital is a stronger predictor of bank default than liquidity. When considered in isolation from each other, capital ratios – both the total regulatory capital ratio (TRCR) and the total equity over total assets ratio (TE/TA) – have a negative impact on the probability of default, while the liquidity ratio – liquid assets over total assets (LA/TA) – has a positive impact. Though counter-intuitive, this second result can be explained by the propensity of failing banks to panic sell their illiquid assets, which mechanically improves their liquidity *prior* to default. *Per se*, this result therefore does not contradict the idea at the heart of the Liquidity Coverage Ratio (LCR). When capital and liquidity are considered in interaction, we however notice that the effect of capital on the probability of default completely outweighs that of liquidity. As a consequence, from a prudential perspective, it may be preferable to rule the LCR out to focus on capital requirements. This is in line with a recommendation made by Thakor (2018).
- Basel III capital requirements are set at a too low level. Non-linear models – such as RF, GBC and HGBC – allow to identify two regimes: one characterized by a low capitalization associated with a high probability of default, the other characterized by a larger capitalization associated with a low probability of default. More precisely, we manage to approximate the threshold values of the capital ratios above which banks enter the second regime: when the leverage ratio (TE/TA) is greater than 10% and the risk-weighted capital ratio (TRCR) is above 15%, banks enter the low-risk-of-default regime. Increasing capital requirements above these two thresholds is shown to

have no further impact on the probability of default. As a consequence, we recommend setting the leverage ratio at 10% and the risk-weighted capital ratio at 15%, which is above what Basel III recommends and consistent with Dagher et al. (2020). Increasing capital requirements to these levels would not hamper banks' activities. On the contrary, as shown by Durand and Le Quang (2022), this would actually have a positive impact on banks' return on assets (ROA). The only adverse effect would be on banks' return on equity (ROE), which would indeed decrease. In other words, setting the capital ratios to these levels would generate a benefit (a significant decrease in banks' probability of default) whose cost is a private cost supported only by shareholders. Notice that these would obviously benefit from the lowering of the probability of default, which eventually could compensate the ROE's decrease.

- Overall, the Basel III framework seems sub-optimally complex. The evidence provided in this paper indeed questions the relevance of both liquidity ratios and bail-in standards. In particular, it seems that properly designed capital requirements (TE/TA > 10% and TRCR > 15%) could render both liquidity regulation and bail-in standards superfluous. Indeed, if capital strongly dominates liquidity in determining the probability of default, then regulators must focus on capital regulation and release the regulatory pressure exerted on liquidity. In addition, increasing capital requirements would increase banks' loss-absorbing capacity in a safer manner than bail-in standards do by resorting to questionable eligible liabilities such as coco bonds (Admati et al., 2013; Persaud, 2014). In other words, banking regulation could be rendered less complex and more efficient by focusing more on capital ratios.

These policy recommendations have recently been reality checked. Indeed, both the failure of Silicon Valley Bank (SVB) and that of Credit Suisse in March 2023 comfort the three points that have just been developed. 1) Credit Suisse's LCR was more than met *prior* to default since Credit Suisse's high quality liquid assets covered more than 150% of the expected outflows. Nonetheless, this did not prevent the bank from bankrupting. In fact, in the age of social media the rate at which depositors withdraw their deposits when they start questioning the solvency of their bank exceeds by far the stress scenarios on which the computation of the LCR is based. As a consequence, even though SVB was not subjected to the LCR, such instrument would have hardly been sufficient to deal with the massive outflows of deposits faced by the bank. During contemporary bankruns, social media indeed act as bankrun catalyst (Cookson et al., 2023) so that cash demand appears as unlimited, which questions the relevance of an approach based on the constitution of a limited reserve of liquid assets. On the contrary, the stress should be put on banks' solvency. 2) In this respect, both SVB's and Credit Suisse's capital ratios were below the levels put forward in this paper: in 2022, Credit Suisse's CET1 leverage ratio was slightly greater than 5% (SVB: around 8%), while its risk-weighted ratio slightly above 14% (SVB: around 16%). 3) The conversion of Credit Suisse's AT1 coco bonds had a destabilizing impact since the conversion risk – resulting from their use as going-concern instruments (Perotti, 2023) – was largely underestimated by investors.

The rest of the paper is organized as follows. The next section reviews the literature to which this paper contributes. Section 3 offers some details on the models that are used in the paper. Section 4 describes our database. Section 5 presents the main results. Robustness checks are provided in section 6 and section 7 concludes.

2 Literature review

This article contributes to four different strands of the literature on banking regulation: that on the determinants of bank default, that on the interaction between liquidity and solvency risks, that on regulatory complexity and that on the design of the optimal capital ratio.

2.1 Determinants of bank default

The literature on the determinants of bank default seeks to exhibit which variables are the best predictors of default. Since information on bank default is hard to obtain, part of the literature on this topic resorts to proxies. The main proxies used are the z-score (Demirgüç-Kunt and Huizinga, 2010; Laeven and Levine, 2009), the NPL ratio (Berger and DeYoung, 1997; Delis and Staikouras, 2011; Salas and Saurina, 2002), the CDS spreads (Alter and Schüler, 2012; Soenen and Vander Vennet, 2021, 2022) or the distance to default (Eichler and Sobanski, 2016).

When information on bank default is available, the bankruptcy prediction problem consists in a simple classification problem. Such a problem can be solved either by resorting to a statistical approach or to an intelligent approach (Ravi Kumar and Ravi, 2007). Statistical methods include well-known logistic regressions and are widely used to deal with classification problems, including bankruptcy prediction for firms (Ohlson, 1980; Jones and Hensher, 2004) and for banks (Martin, 1977; Kolari et al., 2002; Imbierowicz and Rauch, 2014). Intelligent methods consist in machine learning techniques such as for instance neural networks or random forests. Specifically, neural networks are largely used in the bankruptcy prediction literature (Ravi Kumar and Ravi, 2007) and are often shown to perform better than logistic regressions (Tam and Kiang, 1990; Tam, 1991; Salchenberger et al., 1992). Fewer papers resort to random forests to predict firms' failures (Zoričák et al., 2020).

The main challenge associated with bankruptcy prediction is that, by definition, bankruptcies are very rare events. Datasets are thus severely imbalanced with one class (that of bankrupted banks) far less represented than the other (that of non-bankrupted banks). There are several ways to deal with imbalanced datasets: either under-sampling or over-sampling (or mixing the two). Under-sampling aims at reducing the size of the majority class to match that of the minority class. It therefore has the inconvenience to delete information, but is in general less computationally demanding than over-sampling. Over-sampling consists in balancing class distribution by replicating items in the minority class, either by exactly replicating some randomly selected items found in the minority class – that is the logic behind Random Oversampling With Replication (ROWR) (Zhou, 2013) – or by creating new items through the Synthetic Minority Oversampling Technique (SMOTE) proposed by Chawla et al. (2002). If under-sampling could sometimes be preferred to over-sampling when the dataset is weakly imbalanced (Zhou, 2013), there is a consensus in the literature that SMOTE is the best option for severely imbalanced datasets (Chawla et al., 2002; García et al., 2012; Zhou, 2013; Haixiang et al., 2017).

The literature on the determinants of bank default has reached a consensus around several financial ratios that are considered as the main determinants of defaults. Those ratios are the rationale behind the computation of the widely used z-score (Altman, 1968; Altman et al., 1977) and behind the CAMELS ratings.[1] In this respect, the literature provides evidence that capital greatly influences the probability of default (Berger and Bouwman, 2013; Parrado-Martínez et al., 2019). In addition to these financial ratios, the literature identifies several structural and environmental factors that significantly impact the default risk. These factors consist, for instance, in the monetary policy led by the central bank (Soenen and

[1] Capital adequacy, Assets quality, Management, Earnings, Liquidity, and Sensitivity constitute the six factors used by regulatory authorities to classify financial institutions according to their soundness.

Vander Vennet, 2021, 2022) or in specificities related to national politics (Eichler and Sobanski, 2016). Ravi Kumar and Ravi (2007) provide an exhaustive review of the variables found as predictors of bank default in papers published from 1968 to 2005.

2.2 Interaction between liquidity and solvency risks

While the empirical literature on the determinants of bank default provides great insight concerning the main determinants of the default risk, it does not allow to precisely disentangle how those determinants interact. In particular, despite the stress put on the liquidity risk after the 2007-08 crisis (Acharya and Mora, 2015), few empirical papers have managed to show how liquidity and solvency risks interact. From a theoretical point of view, this interaction lies at the core of the seminal paper by Diamond and Dybvig (1983). These authors indeed manage to show that the liquidity risk, when modeled as early withdrawals of cash by depositors, can sometimes precipitate the failure of an otherwise solvent bank. If such a run is rendered far less likely in a context where most advanced economies have implemented a deposit insurance scheme, the Diamond-Dybvig framework can nonetheless be adapted so as to model the destabilizing consequences of the strong reliance of banks on very-short term debts (Morris and Shin, 2016).

The empirical literature on the interaction between the solvency and the liquidity risks is rather thin. Imbierowicz and Rauch (2014) explore the relationship between these two risks based on a sample of US banks (4046 non-defaulting banks and 254 defaulting banks) between 1998 and 2010. They show that, if taken separately these two risks always impact the probability of default, the impact of their interaction greatly varies depending on banks' probability of default. DeYoung et al. (2018) provide evidence that capital and liquidity are seen as substitutes by small banks. The idea is that these banks tend to improve the liquidity of their assets when their capital deteriorates: to prevent runs from happening, banks that witness a depletion of their capital switch away from illiquid assets to improve their short-term ability to raise liquidity. The authors therefore conclude that the liquidity risk is naturally mitigated by capital constraints at the level of small community banks, which justifies their exemption from the Basel III liquidity standards.

Disentangling the respective roles of liquidity and capital in banking crises is therefore of the utmost importance for banking regulation. Indeed, depending on the relationship between liquidity and capital, the joint implementation of capital ratios and liquidity ratios could either be seen as a step toward a more stable banking system or as a sub-optimal complexification of banking regulation. However, whether liquidity and capital standards are complements or substitutes to one another is a question that has not reached a consensus in the literature (Clerc et al., 2022). Another way to phrase the issue is to inquire which of the liquidity and of the solvency risks is to be understood as the consequence of the other. In other words, if liquidity dry-ups are shown to sometimes precede capital depletion in banking crises, then adding liquidity constraints to already existing capital constraints makes total sense. On the contrary, if liquidity difficulties only occur when banks are insufficiently capitalized, then the solution is to increase capital requirements and not to implement liquidity ratios. In that case, such ratios only reinforce the complexity of banking regulation without dealing with the main issue at stake: under-capitalization. Empirical evidence seems to suggest that, during the 2007-08 crisis, liquidity dry-ups were most of the time the mere consequence of insufficient capital levels (Thakor, 2018). Indeed, using transaction-level data on short-term unsecured certificates of deposit in Europe between 2008 and 2014, Pérignon et al. (2018) provide evidence that, even if many banks suffered funding dry-ups, no market-wide freeze occurred during this period. Best-capitalized banks actually increased their short-term uninsured funding, while only least-capitalized banks reduced funding. Evidence therefore suggests that stronger capital requirements by themselves could have prevented liquidity dry-ups from happening.

2.3 Regulatory complexity

Regulatory complexity refers to the idea that the implementation of a specific body of rules can generate important economic costs either because those rules are very numerous or because they are particularly difficult to implement. According to this definition, the complexity of banking regulation has importantly increased following the 2007-08 crisis. In fact, the very history of banking regulation seems to be one of ever increasing complexity: Basel I is fully described within 28 pages, whereas Basel II goes over more than 330 pages and the initial version of Basel III over more than 600 pages (Haldane and Madouros, 2012; Herring, 2018). If it can be argued that sophisticated rules are only the natural response to an increasingly complex banking system, the study of complex systems however leads to the very opposite conclusion: "complex systems typically call for simple control rules" (Haldane, 2011, p.2). In fact, as Kane's description of the regulatory dialectic made it quite clear (Kane, 1977, 1981), regulated institutions are incentivized to adapt their behaviors so as to avoid regulatory restrictions. With a lag, regulators are then incentivized to complexify these restrictions in order to make them less easily avoidable, which in turn nourishes new avoidance strategies, etc. In the end, both the regulation and the market it regulates are very complex. In such a game, it seems that regulators most of the time end up captured by the complexity of the industry that they are supposed to regulate. Responding to complexity by simple rules is a way to prevent such "regulatory capture by sophistication" (Hakenes and Schnabel, 2013).

Current banking regulation suffers from two types of complexities. On the one hand, following the 2007-08 crisis numerous new rules were added to the main capital constraints. Liquidity ratios as well as bail-in standards are examples of new standards most banks must now comply with. On the other hand, the very implementation of capital regulation is complex. Indeed, both the definition of the numerator and that of the denominator of the regulatory capital ratio are subject to complexity. Regarding the denominator, the question is how to measure the risk associated with the asset side of banks' balance sheets. The option retained by regulators is to capture this risk through an estimation of the risk-weighted assets (RWA). Such a definition poses at least two problems. The first one is the very estimation of the RWA. Given the complexity of banking activities, banks are allowed, under certain conditions, to resort to the Advanced Internal Rating-Based (A-IRB) approach to compute their RWA. Such an internal approach has been shown to allow banks to underestimate their capital requirements (Mariathasan and Merrouche, 2014). Basel III has thus moved forward by further constraining the use of such internal models. The second problem arising when the capital ratio rests on RWA is that it is not sure that this latter measure efficiently captures bank default risk. As Blundell-Wignall and Roulet (2013) indeed show it, while the un-weighted leverage ratio is a strong predictor of the default risk, the Basel Tier 1 risk-weighted ratio is not. This result is corroborated by Mayes and Stremmel (2014) and Vazquez and Federico (2015). Regarding the numerator of the capital ratio, the question is that of the definition of regulatory capital. While Basel III has moved forward by eliminating certain hybrid instruments from the definition of regulatory capital, it still rests on a distinction within Tier 1 capital between Core Tier 1 (CET1) capital and Additional Tier 1 (AT1) capital. The former refers to common equity and must at least constitute 4.5% of the RWA, while the latter designates hybrid instruments, such as contingent convertible (coco) bonds, and can constitute up to 1.5% of the Tier 1 regulatory minimum constraint. This twofold definition of regulatory Tier 1 capital poses at least two problems. The first one is that there is strong evidence that common equity proves far more effective than hybrid instruments at absorbing losses (Admati et al., 2013; Demirguc-Kunt et al., 2013). If the objective of Tier 1 capital is to allow banks to deal with difficulties as going-concerns, it would therefore have been preferable to focus the definition of Tier 1 capital on the instrument exhibiting the largest loss-absorbing capacity, which is equity. The second problem is that the hybrid instruments eligible as AT1 capital can eventually have an adverse impact on financial stability. Empirical evidence (Bologna et al., 2020) and theoretical works (Goncharenko, 2021; Le Quang, 2022) indeed show that coco bonds could have harmful unintended consequences.

2.4 Optimal capital ratio

The question of the optimal capital ratio lies in the continuity of that of the optimal financial structure. Modigliani and Miller (1958) famously answered this question by stating that no optimal financial structure can be found. The rationale behind this idea is that when more equity is used, the volatility of the return on that equity decreases and as a consequence the safety of debt increases. Thus, the required return on both sources decreases so that the weighted average cost of finance remains unchanged. The Modigliani-Miller theorem has given birth to a huge literature questioning its relevance in general, and for particular industries such as banks (Miller, 1995). Whether this theorem fully holds or not is out of the scope of this article, but the extent to which a change in banks' financial structure modifies their average cost of finance is an important question as far as banking regulation is concerned.

Indeed, the banking industry opposes any strengthening of equity requirements on the basis that they would unambiguously increase their funding cost and thus force them to increase their loan rates, which would impose a cost on society as a whole. Gambacorta and Shin (2018) provide evidence that contradicts this argument. They indeed show that an increase of 1 percentage point in the equity ratio (equity over total assets) yields a decrease of 4 basis points in the cost of debt for a sample of banks located in the G10 countries. Kashyap et al. (2010) provide evidence that, in the long-run, the impact of "substantially heightened" capital requirements on loan rates is expected to be weak. They indeed find that a 10 percentage-point increase in the capital requirement increases loan rates between 25 and 45 basis points. As Admati et al. (2013) point it, the social cost associated with an increase in equity requirements is thus expected to be low. In the same vein, Durand and Le Quang (2022) provide evidence that an increase in the equity ratio has a positive impact on the return on assets (ROA), but a negative impact on the return on equity (ROE) above a threshold of 8% of the equity ratio. They conclude that the sole cost associated with an increase in equity requirements above this threshold is a private cost supported by shareholders.

The question of what exactly is the value of the optimal capital ratio is however rarely explicitly addressed by the literature. Oddly enough, while under Basel II the minimum ratio of risk-weighted Tier 1 capital was set at 4% and the minimum total risk-weighted capital ratio (i.e. Tier 1 plus Tier 2) at 8%, the Basel Committee never provided any rationale for these minimum levels (Herring, 2018). Similarly, no economic rationale is provided for the revised minima put forward by Basel III. Recall that these constrain banks to comply with the following ratios: a minimum 8.5% Tier 1 ratio made at least of 7% of CET1 capital and a minimum Tier 1 leverage ratio (i.e. not risk-weighted) of 3%. Miles et al. (2013) provide evidence that these requirements are not optimal. They indeed show that the optimal amount of capital is likely to be at least twice as great as that defined by Basel III. More precisely, they conclude that a ratio of equity over RWA of 20% is optimal, which corresponds to a financial structure resting on 90% to 93% of debt and 7% to 10% of equity. Setting minimum equity requirements between 7% and 10% of total assets would therefore allow banks to face crises more efficiently without hampering their activity. The idea that Basel III capital requirements lie under their optimal values is supported by the literature. Using a DSGE model, Karmakar (2016) indeed shows that doubling the equity ratio from 8% to 16% is welfare-improving. Egan et al. (2017) show that below the $15 - 18\%$ range, capital requirements are insufficient so that loss of welfare and financial instability arise. Barth and Miller (2018) find an optimal equity ratio of 19%, which corresponds to a risk-weighted ratio around 25%. The "seawall approach" proposed by Dagher et al. (2020) yields a more conservative optimal risk-weighted ratio lying between 15% and 23%. Based on an empirical study concerning the euro area, Soederhuizen et al. (2021) recommend to set the minimum risk-weighted capital ratio at 22%. Mendicino et al. (2021) also find that Basel III capital requirements are set at a too low level and recommend setting them at 15%.

2.5 Contribution of the paper

This article contributes to the above-presented literature in at least three respects:

- **Methodology.** Among the seven models that are run, Logit, RF, GBC and HGBC perform the best. Linear SVC, MLP and KNN are, on the contrary, lagging behind.
- **Optimal capital requirements.** We provide evidence that the risk-weighted capital ratio and the leverage ratio actually complement each other. Indeed, to lower the most the probability of default, it appears that the best option is to set the leverage ratio at 10% and the risk-weighted ratio at 15%. Above these levels, no further impact on the probability of default is found. One hypothesis allowing to make sense of the complementarity between these two ratios is to understand the leverage ratio as providing a floor preventing manipulations of the risk-weighted ratio from releasing too much the capital constraint, and this latter as providing incentives for banks not to engage in overly risky investment strategies. The values of the capital ratios found in this paper are consistent with those found by Karmakar (2016); Egan et al. (2017); Dagher et al. (2020); Mendicino et al. (2021).
- **Interaction between liquidity and capital.** We provide evidence that liquidity is not *per se* a strong driver of bank default. Consistently with Pérignon et al. (2018), we therefore question the idea that liquidity risk can act as a driver of financial instability independently of banks being under-capitalized. In fact, we notice that liquidity has a positive impact on the probability of default: failed banks are likely to exhibit more liquid asset portfolios than unfailed banks. In line with the evidence put forward by DeYoung et al. (2018), this may be because banks whose capital deteriorates are forced to sell their illiquid assets. As a consequence, banks whose situation deteriorates up to default might very well be more liquid than sound banks. Such evidence questions the implementation of the LCR in addition to capital requirements. In line with Thakor (2018), we therefore recommend to focus on capital requirements and to release the regulatory pressure put on liquidity. This would allow to reduce the complexity of banking regulation, to reduce the regulatory pressure exerted on safe assets (Caballero et al., 2017), and to prevent the shortening of banks' investment time horizon, which could prove detrimental to the funding of low-carbon sectors as argued by Campiglio (2016).

3 Models, performance measures and interpretation

In this section, we briefly present the methodology this paper is based on. In particular, we provide insight on how the SMOTE procedure works, on the general idea behind each model, on the performance measures we rely on, and finally on the tools we use to draw economic interpretations out of the models. These all aim at estimating a function $f(.)$ defined as follows:

$$\Pr(y = 1|X = x) = f(X) + \epsilon,$$

where $\Pr(y = 1|X = x)$ is the probability that a specific bank belongs to class 1 (default) knowing its specific characteristics $X = x$ (where x is the realization of X), and ϵ the error term. Since data are unbalanced,[2] we find ourselves in a cross-sectional analysis: when $t \neq t'$, a bank j observed at date t is considered as different from the same bank as observed at date t'. To take time dynamics into account and allow for causal relationship to be established, a one-period time lag between X and y is introduced. The function $f(.)$ estimated for bank j at time t can thus be rewritten as follows:

$$\Pr(y_{j,t} = 1|X_{j,t-1} = x_{j,t-1}) = f(X_{j,t-1}) + \epsilon_{j,t}.$$

The following procedure is applied throughout the paper to ensure models' interpretability: 1) k-cross-validation on out-of-sample macro recall is used to determine the hyperparameters of the models, 2) Shapley values are then computed to provide an idea of the significance and of the nature of the impact of the considered variables on the probability of default, 3) partial dependence plots are provided to assess the marginal impact of a given feature on the predicted probability of default.

3.1 SMOTE

Introduced by Chawla et al. (2002), the Synthetic Minority Over-sampling Technique (SMOTE) is inspired by Ha and Bunke (1997). It is built in such a way that it replicates the initial data distribution. More precisely, SMOTE uses the k-nearest neighbors of all the instances found in the minority class (failed banks in our case) to synthesize new minority class instances: synthetic observations are created on the line between existing ones. Using nearest neighbors ensures that the distribution of the balanced sample is the same as that of the original imbalanced sample.

3.2 Models

We very briefly detail here the main idea behind each of the seven models that are used in the paper. In addition, we highlight their main *hyperparameters* (reported in appendix B) which constrain the optimization process and are chosen according to performance scores (detailed in section 3.3). Models are implemented in Python thanks to Scikit-learn (Pedregosa et al., 2011).

Logistic regression (LR) (or **Logit**) is a linear model that estimates the probability with which each observation enters either of the classes of the outcome variable. In addition to the standard statistical model, one can also include some penalties into the cost function (Lasso, Ridge or a combination of the two: Elastic Net) which shrink the estimated coefficients (Hastie et al., 2009). The intensity of this regularization is itself a hyperparameter.

Random forest (RF) (Breiman, 2001) is an ensemble method based on the aggregation of decision trees (bagging). More precisely, trees are run in parallel so that the prediction of the model consists in the average of their outcomes. In order to avoid overfitting, trees can be constrained by limiting their maximum depth or the maximum number of variables (or features) considered when splitting observations, as well as increasing the minimum number of samples a leaf should have. The size of the forest is an

[2] While "imbalanced" refers to the fact that the minority class (failed banks) is far less represented than the majority class (unfailed banks), "unbalanced" here refers to the fact that the total number of observations is inferior to the total number of banks times the number of periods.

important hyperparameter of this model: the number of aggregated trees (or estimators) affects the ability of the model to generalize to unobserved data as opposed to single decision trees.

Gradient boosting classifier (GBC) (Friedman, 2001) is based on the aggregation of simple models (here decision trees). While RF simply aggregates several trees with bagging, gradient boosting draws a prediction out of a simple model and improves it by implementing another simple model (boosting). This procedure is repeated several times, focusing on residual errors of the previous model. As compared to RF, an additional hyperparameter is the learning rate that affects how fast the algorithm improves the decision trees, with a lower value usually limiting overfitting. We also consider **Histogram-based Gradient Boosting classifiers (HGBC)** which implement optimized versions of GB, namely XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017).

Support vector classifier (SVC) (Boser et al., 1992) (or Support vector machine) consists in finding the hyperplane in a N-dimensional space (N being the number of features) that best classifies the data points. One of the main hyperparameters is the regularization parameter (denoted C) that impacts the margins of error during the training: a larger C restricts the number of misclassifications but also reduces out-of-sample performances. In addition, one might allow for nonlinear separations of the dataset with the risk of overfitting as the polynomial degree increases.

Multi-layer perceptron (MLP) is a class of artificial neural network (McCulloch and Pitts, 1943; Hastie et al., 2009). We only consider here this feedforward type of network that may be composed of one or several layers of neurons, activated according to some usual functions, namely logistic, hyperbolic tangent (tanh) or Rectified Linear Unit (ReLU). The flexibility of MLP comes at the cost of an important risk of overfitting that one should control for with a combination of hyperparameters: a shrinkage of parameters, early stopping of the training and/or a maximum number of iterations in the backpropagation process.

K-nearest neighbors (KNN) is a non-parametric supervised learning classifier which classifies points based on their proximity. It is computed such that an observation is assigned to the class that is the most frequent within its neighbors. In addition to the considered number of neighbors K, one may also choose the metric as well as attribute uniform or higher weights to closer points.

3.3 Performance measures

Each of the models presented above generates predictions that allow to classify banks in either of the two considered categories (i.e. failed or unfailed). The variety of predictors implies the need to efficiently compare their performance. For a classification task, the usual criteria rely on the well-known confusion matrix (Hand, 2012) that consists in: the number of true positives (TP, failed banks identified as failed banks), the number of true negatives (TN, unfailed banks correctly identified), the number of false positives (FP, unfailed banks identified as failed ones), and the number of false negatives (FN, failed banks identified as unfailed ones). From these four categories, immediate performance scores can be computed: *accuracy* that is defined as the proportion of accurate predictions among all predictions, *recall* that is defined as the proportion of a given class that is properly identified, *precision* that is defined as the proportion of predictions for a given class that actually belongs to this class, and $F1$-score that is computed as the harmonic mean of recall and precision.[3]

There generally is a tradeoff between recall and precision, especially when the considered model is highly flexible and that the convergence criterion leads to focus on one of these scores. As an extreme

[3]These scores are given by the following formula: Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$, Recall = $\frac{TP}{TP+FN}$, Precision = $\frac{TP}{TP+FP}$ and $F1$-score = $2\frac{\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}}$.

example, a model may very well classify all instances in a given class and therefore reach perfect recall. In that case, such model would display a very low precision. The $F1$-score tackles this problem by assessing models with the average of these two scores. However, with severely imbalanced datasets come two additional challenges. First, the accuracy score overestimates the performance of the model if the overrepresented class is well predicted. Second, precision and recall scores imply another tradeoff regarding the class to predict. Indeed, even in a binary classification problem, the performance measure of the model is affected by the class defined as a *positive* instance. In our case, we do observe that models exhibit very high $F1$-scores when considering the prediction of unfailed banks as they easily predict a large number of TP as compared to both FN and FP. On the contrary, considering very low $F1$-scores in the prediction of failed banks actually underestimates the performance of models. As a result, a relevant performance measure is the average between the proportion of 1 (failed banks) that is well predicted and the proportion of 0 (unfailed banks) that is well predicted, which is the very definition of the average between the recall scores of the two classes, also known as *macro recall*. It is worth noticing that, when the classification problem is a binary problem, the area under the ROC curve can be computed the same way as the macro recall (Fawcett, 2006; Sokolova and Lapalme, 2009; Muschelli, 2020). In that case, these two metrics are thus equal.

3.4 k-cross validation

A complementary point should be made when comparing the performance of the models: they may not generalize identically on unobserved data. This is where cross-validation is useful (Ojala and Garriga, 2010). k-cross validation is a resampling method that uses different portions of the dataset to train and evaluate a given model on different iterations. More precisely, it consists in splitting observations into k folds and training k times the same model leaving each time a different kth part of the data to perform score measures. Cross-validation helps in the determination of the relevant classification model in two ways. First, by averaging each score reached on the k validation sets any score becomes a more robust performance measure. Second, it helps to identify hyperparameters that limit the gap between the train and validation scores, that is to limit overfitting.

3.5 Interpretation

As already mentioned, we resort to two tools to draw economic interpretation out of the models: Shapley value and partial dependence plots.

Shapley value. For all models, we resort to the Shapley value (Shapley, 1953; Strumbelj and Kononenko, 2013). The rationale behind the computation of this latter is grounded in game theory: all features are assumed to be players engaged in a game where the payout is the prediction. In this context, the Shapley value indicates how this payout is distributed among the features given their contribution. See Molnar (2020) for more details on computation. In addition, as a robustness check, we rely on permutation feature importance (Molnar, 2020).

Partial dependence plots (PDPs). Economic interpretations of our results mostly rely on PDPs (Friedman, 2000; Hastie et al., 2009). PDPs average the Individual Conditional Expectation (ICE) of all individuals. Considering the i-th individual and the variable X_j and fixing all the other variables to their level taken for the i-th individual, the ICE corresponds to the predictions of the model when X_j varies from its minimum to its maximum value with step k. However, PDPs can be biased when features are strongly correlated. To take potential biases into account, ALEs are provided in section 6.2.

4 Data and descriptive statistics

4.1 Data

Our sample consists in US bank balance sheet variables covering the 2000-2018 period. Data come from the FitchConnect database. Failed banks are identified thanks to the list of failed banks as provided by the Federal Deposit Insurance Corporation (FDIC). This list gathers failures of banks that were covered by the FDIC deposit insurance scheme. After data treatment for missing values, we managed to keep 23 variables (see appendix A), 4707 banks among which 454 have defaulted. Table 1 displays the evolution of the number of banks and defaults per year.

Table 1: Evolution of the number of observations and defaults per year

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Nb. obs.	3866	3961	4029	4067	4076	4305	4337	4435	4516	4465
Nb. defaults	0	1	1	3	0	0	1	18	115	138

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018
Nb. obs.	4361	4291	4226	4207	4194	4198	4191	4196	4197
Nb. defaults	80	42	21	13	7	5	6	0	3

Source: Authors' calculations.

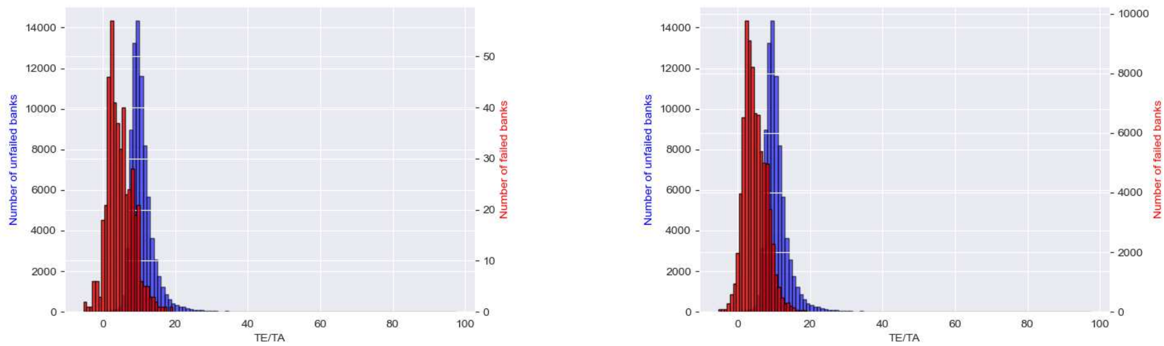
4.2 Descriptive statistics

4.2.1 Variables' distributions

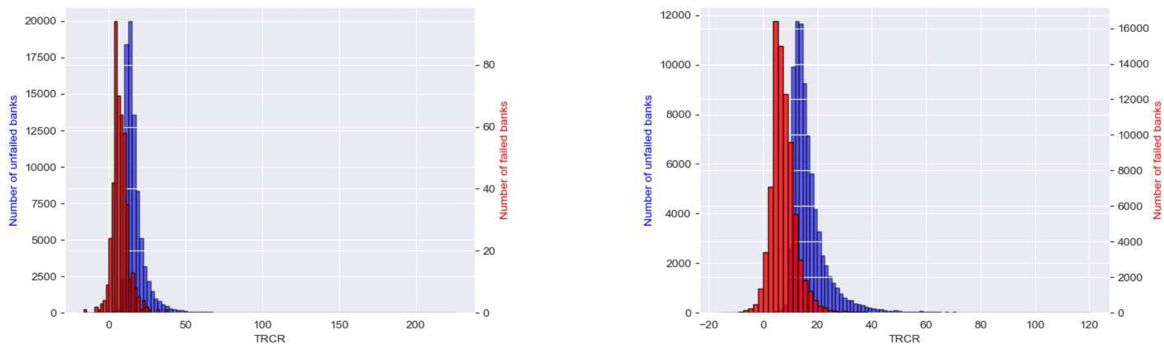
The main purpose of the models that are presented in this paper is to classify banks, that is to separate unfailed from failed banks. We therefore focus on the characteristics of these two groups in order to exhibit significant differences. In particular, we concentrate on the following variables: total equity over total assets (TE/TA), total regulatory capital ratio (TRCR) and liquid assets over total assets (LA/TA). These variables account for regulatory ratios aiming at ensuring banks' solvency (TE/TA and TRCR) and liquidity (LA/TA). Variables' distributions are presented in Figure 1. In line with the CAMELS approach, we notice that better capitalized banks (i.e. banks with higher values of TE/TA and TRCR) are less likely to go bankrupt than less capitalized banks. On the contrary, there is no clear difference between failed and unfailed banks from the viewpoint of the liquidity of their asset portfolios. In fact, it seems that failed banks look slightly more liquid than unfailed banks. This unclear relationship between banks' probability of default and banks' liquidity may however by rendered intelligible keeping in mind the two contradictory mechanisms at work. On the one hand, in line with the CAMELS approach, liquid banks are less likely to go bankrupt since it is easier (less costly) for them to cope with the liquidity outflows required by their creditors. On the other hand, failing banks – or banks close to default – are very often constrained to panic sell their illiquid assets, which mechanically increases the liquidity of their asset portfolios *prior* to their default. As a consequence, failed banks are likely to be characterized by large LA/TA ratios because of the panic sales they may have been forced to engage in. It is moreover worth noticing that applying SMOTE to our dataset does not modify the distributions of the variables that are here presented.

Figure 1: Main regulatory variables' distributions before (left) and after (right) SMOTE

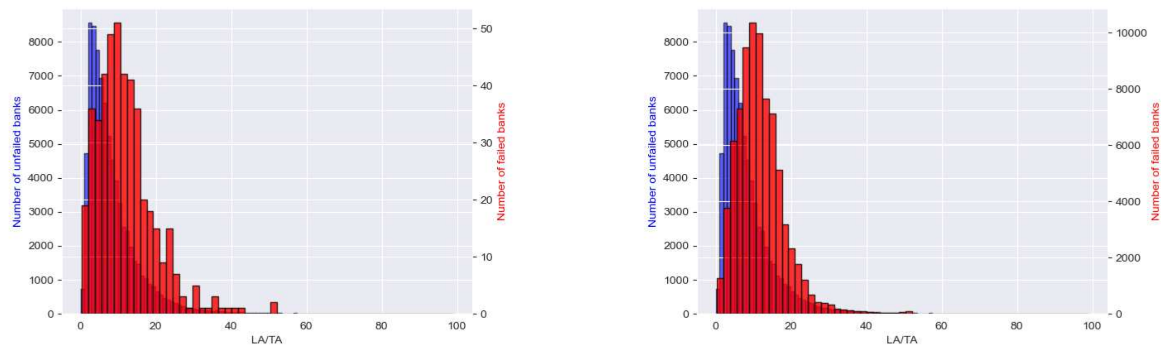
(a) TE/TA



(b) TRCR



(c) LA/TA



Source: Authors' calculations.

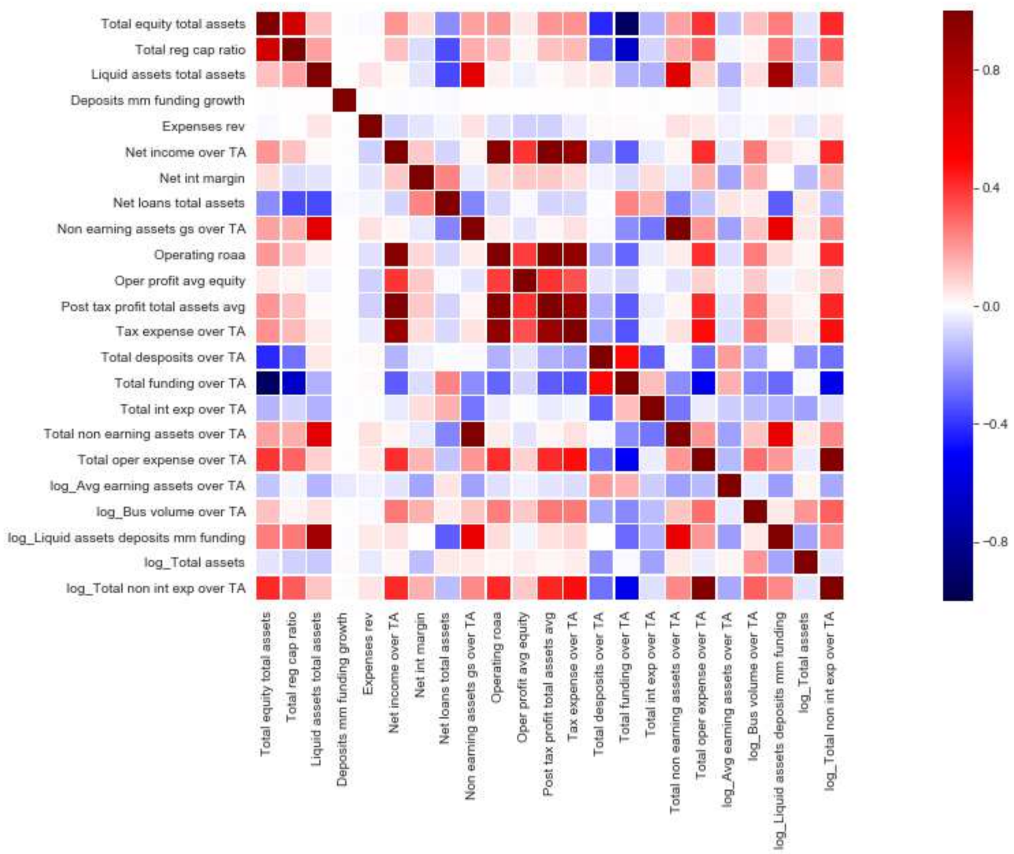
4.2.2 Descriptive statistics before and after SMOTE

We conduct comparison tests on means and variances to assess the impact of the SMOTE procedure on the distribution of the predictors. Specifically, we compare the statistical difference between the means (Student test) and variances (Fisher test) of the distribution of all the predictors before and after applying SMOTE for both the full sample and a subsample comprising only failed banks (i.e., the very group on which the SMOTE procedure is applied). Our results show that when considering the full dataset, several variables (among which ROAA, TRCR, TE/TA and LA/TA) exhibit statistical differences in means. This can be explained by the fact that the SMOTE procedure naturally increases the weight of the mean of the over-sampled class. As a consequence, the means of the variables that significantly impact the probability of default (i.e., the variables whose means are expected to be significantly different between the two classes) are expected to change after the SMOTE procedure is applied. In other words, if the two classes are characterized by radically different means, SMOTE will likely accentuate this difference. When specifically focusing on the subsample which comprises only failed banks, we notice that only two variables display significant differences in means: Expenses revenues and Net interest margin. These differences are due to outliers. In both samples and for a large part of the predictors, we notice significant differences in variances to a one percent critical threshold. Precisely, the variance of the over-sampled dataset is significantly smaller in most cases. This was also expected since the SMOTE procedure creates points based on the existing distribution. As a consequence, all the points created are comprised between the minimum and the maximum of the original distribution, which necessarily decreases the variance of the latter.

4.2.3 Variables' correlations

Since Partial Dependence Plots (PDPs) can be biased when independent variables are highly correlated with each other, we need to have a look at variables' correlations. Figure 2 provides a correlation heatmap. As can be seen, some variables are strongly correlated with each other. To avoid any bias in our estimations, we remove those variables that are the most correlated. More precisely, the following variables have been removed: Net income over Total Assets, Operating profit avg equity, Post tax profit total assets avg. We are therefore left with 20 variables.

Figure 2: Correlation heatmap



Source: Authors' calculations.

5 Results

5.1 Models' performance

Let us deal first with the question of the performance of the various models we resort to. To do so, we compute the performance measures presented in the methodological section. More specifically, since the purpose of the paper is to inquire the determinants of bank default, we pay particular attention to recall. Since models can boost recall by simply over-identifying the number of failed banks, we do not exactly focus on recall but on macro recall. This latter measure is the simple average between the recalls computed for each class (failed banks, unfailed banks). In addition, in the case of a binary classification problem, this measure is equal to the area under the ROC curve (AUROC). For each model, Table 7 in appendix B presents the values of the hyperparameters which maximize the macro recall. Table 2 displays the values of the recalls and macro recalls for the models. Paying attention to the macro recalls computed on the test sample, we notice that four models display scores above 88%: Logit,[4] RF, GBC and HGBC. KNN, Linear SVC and MLP are lagging behind. KNN and MLP are particularly bad at identifying failed banks (low recalls), while Linear SVC over-identifies failed banks, which explains the large recall.[5]

Table 2: Models' performance

	Logit		RF		KNN		GBC	
	Train	Test	Train	Test	Train	Test	Train	Test
Recall	84.55	82.64	90.93	80.99	100	71.90	90.73	80.99
Macro recall (AUROC)	90.12	89.13	93.74	88.63	100	82.29	93.49	88.52

	HGBC		Linear SVC		MLP	
	Train	Test	Train	Test	Train	Test
Recall	96.80	80.99	93.92	86.77	90.04	76.56
Macro recall (AUROC)	96.82	88.72	84.74	81.25	92.94	86.11

Source: Authors' calculations.

Let us have a look at the confusion matrices (Table 3) for the four models that perform best. Doing so, we notice that these models manage to properly identify the vast majority of the failed banks in the test sample: among the 121 failed banks, Logit identifies 100 defaults, HGBC, RF and GBC identify 98 defaults. What is however less convincing is the number of false positives (FP) : 1047 for Logit, 888 for RF, 848 for HGBC and 941 for GBC. Recall however that our dataset is made of bank-year observations. A failed bank is thus identified as a true positive only the year before the actual default occurred. As a consequence, at date $t-2$ this specific bank is identified as an unfailed one even if it might already exhibit the characteristics of a failed bank. To disentangle what the false positives displayed in Table 3 are made of, we try to identify whether or not they in fact consist in banks that at some point go bankrupt. Results are presented in Table 4. More precisely, we 1) identify all FP (Nb. of FP), 2) identify within these FP

[4]To avoid the problems that arise when too many variables are considered in a logistic regression, only five variables are considered in the Logit: ROAA, TRCR, LA/TA and TE/TA. Notice that the performance of the model is only marginally improved when all the variables are taken into consideration.

[5]Linear SVC's precision on the test sample is 1.77%.

the actual number of banks keeping in mind that the same bank can be wrongly identified several times (Nb. of banks in FP), 3) look among these banks for those that actually go bankrupt at some point in the considered time period (Nb. of failed banks in FP), 4) compute the proportion these banks represent among all the banks at least once wrongly identified as failed ones (Prop. of failed banks in FP).

Table 3: Confusion matrices

Train	Logit		HGBC		RF		GBC	
SMOTE	Pred. 0	Pred. 1	Pred. 0	Pred. 1	Pred. 0	Pred. 1	Pred. 0	Pred. 1
True 0	53353	2396	53988	1761	53832	1917	53660	2089
True 1	8609	47140	1779	53970	5055	50694	5163	50586
Test	**Logit**		**HGBC**		**RF**		**GBC**	
no SMOTE	Pred. 0	Pred. 1	Pred. 0	Pred. 1	Pred. 0	Pred. 1	Pred. 0	Pred. 1
True 0	22868	1047	23067	848	23027	888	22974	941
True 1	21	100	23	98	23	98	23	98

Source: Authors' calculations.

Table 4: False positives (FP)

Model	Nb. of FP	Nb. of banks in FP	Nb. of failed banks in FP	Prop. of failed banks in FP (%)
Logit	1047	731	153	14.61
HGBC	848	650	159	18.75
RF	888	668	156	17.56
GBC	941	692	164	17.43

Source: Authors' calculations.

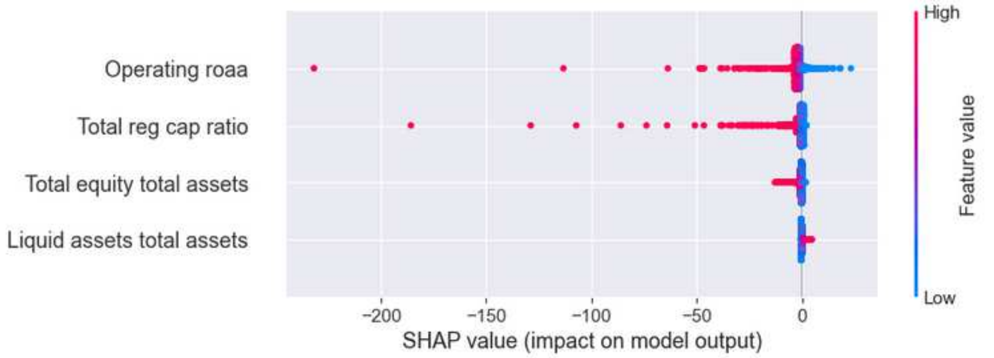
Having a look at Table 4, we notice that, depending on the considered model, between 14.61% and 18.75% of the false positives are actually previous observations of banks that do eventually go bankrupt. Even if the models make mistakes by identifying as failed banks some banks that are actually sound, a significant proportion of these mistakes concern banks that at some point do indeed go bankrupt.

5.2 Determinants of bank default

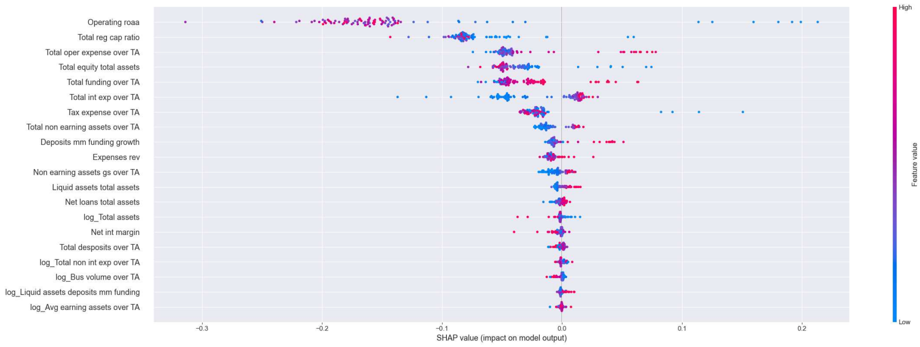
To determine which variables impact the most the probability that banks go bankrupt, we compute the Shapley value (Shapley, 1953; Strumbelj and Kononenko, 2013). Figure 3 displays the Shapley values for Logit, RF, GBC and HGBC. Features are ranked according to the importance of their impact on the prediction made by the models. We focus in particular on variables TRCR, TE/TA and LA/TA. Having a look at Figure 3, we first notice that, in all of the cases, operating ROAA is the most important feature, which is consistent. In addition, in line with the CAMELS approach, we remark that capital and liquidity are most of the time significant predictors of bank default. Table 5 summarizes the ranking of the three variables regarding their impact on the probability of default.

Figure 3: Shapley values

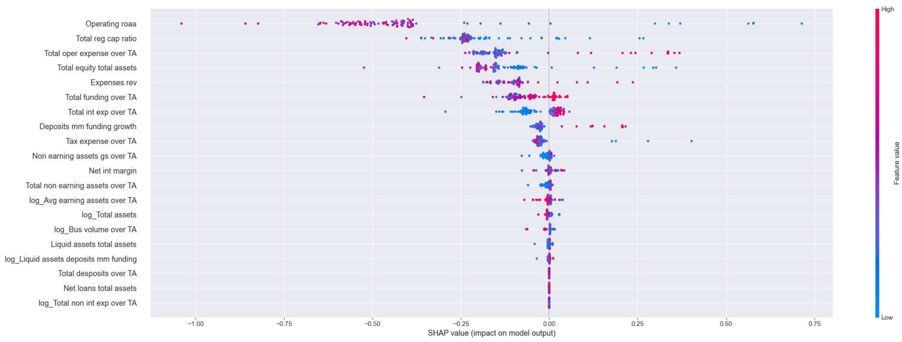
(a) Logit



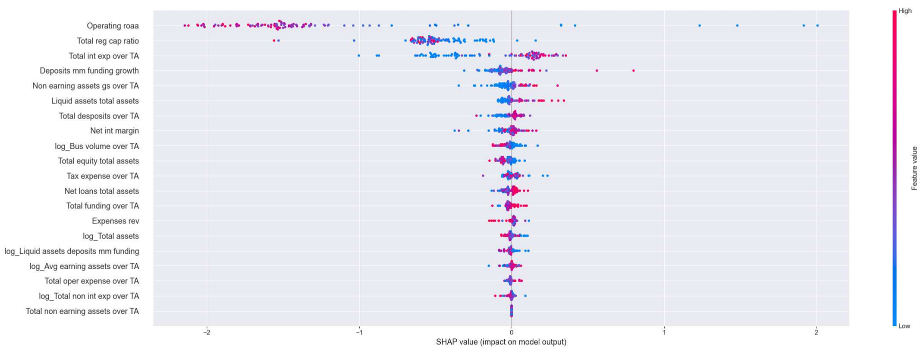
(b) RF



(c) GBC



(d) HGBC



Source: Authors' calculations.

Table 5: Capital versus liquidity

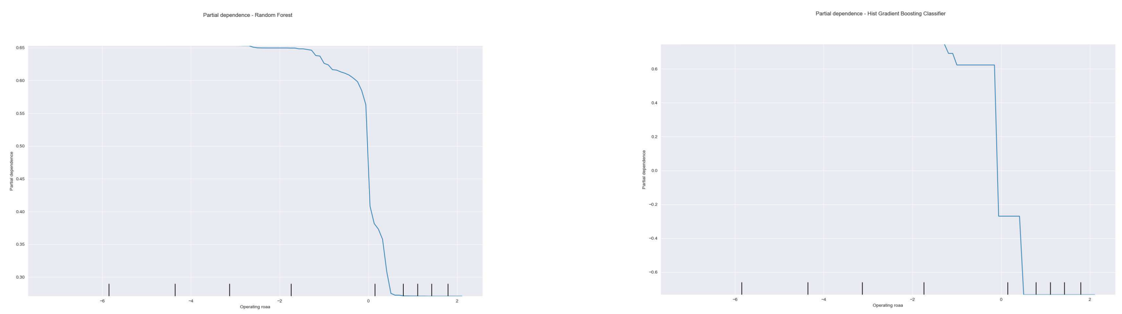
Model	Features' ranking
Logit	TRCR > TE/TA > LA/TA
HGBC	TRCR > LA/TA > TE/TA
RF	TRCR > TE/TA > LA/TA
GBC	TRCR > TE/TA > LA/TA

We notice that variables associated with capital (TRCR and TE/TA) prove better predictors of default than LA/TA. This point will be further developed in section 5.4. Concerning the relative importance of TRCR and TE/TA, it seems that TRCR is a greater determinant of the probability of default than TE/TA. Shapley values also allow to have a look at the nature of the impact of each variable on the probability of default. In line with the CAMELS approach, we notice that variables accounting for capital are associated with a negative impact on the probability of default. Having, for instance, a look at RF we notice that, for most of their values, TRCR and TE/TA have a negative impact on the probability of default: Shapley values mostly lie on the left of the 0. As for LA/TA, we notice that values associated with a positive impact on the probability of default are the largest values of LA/TA, which is in line with the idea that failed banks panic sell their illiquid assets *prior* to default.

The nature of the impact of the variables on the probability of default can be further specified thanks to PDPs. Figure 4 plots PDPs for ROAA as derived from RF and HGBC. We focus here on the best performing non-linear models since they allow to exhibit threshold effects. In particular, we notice that PDPs display two regimes depending on the value of ROAA: when ROAA < 0, the probability of default is high, while it is low when ROAA > 0. This result is not surprising at all since ROAA is defined as net income over total assets and thus provides a direct measure of the solvency of the bank. As a consequence, these PDPs are a good sign as to the ability of the machine learning interpretation tools we resort to in this paper to generate economic meaningful outputs.

Figure 4: Partial Dependence Plots (PDPs) – ROAA

(a) RF (b) HGBC

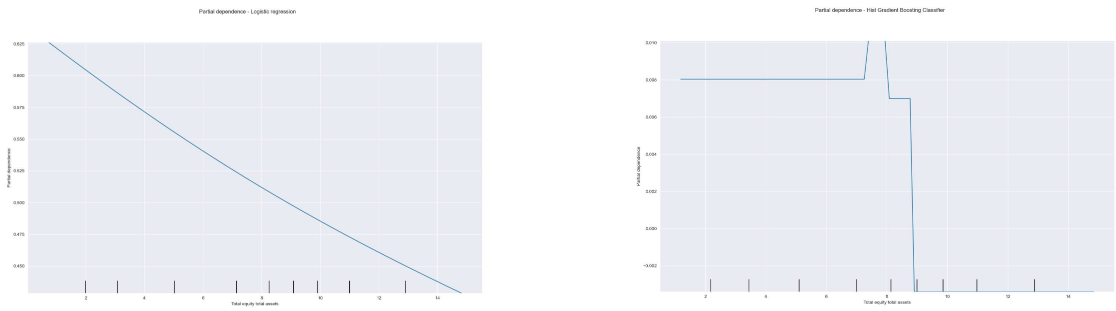


Source: Authors' calculations.

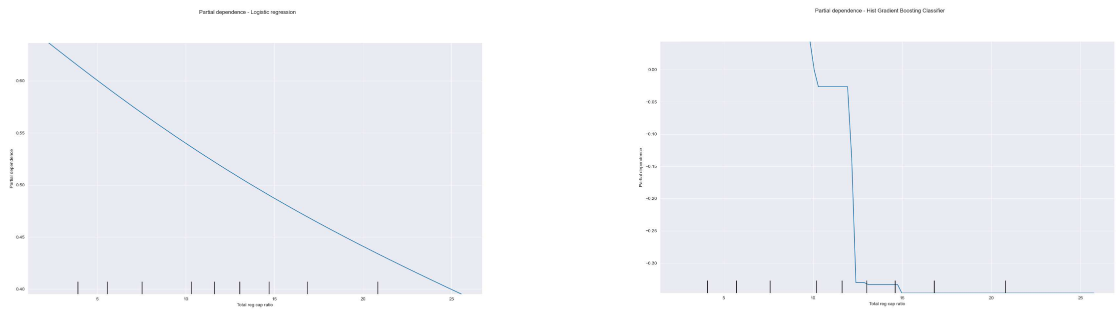
Let us now turn to the nature of the impact of the variables we consider on the probability with which the models identify a bank as a failed one. According to the descriptive statistics presented in section 4.2.1, we expect TE/TA and TRCR to have a negative impact on the probability of default, while the impact of LA/TA is expected to be ambiguous. Figure 5 displays PDPs for Logit and HGBC. Results for GBC and RF are qualitatively the same. We notice that TE/TA negatively impacts the probability that a bank goes bankrupt: for larger values of TE/TA, the probability of default is indeed lower than for smallest values of TE/TA. The same goes for TRCR. This is consistent with the CAMELS approach. Finally, LA/TA has a positive impact on the probability with which a bank is identified as a failed one, which is consistent with the idea that banks facing difficulties very often engage in panic sales and therefore end up holding very liquid asset portfolios.

Figure 5: Partial Dependence Plots (PDPs) – Logit (left) and HGBC (right)

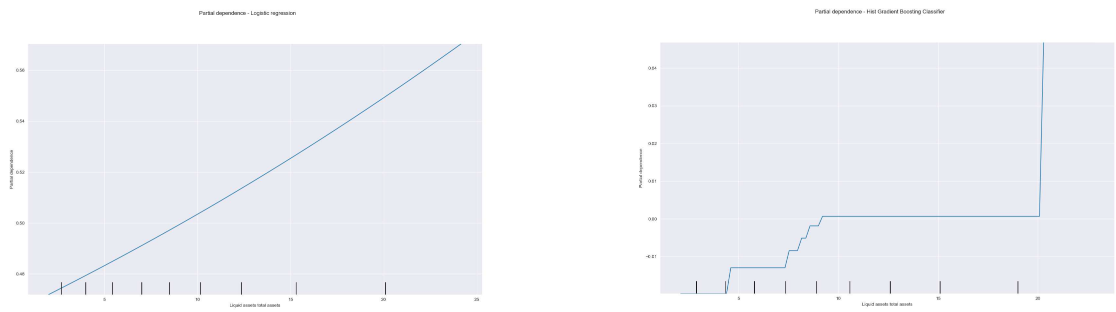
(a) TE/TA



(b) TRCR



(c) LA/TA



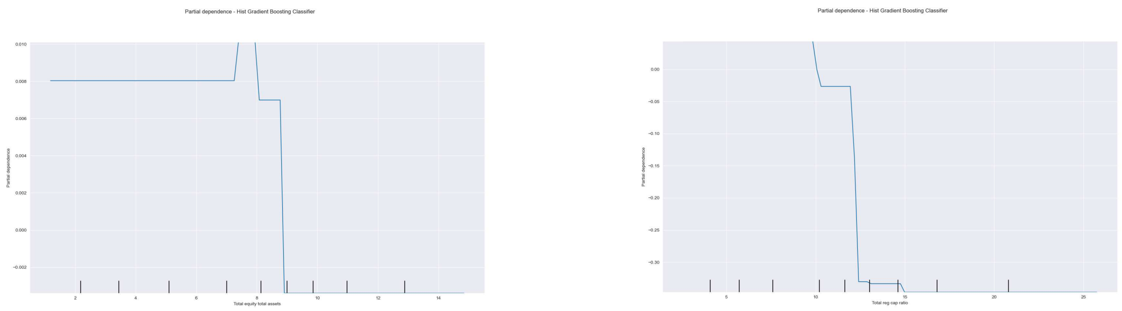
Source: Authors' calculations.

5.3 Optimal capital ratio

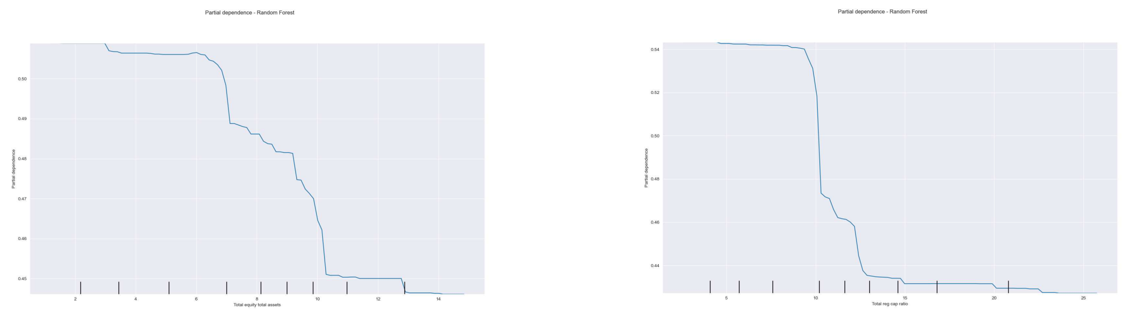
Determining optimal capital requirements requires to exhibit the potential non-linearities characterizing the relationship between capital variables (TE/TA and TRCR) and the probability of default. The idea is that increasing capital requirements is expected to reduce the probability of default up to a certain threshold above which further increasing these requirements could be counter-productive. Exhibiting such non-linearites requires to resort to non-linear models. According to Table 2, the non-linear models that perform the best are HGBC, RF and GBC. We therefore focus on these three models. PDPs are reported in Figure 6. In the three cases, we notice two regimes : one characterized by a high probability of default and a low capitalization, the other characterized by a low probability of default and a large capitalization. More specifically, banks enter this second regime when at least 10% of their assets are funded through core equity (left plots), which corresponds to a regulatory capital ratio greater than approximately 15% (right plots). This is consistent with the idea that current capital requirements are set at a too low level. To dig further into the question of optimal capital requirements, we need to inquire into the impact of the interaction between the two capital ratios here at work. Indeed, as far as preventing bank default is concerned, it may be that keeping only one of the two ratios and setting it at its optimal level is better than keeping the two.

Figure 6: Partial Dependence Plots (PDPs) – TE/TA (left) and TRCR (right)

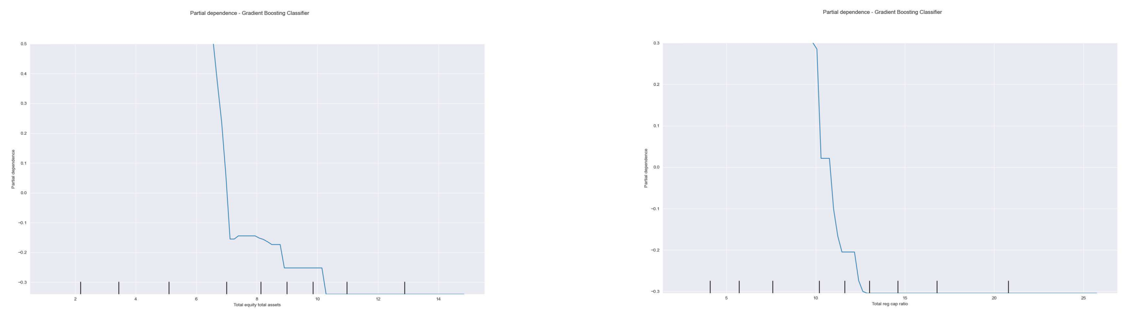
(a) HGBC



(b) RF



(c) GBC



Source: Authors' calculations.

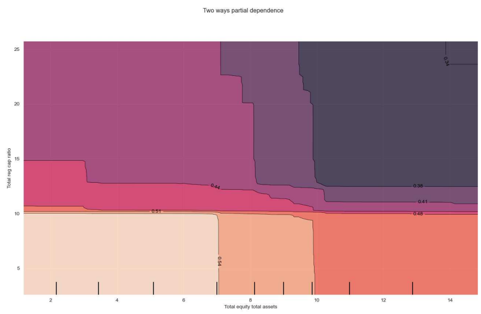
Two-way PDPs (Figure 7) allow to grasp the impact of the interaction between TRCR and TE/TA on the probability of default. Having a look at Figure 7, we notice four situations defined by the threshold values of TE/TA and TRCR as identified above:

- Situation A (lower-left corner): TE/TA < 10% and TRCR < 15%, the probability of default is the highest.
- Situation B (lower-right corner): TE/TA > 10% and TRCR < 15%, the probability of default is lower than in situation A.

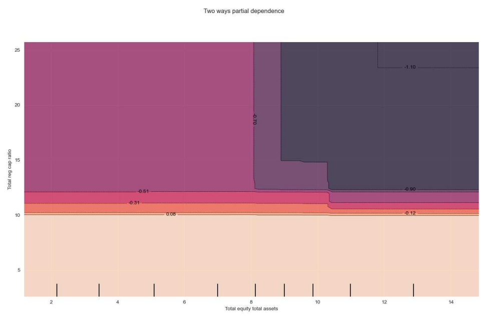
- Situation C (upper-left corner): TE/TA < 10% and TRCR > 15%, the probability of default is lower than in situation A.
- Situation D (upper-right corner): TE/TA > 10% and TRCR > 15%, the probability of default is the lowest.

Figure 7: Two-way Partial Dependence Plots (PDPs) between TE/TA and TRCR

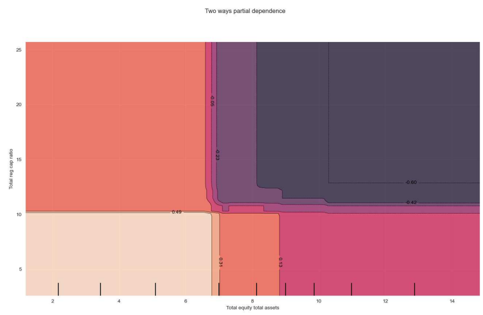
(a) RF



(b) HGBC



(c) GBC



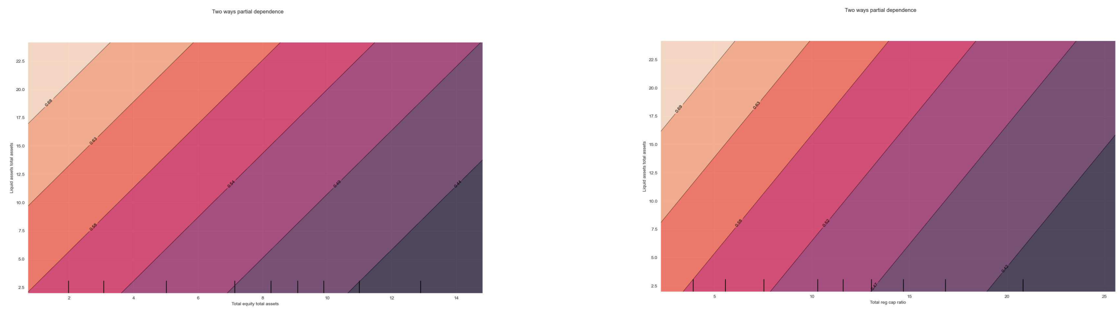
Source: Authors' calculations.

HGBC, RF and GBC allow to conclude that situation A is where the probability of default is maximal, while situation D is where it is minimal. Whether situation B outperforms situation C or not cannot be decided from Figure 7. Finally, we can conclude that implementing two capital ratios at the same time is preferable than implementing only either a risk-weighted ratio or a simple leverage ratio. The rationale behind this result may be that the leverage ratio provides a floor preventing manipulations of the risk-weighted ratio from releasing too much the capital constraint, while taking assets' risk into account prevents banks from engaging in overly risky investment strategies. The two capital ratios may thus be complementary to each other. In addition, the empirical evidence provided in this section indicates that the leverage ratio should be set at least at 10%, while the risk-weighted ratio should be set at 15%. This is above the threshold values defined by Basel III and consistent with those put forward by Karmakar (2016); Egan et al. (2017); Dagher et al. (2020); Mendicino et al. (2021). Setting capital requirements at these levels would not hamper banks' activities. Indeed, evidence is provided by Durand and Le Quang (2022) that increasing capital requirements to these levels would actually have a positive impact on banks' ROA, the only negative impact being on banks' ROE. In other words, shareholders would bear the entire cost of such strengthening of capital requirements. However, such a cost is very likely to be compensated by the reduction of the default risk resulting from better capitalization. Eventually, strengthening capital requirements could thus yield a win-win situation.

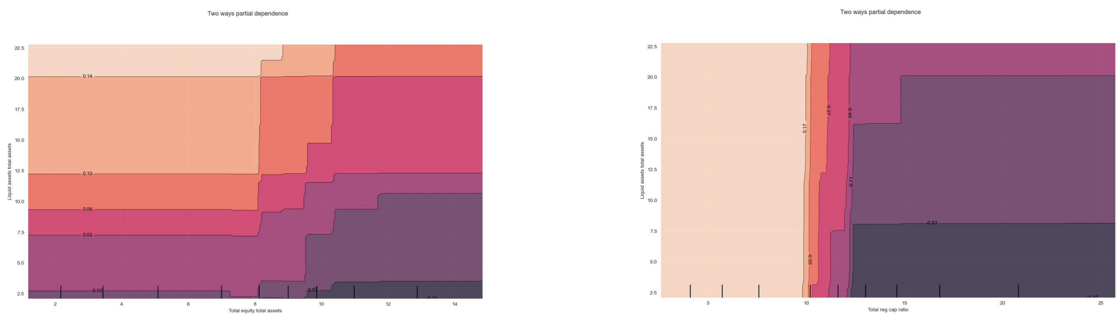
5.4 Interaction between liquidity and solvency risks and regulatory complexity

Figure 8: Two-way Partial Dependence Plots (PDPs) – TE/TA and LA/TA (left) and TRCR and LA/TA (right)

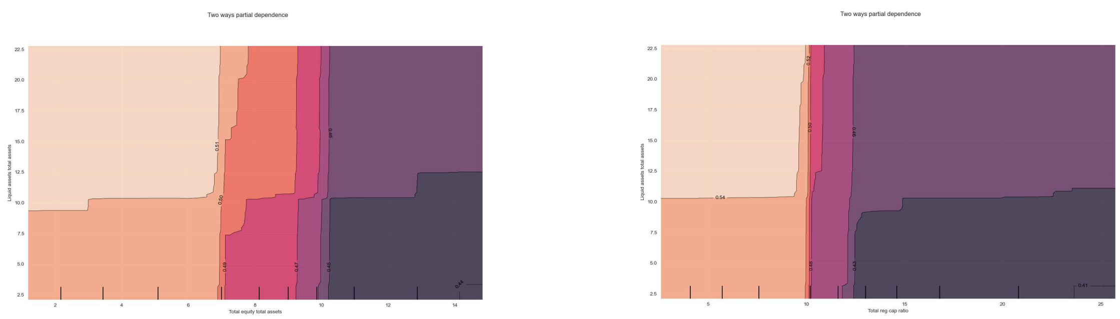
(a) Logit



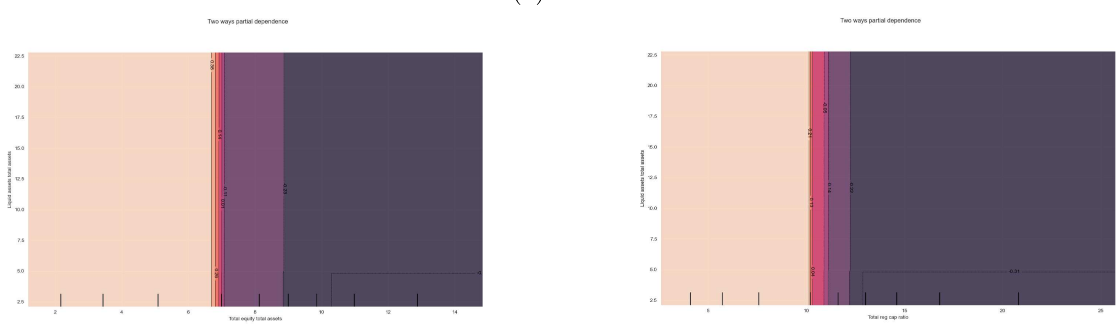
(b) HGBC



(c) RF



(d) GBC



Source: Authors' calculations.

Basel III has introduced two liquidity ratios. The purpose of these ratios is to prevent the liquidity spirals observed during the 2007-08 crisis from materializing. If such spirals were indeed witnessed at that time, it is not so clear that they were actually completely unrelated to banks being generally under-capitalized. Indeed, as Pérignon et al. (2018) show it for Europe, no actual market-wide liquidity dry-up occurred between 2008 and 2014 and banks suffering from liquidity dry-ups during this period were often the least-capitalized ones. In this section, we study the interaction between capital and liquidity. In particular, we try to determine which of the two drives the most the probability of default. We have actually already offered a hint in section 5.2. Indeed, Table 5 shows that capital (TRCR or TE/TA) is a greater predictor of bank default than liquidity. Let us provide additional evidence to this idea. Figure 8 displays two-way PDPs for four models: Logit, HGBC, RF and GBC. These are the models that perform the best according to Table 2. Having a look at the impact of the interaction between TE/TA and LA/TA (left plots) and between TRCR and LA/TA (right plots) on the probability of default we notice that it seems to be mostly driven by the variable accounting for capital (TE/TA and TRCR). This is particularly clear when we focus on the interaction between TRCR and LA/TA in the case of RF, GBC and HGBC. In these three cases, the probability of default is indeed completely determined by the value of TRCR. Things are less clear cut when the interaction between TE/TA and LA/TA is considered. While GBC and RF strongly support the idea that the impact of LA/TA on the probability of default is limited when compared to that of TE/TA, results drawn from Logit and HGBC are more ambiguous. However, in all of the cases, we notice that it is only when TE/TA $> 10\%$ that the lowest probability of default can be reached. Similarly, when TRCR $< 15\%$ (right plots in Figure 8), the probability of default is large no matter the value of LA/TA, while when TRCR $> 15\%$, the probability of default is low no matter the value of LA/TA. As a consequence, it seems that liquidity is not *per se* a strong driver of bank default. We therefore recommend the liquidity ratios to be re-assessed and call into question their very *raison d'être*.

6 Robustness

6.1 Fine-tuning the decision threshold?

To classify banks, models predict a probability for a given vector of X: if this probability is superior (respectively inferior) to 0.5, the predicted class is 1 (respectively 0). Therefore, to increase the performance of a specific model, we could have fine tuned the decision threshold so as to choose the value that maximizes a given performance measure. Having a look at the distributions of the predicted probabilities for all the models (see Figure 9 in appendix C.1) we notice that all the distributions exhibit two clear-cut mods: one around 0 and the other around 1. In other words, changing the value of the decision threshold would only marginally modify the performance of the models (it would actually mostly re-balance the proportion of false positives and that of false negatives). More precisely, having a look at the distribution of each model and differentiating between true (in blue) and false (in red) predictions, we notice that the probabilities having a value between 0.4 and 0.6 are equally distributed among true and false (see Figure 10 in appendix C.1 for RF) . This result outlines the fact that changing the decision threshold would not change much the global quality of the models.

6.2 Accumulated local effects (ALEs)

One of the most important issues associated with PDPs is that they assume that the predictor for which the partial dependence is computed and the other ones are independent. As a consequence, the existence of strong linear correlations between some features (see section 4.2.3) may bias the computation of PDPs. In addition, making X_j vary across all its distribution creates the risk to overfit regions with almost no data. In order to take these issues into account, we compute Accumulated Local Effects (ALEs) (Datta et al., 2016) as a robustness check. By difference from PDPs, ALEs are unbiased even when features are correlated and they are computed over actual data intervals of the explanatory variables. ALEs are only reported for HGBC and RF. They are displayed in appendix C.2. Let us have a look at the different subplots presented in Figure 11.

Subplots (a): TRCR indeed has a negative impact on the probability of default. In addition, the two regimes identified in section 5.3 are clearly distinguishable: the probability of default is large when TRCR $< 15\%$ and low above this threshold. The idea that setting the risk-weighted ratio at 15% is optimal from a prudential perspective is thus confirmed.

Subplots (b): TE/TA indeed has a negative impact on the probability of default. In addition, the two regimes identified in section 5.3 are clearly distinguishable: the probability of default is large when TE/TA $< 10\%$ and low above this threshold. The idea that setting the leverage ratio at 10% is optimal from a prudential perspective is thus confirmed. Results are less clear for HGBC. Recall however that TE/TA is not identified as a strong predictor of default when HGBC is considered (see Table 5).

Subplots (c): LA/TA indeed has a positive impact on the probability of default. Once again, this positive impact can be rendered intelligible when keeping in mind that failing banks are very often constrained to fire sell some of their illiquid assets *prior* to default.

6.3 Without over-sampling

Using the SMOTE procedure to re-balance the dataset allows to increase the performance of the models. However, since it consists in creating new instances in the minority class, it significantly modifies the information the classifiers find in the dataset. To check the robustness of the results presented in this paper, we therefore re-run the models without over-sampling the train sample. Results are presented for all the models in Table 8 in appendix C.3. We do notice that the models underperform when the train sample is imbalanced. More specifically, they get lower scores on the train sample and over-fit the zero class and, as a consequence, fail to properly identify failed banks. Macro recalls are however not so low on both the train and test samples. This can be explained by the propensity of the models to over-predict zeros. KNN is a great illustration: all banks but one are classified as 0 (unfailed banks). In this case, the precisions for both zeros and ones reach 100%, while the recall for ones is only 1%. This is by the way a further argument in favor of using macro recall as the reference score to assess the performance of the models.

6.4 Standardized predictors

In order to control for the impact of (un)standardized features on our models' quality, we ran all hyperparameters pipelines allowing for standardization.[6] Our goal is to see if an increase appears in the models' macro recall. We find that this is the case for KNN (+ 8.48 % with RobustScaler), Linear SVC (+ 8.48 % with RobustScaler), MLP (+ 10.91 % with StandardScaler) and GBC (+ 2.75 % with RobustScaler). However, no standardization is optimal for Logit, RF and HGBC while potential increases imply a loss in interpretability.

[6] We used "StandardScaler" and "RobustScaler" modules from Scikit-Learn as extra hyperparameters.

7 Conclusion

During the last decade, banking regulation has evolved in order to account for the main destabilizing dynamics observed during the 2007-08 crisis. New rules have therefore been implemented (liquidity ratios, bail-in standards, simple leverage ratio) and existing rules have been revised. What is however striking is the lack of consensus in the academic literature concerning both the design of these rules and the potential unintended consequences of their joint-implementation. Moreover, the banking crisis experienced both in the US and in Europe in 2023 suggests that the rules implemented after the 2007-08 crisis may not be up to the task. This paper provides evidence that allows to better assess the current state of banking regulation. Implementing various bankruptcy prediction models on a database comprising 4707 US banks and 454 observations of default on the period 2000-2018, we indeed manage to offer new insights on banking regulation.

The first challenge tackled by the paper is a methodological one. Dealing with a severely imbalanced dataset, we first compare the performance of seven models at predicting bank default: Logit, Random Forests (RF), K-Nearest Neighbors (KNN), Gradient Boosting Classification (GBC), Histogram-based Gradient Boosting Classification (HGBC), Linear Support Vector Classification (Linear SVC) and Multi Layer Perceptron (MLP). Balancing the dataset thanks to the SMOTE procedure, we show that Logit, RF, GBC and HGBC perform the best.

Focusing on the four best models (Logit, RF, HGBC and GBC) we then provide answers to two key questions for banking regulation: that of optimal capital requirements and that of the impact of the interaction between liquidity and capital on the probability of default. We show that capital requirements as defined by Basel III are below their optimal level. Indeed, we provide evidence that setting the leverage ratio at 10% and the risk-weighted ratio at 15% would significantly decrease the default risk. In addition, according to the evidence provided by Durand and Le Quang (2022), strengthening capital requirements up to these levels would not hamper banks' activities. Concerning the liquidity ratios, we provide evidence that liquidity cannot be considered as a strong predictor of bank default. As a consequence, we recommend releasing the regulatory pressure put on liquidity to focus it on capital regulation. Doing so would also allow to reduce the complexity of banking regulation.

References

V. V. Acharya and N. Mora. A Crisis of Banks as Liquidity Providers: A Crisis of Banks as Liquidity Providers. *The Journal of Finance*, 70(1):1–43, Feb. 2015. ISSN 00221082. doi: 10.1111/jofi.12182. URL `https://onlinelibrary.wiley.com/doi/10.1111/jofi.12182`. 4

A. R. Admati, P. M. DeMarzo, M. F. Hellwig, and P. C. Pfleiderer. Fallacies, irrelevant facts, and myths in the discussion of capital regulation: why bank equity is not socially expensive. *Max Planck Institute for Research on Collective Good*, 2013. 2, 6

A. Alter and Y. S. Schüler. Credit spread interdependencies of European states and banks during the financial crisis. *Journal of Banking & Finance*, 36(12):3444–3468, Dec. 2012. ISSN 03784266. doi: 10.1016/j.jbankfin.2012.08.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378426612002221`. 3

E. I. Altman. Financial ratios, dirscriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609, Sept. 1968. ISSN 00221082. 3

E. I. Altman, R. G. Haldeman, and P. Narayanan. ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1):29–54, June 1977. ISSN 03784266. 3

J. R. Barth and S. M. Miller. Benefits and costs of a higher bank leverage ratio. *Journal of Financial Stability*, 38:37–52, Oct. 2018. ISSN 15723089. doi: 10.1016/j.jfs.2018.07.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S1572308917306526`. 7

A. N. Berger and C. H. Bouwman. How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109(1):146–176, July 2013. ISSN 0304405X. doi: 10.1016/j.jfineco.2013.02.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S0304405X13000512`. 3

A. N. Berger and R. DeYoung. Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, 21(6):849–870, June 1997. ISSN 03784266. doi: 10.1016/S0378-4266(97)00003-4. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378426697000034`. 3

A. Blundell-Wignall and C. Roulet. Business models of banks, leverage and the distance-to-default. *OECD Journal: Financial Market Trends*, 2012(2):7–34, Feb. 2013. ISSN 19952872. doi: 10.1787/fmt-2012-5k4bxlxbd646. URL `https://www.oecd-ilibrary.org/finance-and-investment/business-models-of-banks-leverage-and-the-distance-to-default_fmt-2012-5k4bxlxbd646`. 5

P. Bologna, A. Miglietta, and A. Segura. Contagion in the CoCos Market? A Case Study of Two Stress Events. *International Journal of Central Banking*, 16(6):137–184, December 2020. URL `https://ideas.repec.org/a/ijc/ijcjou/y2020q5a4.html`. 6

B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, pages 144–152, Pittsburgh, Pennsylvania, United States, 1992. ACM Press. ISBN 978-0-89791-497-0. doi: 10.1145/130385.130401. URL `http://portal.acm.org/citation.cfm?doid=130385.130401`. 9

L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 08856125. doi: 10.1023/A:1010933404324. URL `http://link.springer.com/10.1023/A:1010933404324`. 8

M. K. Brunnermeier and L. H. Pedersen. Market Liquidity and Funding Liquidity. *The Review of Financial Studies*, 22(6):2201–2238, 2009. ISSN 08939454, 14657368. 1

R. J. Caballero, E. Farhi, and P.-O. Gourinchas. The Safe Assets Shortage Conundrum. *Journal of Economic Perspectives*, 31(3):29–46, Aug. 2017. ISSN 0895-3309. doi: 10.1257/jep.31.3.29. URL `https://pubs.aeaweb.org/doi/10.1257/jep.31.3.29`. 7

E. Campiglio. Beyond carbon pricing: The role of banking and monetary policy in financing the transition to a low-carbon economy. *Ecological Economics*, 121:220–230, Jan. 2016. ISSN 09218009. doi: 10.1016/j.ecolecon.2015.03.020. URL `https://linkinghub.elsevier.com/retrieve/pii/S0921800915001056`. 7

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002. 3, 8

T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016. doi: 10.1145/2939672.2939785. URL `https://doi.org/10.1145%2F2939672.2939785`. 9

L. Clerc, C. Lubochinsky, C. Pouvelle, and A. Tarazi. Interactions of bank capital and liquidity standards within the Basel 3 framework: A literature review. *Note du Comité scientifique de l'ACPR*, May 2022. URL `https://acpr.banque-france.fr/sites/default/files/medias/documents/20220401_def_38_revue_litterature_bale_3.pdf`. 1, 4

J. A. Cookson, C. Fox, J. Gil-Bazo, J. F. Imbet, and C. Schiller. Social Media as a Bank Run Catalyst. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4422754. URL `https://www.ssrn.com/abstract=4422754`. 2

J. Dagher, G. Dell'Ariccia, L. Laeven, M. L. Ratnovski, and M. H. Tong. Bank Capital: A Seawall Approach. *International Journal of Central Banking*, 16(2):249–291, Mar. 2020. URL `https://ideas.repec.org/a/ijc/ijcjou/y2020q1a7.html`. 1, 2, 7, 20

A. Datta, S. Sen, and Y. Zick. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. pages 598–617, 05 2016. 23

M. D. Delis and P. K. Staikouras. Supervisory Effectiveness and Bank Risk. *Review of Finance*, 15(3): 511–543, July 2011. ISSN 1573-692X, 1572-3097. doi: 10.1093/rof/rfq035. URL `https://academic.oup.com/rof/article-lookup/doi/10.1093/rof/rfq035`. 3

A. Demirgüç-Kunt and H. Huizinga. Bank activity and funding strategies: The impact on risk and returns. *Journal of Financial Economics*, 98(3):626–650, Dec. 2010. ISSN 0304405X. doi: 10.1016/j.jfineco.2010.06.004. URL `https://linkinghub.elsevier.com/retrieve/pii/S0304405X10001534`. 3

A. Demirguc-Kunt, E. Detragiache, and O. Merrouche. Bank Capital: Lessons from the Financial Crisis. *Journal of Money, Credit and Banking*, 45(6):1147–1164, Sept. 2013. ISSN 00222879. doi: 10.1111/jmcb.12047. URL `https://onlinelibrary.wiley.com/doi/10.1111/jmcb.12047`. 6

R. DeYoung, I. Distinguin, and A. Tarazi. The joint regulation of bank liquidity and bank capital. *Journal of Financial Intermediation*, 34:32–46, Apr. 2018. ISSN 10429573. doi: 10.1016/j.jfi.2018.01.006. URL `https://linkinghub.elsevier.com/retrieve/pii/S1042957318300123`. 4, 7

D. W. Diamond and P. H. Dybvig. Bank runs, deposit Insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419, June 1983. ISSN 0022-3808, 1537-534X. doi: 10.1086/261155. URL `https://www.journals.uchicago.edu/doi/10.1086/261155`. 4

P. Durand and G. Le Quang. Banks to basics! Why banking regulation should focus on equity. *European Journal of Operational Research*, 301(1):349–372, Aug. 2022. ISSN 03772217. doi: 10.1016/j.ejor.2021.10.057. URL `https://linkinghub.elsevier.com/retrieve/pii/S0377221721009061`. 2, 6, 20, 24

M. Egan, A. Hortaçsu, and G. Matvos. Deposit Competition and Financial Fragility: Evidence from the US Banking Sector. *American Economic Review*, 107(1):169–216, Jan. 2017. ISSN 0002-8282. doi: 10.1257/aer.20150342. URL `https://pubs.aeaweb.org/doi/10.1257/aer.20150342`. 7, 20

S. Eichler and K. Sobanski. National politics and bank default risk in the eurozone. *Journal of Financial Stability*, 26:247–256, Oct. 2016. ISSN 15723089. doi: 10.1016/j.jfs.2016.07.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S1572308916300614`. 3, 4

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006. ISSN 01678655. doi: 10.1016/j.patrec.2005.10.010. URL `https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X`. 10

J. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 11 2000. 10

J. H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189–1232, 2001. ISSN 00905364. URL `http://www.jstor.org/stable/2699986`. Publisher: Institute of Mathematical Statistics. 9

L. Gambacorta and H. S. Shin. Why bank capital matters for monetary policy? *Journal of Financial Intermediation*, 35(B):17–29, July 2018. 6

V. García, J. S. Sánchez, R. Martín-Félez, and R. A. Mollineda. Surrounding neighborhood-based SMOTE for learning from imbalanced data sets. *Progress in Artificial Intelligence*, 1(4):347–362, Dec. 2012. ISSN 2192-6352, 2192-6360. 3

R. Goncharenko. Fighting Fire with Gasoline: CoCos in Lieu of Equity. *Journal of Money, Credit and Banking*, page jmcb.12857, Aug. 2021. ISSN 0022-2879, 1538-4616. doi: 10.1111/jmcb.12857. URL `https://onlinelibrary.wiley.com/doi/10.1111/jmcb.12857`. 6

T. M. Ha and H. Bunke. Off-line, handwritten numeral recognition by perturbation method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):535–539, 1997. 8

G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239, May 2017. ISSN 09574174. 3

H. Hakenes and I. Schnabel. Regulatory Capture by Sophistication. VfS Annual Conference 2013 (Duesseldorf): Competition Policy and Regulation in a Global Economic Order 79991, Verein für Socialpolitik / German Economic Association, 2013. URL `https://ideas.repec.org/p/zbw/vfsc13/79991.html`. 5

A. Haldane and V. Madouros. The dog and the frisbee. *Proceedings - Economic Policy Symposium - Jackson Hole*, pages 109–159, 2012. URL `https://EconPapers.repec.org/RePEc:fip:fedkpr:y:2012:p:109-159`. 1, 5

A. G. Haldane. Capital discipline. Technical report, speech given at the American Economic Association, Jan. 2011. URL `https://www.bis.org/review/r110325a.pdf`. 5

D. J. Hand. Assessing the performance of classification methods. *International Statistical Review*, 80, 12 2012. 9

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009. ISBN 9780387848846. 8, 9, 10

R. J. Herring. The Evolving Complexity of Capital Regulation. *Journal of Financial Services Research*, 53(2-3):183–205, June 2018. ISSN 0920-8550, 1573-0735. doi: 10.1007/s10693-018-0295-8. URL `http://link.springer.com/10.1007/s10693-018-0295-8`. 1, 5, 6

B. Imbierowicz and C. Rauch. The relationship between liquidity risk and credit risk in banks. *Journal of Banking & Finance*, 40:242–256, Mar. 2014. ISSN 03784266. doi: 10.1016/j.jbankfin.2013.11.030. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378426613004512`. 3, 4

S. Jones and D. A. Hensher. Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review*, 79(4):1011–1038, 2004. ISSN 00014826. 3

E. J. Kane. Good Intentions and Unintended Evil: The Case against Selective Credit Allocation. *Journal of Money, Credit and Banking*, 9(1):55–69, 1977. URL https://EconPapers.repec.org/RePEc:mcb:jmoncb:v:9:y:1977:i:1:p:55-69. 5

E. J. Kane. Accelerating Inflation, Technological Innovation, and the Decreasing Effectiveness of Banking Regulation. *The Journal of Finance*, 36(2):355–367, May 1981. ISSN 00221082. doi: 10.1111/j.1540-6261.1981.tb00449.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1981.tb00449.x. 5

S. Karmakar. Macroprudential regulation and macroeconomic activity. *Journal of Financial Stability*, 25:166–178, Aug. 2016. ISSN 15723089. doi: 10.1016/j.jfs.2016.06.006. URL https://linkinghub.elsevier.com/retrieve/pii/S1572308916300481. 7, 20

A. Kashyap, J. C. Stein, and S. G. Hanson. An Analysis of the Impact of 'Substantially Heightened' Capital Requirements on Large Financial Institutions. 2010. 6

G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf. 9

J. Kolari, D. Glennon, H. Shin, and M. Caputo. Predicting large US commercial bank failures. *Journal of Economics and Business*, 54(4):361–387, 2002. 3

L. Laeven and R. Levine. Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93(2):259–275, Aug. 2009. ISSN 0304405X. doi: 10.1016/j.jfineco.2008.09.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X09000816. 3

G. Le Quang. Mind the conversion risk: contingent convertible bonds as a transmission channel of systemic risk. *Finance*, forthcoming, 2022. 6

M. Mariathasan and O. Merrouche. The manipulation of basel risk-weights. *Journal of Financial Intermediation*, 23(3):300–321, July 2014. ISSN 10429573. doi: 10.1016/j.jfi.2014.04.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1042957314000254. 5

D. Martin. Early warning of bank failure. *Journal of Banking & Finance*, 1(3):249–276, Nov. 1977. ISSN 03784266. 3

D. G. Mayes and H. Stremmel. *The effectiveness of capital adequacy measures in predicting bank distress*. Number 2014,1 in SUERF studies. SUERF - The Europ. Money and Finance Forum [u.a.], Vienna, 2014. ISBN 978-3-902109-72-9. OCLC: 891506164. 5

W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943. 9

C. Mendicino, K. Nikolov, J.-R. Ramirez, J. Suarez, and D. Supera. How much capital should banks hold? *Research Bulletin ECB*, 80, 2021. URL https://ideas.repec.org/a/ecb/ecbrbu/20210080.html. 7, 20

D. Miles, J. Yang, and G. Marcheggiano. Optimal Bank Capital. *The Economic Journal*, 123(567): 1–37, Mar. 2013. ISSN 0013-0133, 1468-0297. doi: 10.1111/j.1468-0297.2012.02521.x. URL https://academic.oup.com/ej/article/123/567/1-37/5080596. 6

M. H. Miller. Do the M & M propositions apply to banks? *Journal of Banking & Finance*, 19(3-4):483–489, June 1995. ISSN 03784266. doi: 10.1016/0378-4266(94)00134-O. URL https://linkinghub.elsevier.com/retrieve/pii/037842669400134O. 6

F. Modigliani and M. H. Miller. The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, 48(3):261–297, 1958. ISSN 00028282. URL http://www.jstor.org/stable/1809766. Publisher: American Economic Association. 6

C. Molnar. *Interpretable machine learning: a guide for making Black Box Models interpretable.* 2020. ISBN 978-0-244-76852-2. URL https://christophm.github.io/interpretable-ml-book/. OCLC: 1105728953. 10

S. Morris and H. S. Shin. Illiquidity component of credit risk – the 2015 Lawrence R. Klein Lecture. *International Economic Review*, 57(4):1135–1148, Nov. 2016. ISSN 00206598. doi: 10.1111/iere.12192. URL http://doi.wiley.com/10.1111/iere.12192. 4

J. Muschelli. ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. *Journal of Classification*, 37(3):696–708, Oct. 2020. ISSN 0176-4268, 1432-1343. doi: 10.1007/s00357-019-09345-1. URL http://link.springer.com/10.1007/s00357-019-09345-1. 10

J. A. Ohlson. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1):109, 1980. ISSN 00218456. 3

M. Ojala and G. C. Garriga. Permutation tests for studying classifier performance. *J. Mach. Learn. Res.*, 11:1833–1863, aug 2010. 10

P. Parrado-Martínez, P. Gómez-Fernández-Aguado, and A. Partal-Ureña. Factors influencing the European bank's probability of default: An application of SYMBOL methodology. *Journal of International Financial Markets, Institutions and Money*, 61:223–240, July 2019. ISSN 10424431. doi: 10.1016/j.intfin.2019.04.003. URL https://linkinghub.elsevier.com/retrieve/pii/S1042443118304700. 3

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011. 8

E. Perotti. The Swiss authorities enforced a legitimate going concern conversion, Mar. 2023. URL https://cepr.org/voxeu/columns/swiss-authorities-enforced-legitimate-going-concern-conversion. 2

A. D. Persaud. Why bail-in securities are fool's gold. *Peterson Institute for International Economic Policy Brief 14-23*, 2014. doi: https://www.piie.com/sites/default/files/publications/pb/pb14-23.pdf. 2

C. Pérignon, D. Thesmar, and G. Vuillemey. Wholesale Funding Dry-Ups. *The Journal of Finance*, 73(2): 575–617, 2018. doi: https://doi.org/10.1111/jofi.12592. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12592. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12592. 4, 7, 22

P. Ravi Kumar and V. Ravi. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180(1):1–28, July 2007. ISSN 03772217. 3, 4

V. Salas and J. Saurina. Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks. *Journal of Financial Services Research*, 22(3):203–224, 2002. ISSN 09208550. doi: 10.1023/A:1019781109676. URL `http://link.springer.com/10.1023/A:1019781109676`. 3

L. M. Salchenberger, E. M. Cinar, and N. A. Lash. Neural Networks: A New Tool for Predicting Thrift Failures. *Decision Sciences*, 23(4):899–916, July 1992. ISSN 0011-7315, 1540-5915. 3

L. S. Shapley. A Value for n-Person Games. *Princeton University Press*, pages 307–317, 1953. 10, 15

B. Soederhuizen, B. Kramer, H. v. Heuvelen, and R. Luginbuhl. Optimal capital ratios for banks in the euro area. CPB Discussion Paper 429, CPB Netherlands Bureau for Economic Policy Analysis, Sept. 2021. URL `https://ideas.repec.org/p/cpb/discus/429.html`. 7

N. Soenen and R. Vander Vennet. Determinants of European banks' default risk. *Finance Research Letters*, page 102557, Nov. 2021. ISSN 15446123. doi: 10.1016/j.frl.2021.102557. URL `https://linkinghub.elsevier.com/retrieve/pii/S154461232100516X`. 3

N. Soenen and R. Vander Vennet. ECB monetary policy and bank default risk. *Journal of International Money and Finance*, 122:102571, Apr. 2022. ISSN 02615606. doi: 10.1016/j.jimonfin.2021.102571. URL `https://linkinghub.elsevier.com/retrieve/pii/S0261560621002229`. 3, 4

M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, July 2009. ISSN 03064573. doi: 10.1016/j.ipm.2009.03.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0306457309000259`. 10

E. Strumbelj and I. Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41:647–665, 2013. doi: 10.1007/s10115-013-0679-x. 10, 15

K. Tam. Neural network models and the prediction of bank bankruptcy. *Omega*, 19(5):429–445, Jan. 1991. ISSN 03050483. 3

K. Y. Tam and M. Kiang. Predicting bank failures: A neural network approach. *Applied Artificial Intelligence*, 4(4):265–282, Jan. 1990. ISSN 0883-9514, 1087-6545. 3

A. V. Thakor. Post-crisis regulatory reform in banking: Address insolvency risk, not illiquidity! *Journal of Financial Stability*, 37:107–111, 2018. ISSN 1572-3089. doi: https://doi.org/10.1016/j.jfs.2018.03.009. URL `https://www.sciencedirect.com/science/article/pii/S1572308918301396`. 1, 4, 7

F. Vazquez and P. Federico. Bank funding structures and risk: Evidence from the global financial crisis. *Journal of Banking & Finance*, 61:1–14, Dec. 2015. ISSN 03784266. doi: 10.1016/j.jbankfin.2015.08.023. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378426615002344`. 5

L. Zhou. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, 41:16–25, Mar. 2013. ISSN 09507051. 3

M. Zoričák, P. Gnip, P. Drotár, and V. Gazda. Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets. *Economic Modelling*, 84:165–176, Jan. 2020. ISSN 02649993. 3

A Data sources and definitions

Table 6: Data sources and definitions

Data	Definition	Source
Total equity total assets	Ratio of total equity to total assets. This ratio is close to the leverage ratio as defined under Basel agreements.	FitchConnect
Total reg cap ratio	Total regulatory capital ratio as defined under Basel agreements. It is fixed to 8% of the risk weighted assets, plus a conservation buffer (2%).	FitchConnect
Liquid assets total assets	Liquid assets detained by the bank over its total assets	FitchConnect
Net loans total assets	Ratio of net loans to total assets.	FitchConnect
Deposits mm funding growth	Growth rate of deposits to money market funding.	FitchConnect
Expenses rev	Expenses over revenues ratio.	FitchConnect
Net int margin	Returns on invested funds. It is measured by the difference between the interests received and those paid, divided by the average invested assets.	FitchConnect
Non earning assets gs over TA	All assets that do not generate income over total assets.	FitchConnect
Operating roaa	Ratio of net income to average total assets. It measures the profitability of assets, meaning how a firm uses the resources it owns to generate profit. It refers to the returns on the assets purchased using each unit of money invested.	FitchConnect
Tax expense over TA	Expense for current and deferred tax for the period over total assets.	FitchConnect
Total desposits over TA	Total deposits over total assets.	FitchConnect
Total funding over TA	Total Deposits, Money Market and Short-term Funding + Total Long Term Funding + Derivatives + Trading Liabilities, all over total assets.	FitchConnect
Total int exp over TA	Ratio of total interest expense / Total assets.	FitchConnect
Total non earning assets over TA	All assets that do not generate income, over total assets.	FitchConnect
Total oper expense over TA	Operating costs include administration costs such as staff costs, over total assets	FitchConnect
log Avg earning assets over TA	Logarithm of year assets that generate income, over total assets.	FitchConnect

Table 6: (continued)

log Total assets	Logarithm of total assets. It gives a proxy for banks' size.	FitchConnect
log Bus volume over TA	Logarithm Total Business Volume = Managed Securitized Assets Reported Off-Balance Sheet + Other off-balance sheet exposure to securitizations + Guarantees + Acceptances and documentary credits reported off-balance sheet + Committed Credit Lines + Other Contingent Liabilities + Total Assets. All over total assets.	FitchConnect
log Liquid assets deposits mm funding	Liquid assets as a deposit.	FitchConnect
log Total non int exp over TA	Non interest expenses over total assets.	FitchConnect

B Hyperparameters

The list of the hyperparameters and their values are presented in Table 7. For each model, the optimal values of the hyperparameters are displayed in bold characters. The last column presents the mean effect (if any) of each hyperparameter (from the left of the list) on the out-of-sample macro recall score.

Table 7: Hyperparameters per model and macro recall

Model	Hyperparameter	value range	mean effect
LR	C	[0.01, 0.1, ***1***, 10, 100, 1000]	inv. U-shape
	penalty	[None, ***l1***, *l2*, ElasticNet]	inv. U-shape
	l1_ratio (if EN)	[0.1, 0.3, 0.5, 0.7, 0.9]	n.s.
	solver	[lbfgs, ***liblinear***, newton-cg, sag, saga]	decreasing
RF	n_estimators	[5, 10, 50, ***100***, 500, 1000]	inv. U-shape
	max_depth	[***5***, 10, 50, None]	decreasing
	max_features	[***5***, 10, sqrt, log2, None]	U-shape
	min_samples_split	[***5***, 10, 20]	increasing
GBC	learning_rate	[0.01, ***0.1***, 0.5]	decreasing
	n_estimators	[5, ***10***, 50,100, 500, 1000]	decreasing
	max_depth	[***5***, 10, 50, None]	decreasing
	max_features	[***5***, 10, sqrt, log2, None]	U-shape
	min_samples_split	[***5***, 10, 20]	decreasing
HGBC	learning_rate	[0.01, ***0.1***, 0.5]	decreasing
	max_iter*	[5, ***10***, 50,100, 500, 1000]	decreasing
	max_depth	[5, ***10***, 50, None]	inv. U-shape
	max_features	[***5***, 10, sqrt, log2, None]	U-shape
	l2_regularization	[***0***, 0.1]	decreasing
SVC	C	[0.1, ***1***, 10, 100, 1000]	inv. U-shape
	kernel	[***linear***, poly, rbf]	U-shape
	degree (if poly)	[***2***, 3]	decreasing
	gamma	[***0.0001***, 0.001]	decreasing
MLP	activation	[logistic, relu, ***tanh***]	inv. U-shape
	alpha	[***0.005***,0.01]	increasing
	max_iter	[10, ***50***, 100]	decreasing
	early_stopping	[False, ***True***]	increasing
	hidden_layers	39 combinations of [10, 100, 500] a.n.*	n.r.
	alpha	[***0.05***, 0.01]	increasing
KNN	n_neighbors	[4, 15, 25, 30, ***40***, 50]	increasing
	metric	[euclidean, ***manhattan***, minkowski]	inv. U-shape
	weights	[uniform, ***distance***]	increasing

Source: Authors' calculations.

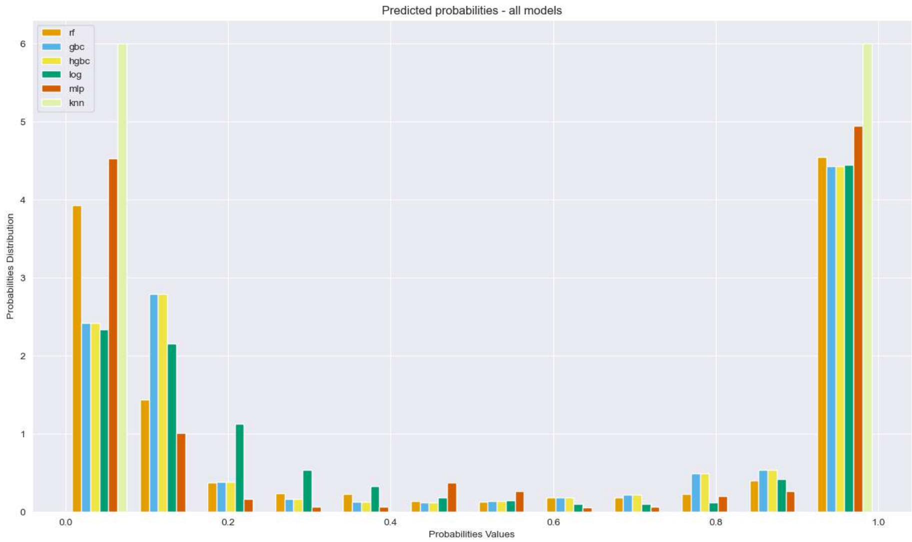
* max_iter in HGBC corresponds to n_estimators in RF and GB.

** The smallest possible MLP is chosen with a single layer of 10 artificial neurons.

C Robustness outputs

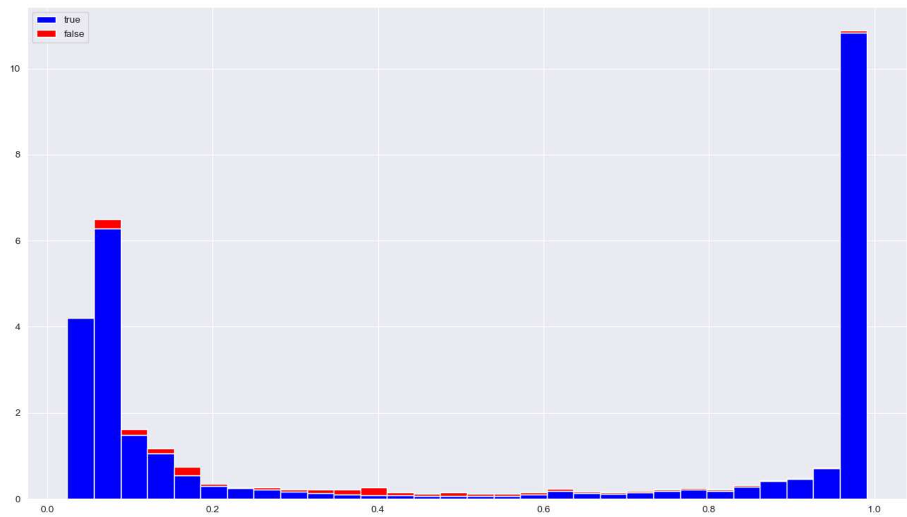
C.1 Fine-tuning the decision threshold?

Figure 9: Predicted probability distribution - all models



Source: Authors' calculations.

Figure 10: Predicted probability distribution - Random Forest

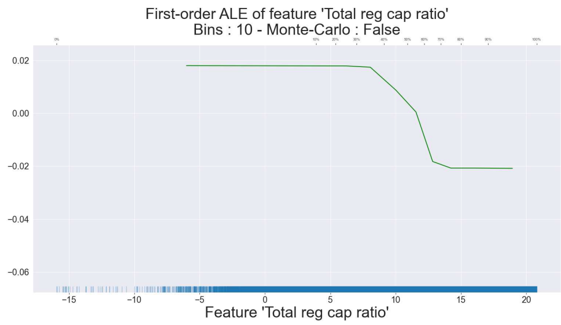


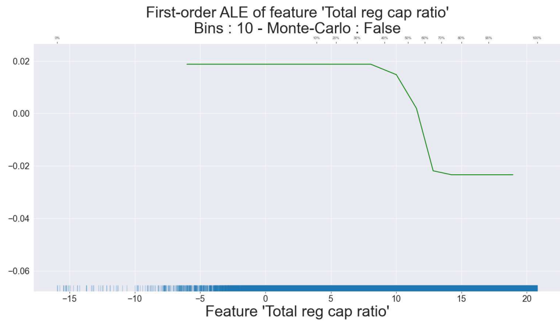
Source: Authors' calculations. Interpretation example: around 7% of the observations have a predicted probability around 0.05. Among those, more than 6% are well predicted (blue part) while 0.5% are wrongly predicted (red part).

C.2 ALEs

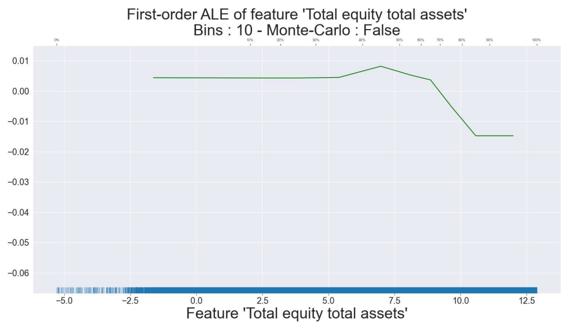
Figure 11: ALEs - RF (left) and HGBC (right)

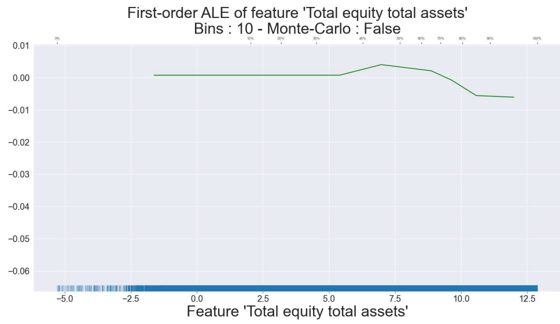
(a) TRCR



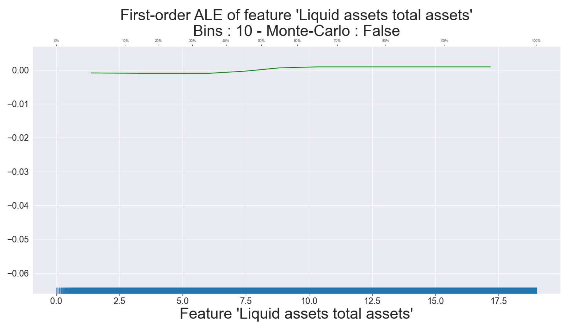


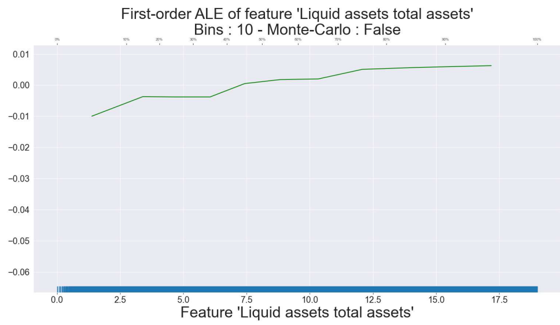
(b) TE/TA





(c) LA/TA





Source: Authors' calculations.

C.3 Without over-sampling

Table 8: Models' performance -

	Logit		RF		KNN		GBC	
	Train	Test	Train	Test	Train	Test	Train	Test
Recall	39	0	61	49	100	1	75	51
Macro recall (AUROC)	69	50	80	74	51	51	87	76

	HGBC		Linear SVC		MLP	
	Train	Test	Train	Test	Train	Test
Recall	55	13	31	44	44	46
Macro recall (AUROC)	77	57	65	71	72	73

Source: Authors' calculations.