



**HAL**  
open science

# Banks to basics! Why banking regulation should focus on equity

Pierre Durand, Gaëtan Le Quang

► **To cite this version:**

Pierre Durand, Gaëtan Le Quang. Banks to basics! Why banking regulation should focus on equity. European Journal of Operational Research, 2022, 301 (1), pp.349-372. 10.1016/j.ejor.2021.10.057 . hal-03411360

**HAL Id: hal-03411360**

**<https://hal.science/hal-03411360v1>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Banks to basics! Why banking regulation should focus on equity

Pierre Durand <sup>a</sup>, Gaëtan Le Quang <sup>b</sup>

---

---

## Abstract

Banking regulation faces multiple challenges that call for rethinking the way it is designed. In this paper, we argue that regulators should focus more on simple equity requirements instead of elaborating complex rules. Such a constraint in equity is however opposed by the banking industry because of its presumed adverse impact on banks' performance. Using various techniques (Lasso, Random Forest, Support Vector Regression, Artificial Neural Network) on a large dataset of banks' balance sheet variables, we show that the equity ratio (equity over total assets) has a clear positive effect on banks' performance when measured by the return on assets, while the impact of this ratio on the return on equity is most of the time negative. Strong equity requirements do not therefore impede banks' performance, but do reduce the shareholder value. This may be the reason why the banking industry so fiercely opposes strong equity requirements. In addition, from a methodological perspective, we provide evidence that Random Forest performs better than other techniques at dealing with banks' balance sheet data. Doing so, we provide avenues for future research dealing with these kind of data.

**Keywords:** Banking regulation; Random Forest; Support Vector Machine; Artificial Neural Network

**JEL classification :** C44, G21, G28

**Acknowledgements:** The authors are grateful to the editor, Emanuele Borgonovo, and to two anonymous referees for their helpful comments. The authors also want to thank their colleagues from the EconomiX research center. They are particularly grateful to Laurence Scialom and Valérie Mignon for their guidance and advice. Gaëtan Le Quang acknowledges the support of the Chair Energy and Prosperity, under the aegis of La Fondation du Risque.

<sup>a</sup>Université Paris Est Créteil, ERUDITE, 94010 Créteil Cedex, France

<sup>b</sup>Univ Lyon, Université Lumière Lyon 2, GATE UMR 5824, F-69130 Ecully, France

Corresponding author: [g.le-quang@univ-lyon2.fr](mailto:g.le-quang@univ-lyon2.fr)

## 1. Introduction

Banking regulation proved inadequate to prevent the banking system from collapsing at the end of the 2000s. Massive public bailouts thus occurred and banking regulators began to work on a new framework meant to make it possible to contain the risk inherent to banking activities. This risk is now known as being systemic and as striking both sides of the balance sheet. Macroprudential tools – such as regulatory countercyclical buffers – and liquidity requirements – such as the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR) – have therefore been put in place to tackle the threats posed by the specific nature of banking risk. In addition, bail-in instruments have been designed to prevent costly bailouts from occurring whenever a bank deemed too big to fail goes bankrupt.

Banking regulation has thus evolved after the 2007-2008 financial collapse. The macroprudential shift that began right after the crisis seems to lead banking regulation in the right direction. However, knowing that the previous financial crisis was, among other reasons, due to the very complexity of the financial system,<sup>1</sup> it is quite surprising to notice that the regulatory framework newly designed is very complex. The first pillar of Basel III indeed states that banks have to comply with two liquidity ratios, one leverage ratio and a risk-based capital ratio. Two bail-in standards – the Total Loss Absorbing Capacity (TLAC) that concerns systemic banks and the Minimum Requirement for own funds and Eligible Liabilities (MREL) that concerns every bank located in the European Union – have additionally been put in place to protect taxpayers from bailouts. This complexity paves the way for criticisms coming from the banking industry, arguing that such a multiplication of rules can only hamper banks' activity.<sup>2</sup> Conversely, it is also surprising to notice that simple rules, such as the simple leverage ratio, are set at a very low level, the latter being set at 3%.

The objective of this paper is to suggest that banking regulation would be better off focusing more on the simple leverage ratio (i.e., equity over total assets) instead of relying on complex and less transparent rules. We argue that implementing a strong constraint in equity could only reinforce the efficiency of banking regulation without impeding banks' performance. In particular, based on banks' balance sheet data, we provide evidence that, contrary to the argument put forward by the banking industry to oppose any strengthening of equity requirements, higher equity ratios could have a positive impact on banks' performance.

The main difficulty that arises when dealing with balance sheet data is the great number of variables that have to be taken into consideration and their potential non-linearity and mutual dependency. Dealing with banks' balance sheet data thus first requires to disentangle which variables are worth considering and which are not depending on the question asked. In this respect, statistical learning methods prove of great help, but have surprisingly not attracted much attention from the banking literature so far. To our knowledge, [Petropoulos et al. \(2020\)](#) is a rare exception where authors build an early warning system based on banks' balance sheet data by resorting to various Machine Learning

---

<sup>1</sup>This complexity was the direct consequence of the great promotion of financial innovation during the 2000s ([Brunnermeier, 2009](#)).

<sup>2</sup>In January 2020, the French Parliament voted against the full implementation of Basel III arguing mainly that it would impede banks' activities too much.

(ML) techniques. Our paper proceeds the same way to inquire the impact of different regulatory variables on banks' performance. Specifically, we resort to four different models: Lasso, Random Forest (RF), Support Vector Regression (SVR) and Artificial Neural Network (ANN). These models are applied on a large dataset of banks' balance sheet variables covering the period 2000-2018. Those methods being rarely used in the banking literature, we compare their performance to assess which proves the best to tackle balance sheet data. The models are compared in two different ways: when the question is dealt with through regressions and, as a robustness check, when the question is rephrased as a classification problem. This comparison allows us to focus on the results produced by the model that performs the best.

The contribution of our paper is twofold. We contribute to the banking regulation literature by providing evidence that the ratio equity over total assets ( $\frac{E}{TA}$ ) has a strong positive impact on banks' performance as measured by the return on average assets (ROAA). Far from impeding banks' profitability, high equity requirements could thus foster it. On the contrary, we show that the ratio  $\frac{E}{TA}$  has a non-linear impact on the return on average equity (ROAE). In particular, when the ratio  $\frac{E}{TA}$  is below approximately 8%, the ROAE is positively impacted by an increase in this ratio. When  $\frac{E}{TA}$  is above 8%, the ROAE is, on the contrary, negatively impacted by an increase in this ratio. This may be the main reason why the banking industry opposes higher constraints in equity: above a certain threshold, increasing equity requirements reduces the shareholder value. The main cost associated with an increase in equity requirements is, therefore, supported by shareholders (since the ROAE may decrease when  $\frac{E}{TA}$  increases) and not by other stakeholders (since the ROAA increases when  $\frac{E}{TA}$  increases). From a methodological perspective, our paper contributes to the growing literature on the performance of various empirical techniques at dealing with banks' balance sheet data. In line with previous studies (Tanaka et al., 2016; Petropoulos et al., 2020), we provide evidence that the use of ML techniques allows taking the specificity of these data into account and offers various ways of producing results that can easily be given an economic meaning. More specifically, RF regressions outperforms ANN and SVR in at least two respects: their in- and out-of-sample performances are better and their outputs can easily be interpreted to provide economic insights. The results drawn from regressions are comforted by those drawn from classification methods presented as a robustness check.

The remainder of the paper is structured as follows. Section 2 presents a review of the literature. Section 3 provides a brief overview of the regulatory framework that has been put in place after the crisis and presents the tested hypothesis. Section 4 presents the models. Data are presented in section 5 alongside with some descriptive statistics. Section 6 reports the results. Robustness checks are provided in section 7. Section 8 discusses the methodological contribution of the paper and section 9 concludes.

## 2. Literature review

Our paper lies at the intersection between two research questions that are traditionally treated separately: on the one hand, the question of the determinants of banks' solvability and, on the other hand, that of the determinants of banks' performance. Taking together, those two questions allow

inquiring the optimal regulatory framework, namely the one that would allow preventing bank failures without impeding banks' performance.

The literature on the determinants of banks' solvability tries to determine which variables help the most predicting failures. The rationale behind this approach is that, once those variables properly identified, regulators can elaborate constraints based on them. The literature on bankruptcy prediction 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 CAMELS<sup>3</sup> ratings. If those variables are widely acknowledged as the main determinants of bank failures, there is no consensus concerning the magnitude of their impact and their relative importance. Using a dataset covering 11,000 banks in the U.S. and Europe between 2001 and 2009, Vazquez and Federico (2015) argue that banking regulation would be better off focusing on the leverage ratio, especially for systemically important banks (G-SIBs). Indeed, according to these authors, a 3.5 percentage point increase in the pre-crisis capital buffers of G-SIBs would have caused a 48 percentage point decrease in their probability of failure during the crisis. More specifically, using logistic regressions and discrete survival time analysis, Mayes and Stremmel (2014) show that the simple leverage ratio outperforms the risk-weighted capital ratio at predicting bank failures. This result contributes to the literature that calls for cautiousness when it comes to designing capital requirements as a function of an estimation of the risk associated with bank activities (Mariathasan and Merrouche, 2014; Asmild and Zhu, 2016). On the contrary, focusing on 17 advanced economies and considering data from 1870 to 2015, a recent study by Jordà et al. (2021) shows that the capital ratio does not allow to predict *ex ante* bank failures but is a key determinant of banks' recovery *ex post*. From a methodological perspective, the literature on the determinants of bank failures has started from classic statistical techniques (such as logistic regressions) and now also resorts to more sophisticated approaches borrowing from operational research. Ravi Kumar and Ravi (2007) provide a review of this literature from 1968 to 2005. More recently, some papers have shown that ML techniques perform better than econometrics when it comes to determining which variables impact the most the probability of default. For example, Tanaka et al. (2016) design an early-warnings system for predicting bank failures. Doing so, they show that RF regressions outperform other methods such as logistic regressions. In the same vein, Petropoulos et al. (2020) show that random forests and neural networks outperform other methods<sup>4</sup> when it comes to exhibiting the determinants of banks' insolvency. These authors point out that the main advantages associated with ML techniques are that they allow taking into consideration many explanatory variables – which makes it possible to fully embrace the complexity of banks' balance sheet – without having to assume the shape of the relationship between those variables. Petropoulos et al. (2020) finally conclude that variables associated with capital and earnings are the best predictors of bank failures.

Part of the literature on banks' performance deals with the impact of capital on profitability. No consensus is however to be found. Some papers conclude on the existence of a positive relationship

---

<sup>3</sup>CAMELS stands for Capital adequacy, Assets quality, Management, Earnings, Liquidity, and Sensitivity.

<sup>4</sup>Specifically, the following methods are compared: Logistic Regression, Linear Discriminant analysis, Random Forests, Support Vector Machines, Neural Networks and Random Forests of Conditional Inference Trees.

between capital and profitability (Berger, 1995; Iannotta et al., 2007; Lee and Hsieh, 2013). This positive relationship is explained by the fact that well-capitalized banks face a lower probability of default than less capitalized banks, which allows the former to benefit from lower funding costs than the latter (Admati et al., 2013). Gambacorta and Shin (2018) indeed provide empirical evidence 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. In other words, increasing equity requirements could decrease banks' funding cost and thus eventually improve their performance. Other papers provide evidence of a negative relationship between capital and profitability justified by the existence of a "low-risk anomaly" and difficulties for well-capitalized banks to diversify their activities (Goddard et al., 2013; Baker and Wurgler, 2015). Fewer papers study the impact of liquidity on profitability. On the one hand, holding liquid assets decreases the maturity mismatch and thus the liquidity risk. Banks that hold a large proportion of their asset portfolio in liquid assets thus benefit from low funding costs since they are less likely to go bankrupt, which has a positive impact on their performance (Berger and Bouwman, 2009; Bordeleau and Graham, 2010). On the other hand, liquid assets generate low returns and therefore low revenues. In this perspective, banks that hold a large amount of liquid assets are likely to perform less than banks investing in riskier assets (Goddard et al., 2013; Molyneux and Thornton, 1992). From a methodological perspective, studies of banks' performance resort to a wide variety of methods, including operational research. Fethi and Pasiouras (2010) provide a survey of this literature. More recently, resorting to a stochastic frontier model, Badunenko et al. (2020) study the determinants of banks' performance for European commercial banks and point out that a sustainable business model requires different strategies depending on the time horizon considered (either the long-run or the short-run).

To our knowledge, our paper is the first to use ML methods to investigate the determinants of banks' performance based on balance sheet data. **Doing so we contribute to the existing literature in four respects:**

- in line with Mayes and Stremmel (2014), our results suggest that banking regulation should rather focus on a simple leverage ratio instead of focusing on a risk-weighted ratio,
- in line with Berger (1995); Iannotta et al. (2007); Lee and Hsieh (2013); Admati et al. (2013); Gambacorta and Shin (2018), we show that increasing capital does not impede banks' performance when measured as the ROAA, and provide evidence that the opposition of the banking industry to capital regulation may be grounded in its negative impact on the shareholder value as measured by the ROAE,
- in line with Ravi Kumar and Ravi (2007); Tanaka et al. (2016); Petropoulos et al. (2020), we show that ML methods perform well when it comes to dealing with banks' balance sheet data,
- **in line with Molnar (2020); Aria et al. (2021), we show how interpretability methods can be used so as to render ML results interesting for economists, academic finance and regulators.**

Our paper provides a major debunking of the idea according to which designing strong equity requirements would necessarily impede banks' activity and thus have an adverse impact on society as a

whole. As a consequence, it offers support to the definition of a strong constraint in equity instead of the multiplication of rules that could each have potentially harmful unexpected consequences.

### 3. Regulatory framework and tested hypothesis

This section first provides an overview of the key novelties introduced by banking regulators after the crisis. Then, we formulate the hypothesis that will be tested using various empirical methods.

#### *3.1. A brief overview of the regulatory framework*

As already mentioned, banking regulation has been renewed after the crisis. However, in spite of the diagnosis that was formulated right after the crisis (i.e., the necessity to contain systemic risk through a proper macroprudential policy), the regulatory framework that is currently implemented is somewhat disappointing (Baker, 2013). Capital requirements have indeed barely been strengthened, liquidity regulation is ill-designed, and bail-in instruments could have unexpected harmful consequences. We detail all these points here.

Capital requirements have been modified after the crisis. In particular, the definition of regulatory capital has been tightened and many hybrids that were used to be integrated in regulatory capital are now ruled out. Banks now have to hold at least 7% of their risk-weighted assets in core capital compared to 2% under Basel II. A simple leverage ratio of 3% has in addition been put in place. Core capital requirements have thus been strengthened. However, as Admati et al. (2013) state it, a lot more can be done in this direction. In fact, the implementation of bail-in standards in addition to the capital constraints defined in the first pillar of Basel III can be interpreted as a palliative for low core capital requirements (Persaud, 2014).

Both the TLAC and the MREL define constraints in what are called eligible liabilities. Those are financial instruments that are associated with a high loss-absorbing capacity in the event where the issuing bank goes bankrupt. Contingent convertible bonds (coco bonds) are such eligible liabilities. They are bonds that behave like regular bonds in normal times, but are converted into equity whenever a pre-defined event occurs. The main purpose of such bonds is to make it possible for banks to fund themselves by issuing bonds that are cheaper than equity and to make them benefit nonetheless from a loss-absorbing capacity similar to that of equity. Coco bonds are however not a panacea. They indeed could serve as a channel through which systemic risk could materialize. For instance, Corcuera et al. (2014) show that because of the conversion risk, coco bonds exhibit a death-spiral effect. To hedge the conversion risk, coco bonds' holders may indeed short sell shares. Doing so they may find themselves in a position of selling shares whose price is decreasing and, therefore, they may contribute actively to the materialization of the conversion risk. By hedging the conversion risk, investors thus make it more likely. Hence the spiral effect. Bologna et al. (2020) provide empirical evidence that shows how contagion can spread in the coco bonds' market. Using two stress episodes that affected the European coco bonds' market in 2016, the authors exhibit a significant coco bonds-specific contagion that can be the consequence of the reassessment by investors of coco bonds' riskiness. Coco bonds are thus complex financial instruments that could eventually act as a channel through which systemic risk propagates. Expectations of conversions could indeed nourish self-fulfilling panic sales of coco bonds

and thus precipitate a market-wide panic that could spread to the whole financial system (Le Quang, 2019). It is therefore of the utmost importance to keep in mind that one of the reasons why the global financial crisis was so severe was because the systemic risk associated with securitized products had been largely overlooked. The current will to ensure financial stability through complex financial instruments, such as coco bonds, thus appears as a dangerous oversight of history. This is one of the reasons why contingent capital must not be used as a palliative for equity (Admati et al., 2013).

Liquidity regulation has been implemented through two different rules: the LCR and the NSFR. The LCR states that banks need to hold enough high quality liquid assets (HQLA) to withstand a liquidity crisis lasting 30 days. In other words, banks need to hold enough liquid assets to cope with their short-term liquidity needs. The NSFR states that banks' illiquid assets need to be funded through stable funding instruments. The first question that arises when looking at current liquidity regulation is why two ratios instead of one. Having a closer look at them, we notice that the two ratios are in fact redundant. Let us demonstrate it through an example.<sup>5</sup> Consider a bank whose asset portfolio is made both of liquid and illiquid assets that are funded through runnable and not-runnable liabilities. The balance sheet of the bank is thus as follows:

Asset	Liability
Liquid assets ( $LA$ )	Runnable liabilities ( $RL$ )
Illiquid assets ( $IA$ )	Not-runnable liabilities ( $NRL$ )

The LCR is met whenever  $LA > RL \iff LA - RL > 0$ . The NSFR is met whenever  $NRL > IA \iff NRL - IA > 0$ . The balance sheet identity states that  $LA + IA = NRL + RL$ . In other words, we have  $LA - RL = NRL - IA$ . As a consequence, whenever the LCR is binding, the NSFR is binding too and *vice versa*. The two ratios are thus likely to be redundant since when one is binding, the other is necessarily binding too. Instead of two ratios, liquidity regulation would thus be better off defining only one ratio. This would be a first step towards reducing the complexity of the regulatory framework. Which ratio should then be ruled out and which should remain? The difference between the two ratios is the side of the balance sheet on which they focus: the LCR focuses on the asset side, while the NSFR focuses on the liability side. We believe for two reasons that the perspective that should be adopted is that of the NSFR. The first reason lies in the very definition of the LCR. It is defined as the ratio of HQLA over runnable liabilities. The problem is that the very fact of defining some assets as HQLA could make those assets illiquid in the event of a crisis. This idea is what has been coined as the Goodhart's law based on a quotation from a paper by Charles Goodhart according to which "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart, 1975). Gambling financial stability on the definition of liquid assets is therefore dangerous. AAA securitized products were indeed deemed rather liquid before their market froze. The second reason echoes the current concern of prudential regulators to take climate change into account (Bolton et al., 2020). To do so, financial regulation has to incentivize banks to invest in

---

<sup>5</sup>The reasoning comes from Bolton et al. (2019).



green assets in order to fill the green investment gap. Those assets are in general long-term assets. By constraining banks to invest in short-term liquid assets, the LCR could thus deter banks from investing in green assets. Consequently, such liquidity requirements are expected to have a negative impact on the access to finance of low-carbon sectors (Campiglio, 2016). Liquidity regulation should then focus more on the liability side than on the asset side of the balance sheet.

To summarize, banking regulation has evolved after the great financial crisis to tackle the new risks that materialized at the end of the 2000s. However, the current regulatory framework suffers from obvious flaws. In particular, too many rules are being implemented and none is perfectly well-designed. We think that the risk-weighted capital ratio and the bail-in standards could be replaced by a unique strong equity constraint that would be both more transparent and efficient. In addition, we think that such a constraint would not impede banks' activities. This twofold hypothesis is discussed in more detail in the next section.

### *3.2. Tested hypothesis*

From the discussion led in the previous section, it appears that the regulatory framework currently grants too much weight to complex rules (risk-weighted capital ratio, bail-in standards etc.) and disregards simpler rules (simple leverage ratio). However, strong equity requirements could help achieving the objective of all the rules that have been presented in the previous section. Equity is indeed the best funding instrument to absorb losses and thus to make sure that failing banks will be bailed-in instead of bailed-out. As Admati et al. (2013) note it, equity indeed dominates convertible debt in this matter. Liquidity regulation aims at ensuring the consistency of banks' balance sheet by limiting maturity mismatch. Before the crisis, banks indeed strongly relied on very short-term financial instruments to fund longer-term assets, which exposed them to short-term liquidity risk (Acharya et al., 2011; Morris and Shin, 2016). Defining higher equity requirements would probably make it possible to reduce the maturity mismatch and thus to reduce the exposure of banks to liquidity risk without relying on a questionable definition of which assets are deemed highly liquid and which are not. In this perspective, strong equity requirements appear as a better solution than the definition of liquidity ratios to ensure financial stability and as a better solution than bail-in standards to protect taxpayers from costly bailouts. In this paper, we start from the idea that equity indeed performs better than both bail-in standards and liquidity ratios in fulfilling their goals.

Starting from this idea, we inquire whether the arguments opposed by the banking industry to higher equity requirements are justified or not. The main argument against such requirements is that they would have a negative impact on the ability of banks to conduct their activities by impeding their performance. We aim at providing empirical evidence that the relationship could actually be the reverse. Since higher equity requirements reduce the probability that banks end up bankrupting, it is expected that better-capitalized banks are able to issue bonds at a lower cost. All in all the total funding cost of banks could decrease and their profitability thus increase. If this is indeed the case, nothing opposes the definition of higher equity requirements. This is the hypothesis that is tested in this paper.

## 4. Models and interpretation

To assess the impact of the explanatory variables on banks' performance we resort to four models: Lasso, RF, ANN, and Linear SVR. As mentioned in section 2, these models perform well when dealing with probability of default. In this paper, we investigate the performance of these models in assessing the relationship between balance sheet variables and banks' performance when the research question is phrased as a regression problem. We then provide an economic interpretation of the results associated with the model that performs the best. To this aim, we use several interpretation methods. On the one hand, we resort to variables' importance measures (Breiman et al., 1984; Molnar, 2020) to estimate the predictive power associated with each explanatory variable. On the other hand, we rely on marginal effect methods to determine the nature of the impact of the explanatory variables on the dependent variable. In particular, we resort to Partial Dependence Plots (PDPs) and Accumulated Local Effects (ALEs).

### 4.1. Lasso

Lasso models (Tibshirani, 1996) are frequently resorted to when the number of determinants to deal with is large. In this paper, Lasso is used as the benchmark standard linear model. It works as a shrinkage method by imposing a penalty on coefficients' size. The complexity parameters are chosen so as to minimize an information criterion. In the paper, two criteria are used: the Akaike information criterion (AIC) and the Bayes information criterion (BIC).

### 4.2. Random Forest

Random Forest (Breiman, 2001) is an ensemble method based on the aggregation of decision trees. A decision tree consists in a recursive binary partition of the space of explanatory variables into final regions in which an output value is predicted through a simple model. The number of trees, their depth, and the minimal number of observations per final leaf are chosen through cross-validation. This model gives the possibility to manipulate a large number of explanatory variables without imposing any functional form on the relationship between these variables and the output. The implementation of results' interpretation is relatively straightforward (Hastie et al., 2009; Molnar, 2020).

### 4.3. Artificial Neural Networks

Artificial Neural Networks (McCulloch and Pitts, 1943; Hastie et al., 2009) model links between features and explained variables through the application and composition of non-linear functions. In order to reduce complexity, we resort to the most widely used neural network, namely the single hidden layer back-propagation network.<sup>6</sup> In that case, there is only one hidden layer between the inputs and the output. Cross-validation is used to find the number of neurons, the batch size, and the maximum number of iterations in the optimization process. A standard procedure in data treatment when using ANN models is to normalize variables. Since balance sheet variables are characterized by large values' differences, we run ANN regressions on normalized data. Interpretation methods can nonetheless be implemented, but are less easily given an economic interpretation. The same is done for Linear SVR.

---

<sup>6</sup>The use of a one-layer perceptron in a regression problem can explain why PDPs are linear (see Appendix C.2). In that case, the output function is indeed the identity function (Hastie et al., 2009).

#### 4.4. Linear Support Vector Regression

Support Vector Machines are generally used in classification problems but can be adapted for regression purposes. SVR models (Boser et al., 1992) are non-linear and large-margin regression models in which parameters are estimated under a constraint that allows for a certain error. Linear SVRs use a linear kernel, which allows sparing computational time (Ho and Lin, 2012). The penalty term and the epsilon are found using cross-validation.

#### 4.5. Interpretation: variables' importance

To determine the predictive power of the explanatory variables we use three measures of importance. In RF regressions, we rely on a generalization of Breiman et al. (1984)'s measure of relevance for a single tree. It averages the sum of improvements across all trees, given by each variable when it is chosen as the splitting one. This measure attributes a standardized score to all features, which gives their determining power on labels. For all models, we use permutation importance (Molnar, 2020) that measures to what extent the prediction of the model changes when randomly shuffling each variable. Finally, we compute SHAP summary plots that classify variables according to their importance and give an idea of how the dependent variable is influenced by each of them.

#### 4.6. Interpretation: marginal effects

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  $k$ . 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. 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 rely on ALEs 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.

## 5. Data and descriptive statistics

### 5.1. Data

The objective of our empirical strategy is to determine what the main predictors of banks' performance are in order to decide whether or not increasing equity requirements would have the negative impact put forward by the banking industry to oppose them. To do so, we focus on two key measures of banks' performance: the return on average assets (ROAA) and the return on average equity (ROAE). Both variables are frequently used as measures for banks' performance in the literature.<sup>7</sup> There is however an essential difference between these two variables: the ROAE indicates to what extent this performance could eventually specifically benefit to the shareholders.

---

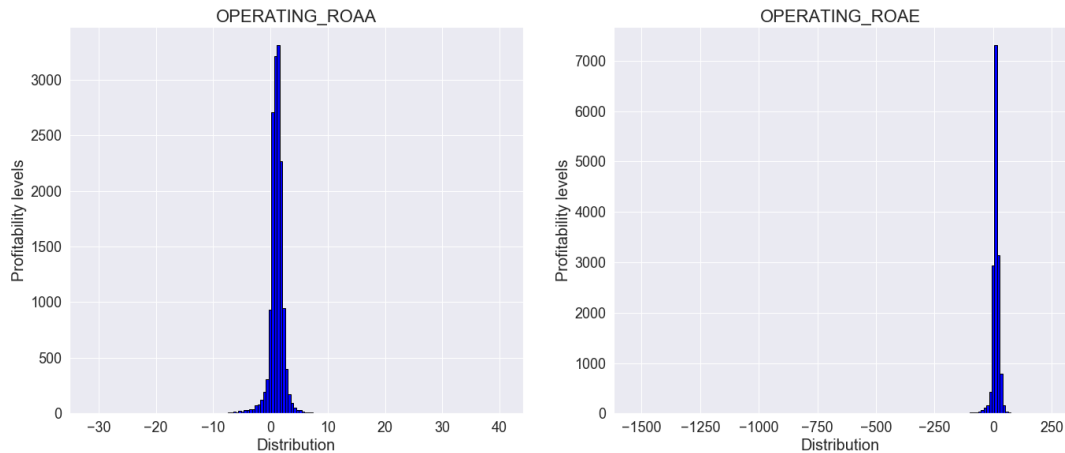
<sup>7</sup>ROAA: Berger et al. (1995); Osborne et al. (2012); Tran et al. (2016); Xu et al. (2019). ROAE: Berger (1995); Osborne et al. (2012); Distinguin et al. (2013); Tran et al. (2016); Xu et al. (2019).

Explanatory variables and the corresponding data sources are presented in [Appendix A](#). The initial database comprised 62 variables and 15310 observations.<sup>8</sup> Rearranging the database to remove twin variables (i.e., those with a very similar definition and measurement), we finally kept 36 variables.<sup>9</sup> Among those variables, three are worth pointing out. The first one is equity over total assets ( $\frac{E}{TA}$ ). This variable accounts for the proportion of the asset portfolio that is funded through equity. The higher the ratio, the better the bank is capitalized. The main purpose of our empirical strategy is to assess to what extent this ratio impacts banks' performance and in which direction. Higher equity requirements would indeed translate into higher values of this ratio. The other variables worth pointing out are the total capital ratio (TCR) and liquid assets over total assets ( $\frac{LA}{TA}$ ). The latter variable is used as a proxy for the LCR: since the LCR has only recently been implemented – in fact it is not fully implemented yet –, data on the LCR are insufficient. The ratio  $\frac{LA}{TA}$  is however often used as a proxy for the LCR.<sup>10</sup>

Data are annual and come from the FitchConnect database. Our database comprises twenty one countries<sup>11</sup> selected for their economic proximity and the fact that they implement similar banking regulation.

## 5.2. Descriptive statistics

Figure 1: Distributions of the dependent variables



Source: Authors' calculations from FitchConnect data.

Before presenting the results, we display some descriptive statistics concerning the variables. Figure 1 shows the distribution of the dependent variables (ROAA and ROAE). As can be seen, the values

<sup>8</sup>An observation is defined as a bank for a given year. The number of banks evolves through time.

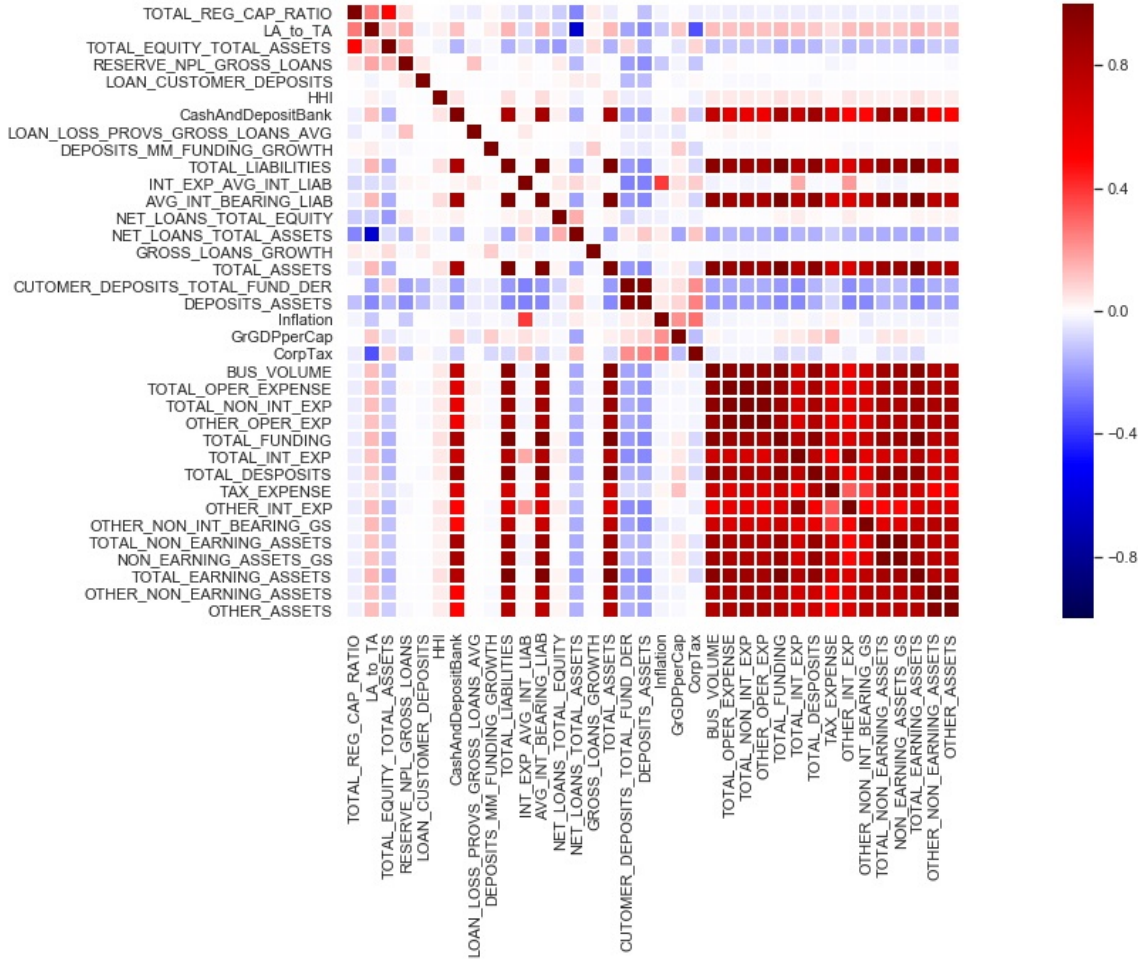
<sup>9</sup>We also constrained the dataset to include some accounting and macroeconomic variables identified by the literature.

<sup>10</sup>As stated in this recent paper published by the Bank for International Settlements: "Studies of Basel III's Liquidity Coverage Ratio (LCR) or Net Stable Funding Ratio (NSFR), for example, have so far been conducted using relatively coarse approximations, such as the ratio of liquid assets or deposits to total assets : banks have started to disclose their Basel III LCR or NSFR in their financial statements only recently." (Boissay et al., 2019, p.57)

<sup>11</sup>Belgium, Canada, China, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Luxembourg, the Netherlands, Norway, Poland, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom and the United States.

of these variables are strongly concentrated around their mean value. Both variables display some outliers, but the methods presented in the previous section are robust even in the presence of such observations. Figure 2 presents the linear correlation between all explanatory variables. We notice that some of them are highly correlated. This strong correlation may bias the computation of PDPs. This is the reason why ALEs are computed as a robustness check.

Figure 2: Linear correlation between features



Source: Authors' calculations from FitchConnect data.

## 6. Results

In this section, we present the results of the different regressions that have been run. We first compare the performance of the models by comparing their  $R^2$  both in- and out-of-sample. We then present the main results regarding the impact of the ratio  $\frac{E}{TA}$  on the ROAA and the ROAE.

### 6.1. Performance of the models

Table 1 displays the in- and out-of-sample coefficients of determination for all the models considered for both dependent variables (ROAA and ROAE). Having a look at the results when ROAA is the

dependent variable, we notice that Lasso regressions perform the least both in- and out-of-sample. ANN and SVR display rather identical performances, while RF strongly outperforms all the other models. When ROAE is the dependent variable, results are not that clear cut. Lasso still underperforms by comparison to the rest of the models and RF regressions outperform ANN and SVR in-sample. However, having a look at the out-of-sample performances of the models, we notice that both ANN and SVR perform better than RF.

Table 1: The performance of the models

(a) ROAA as the dependent variable

Sample	Models				
	Lasso (AIC)	Lasso (BIC)	RF	ANN	Linear SVR
In-sample	0.15	0.15	0.79	0.19	0.18
Out-of-sample	0.12	0.12	0.33	0.17	0.16

(b) ROAE as the dependent variable

Sample	Models				
	Lasso (AIC)	Lasso (BIC)	RF	ANN	Linear SVR
In-sample	0.11	0.11	0.82	0.24	0.23
Out-of-sample	0.11	0.11	0.17	0.23	0.23

Source: Authors' calculations. Tables show the in- and out-of-sample coefficients of determination ( $R^2$ ) for the models. In the Lasso model, variables selection is done using both the Akaike information criterion (AIC) and the Bayes information criterion (BIC).

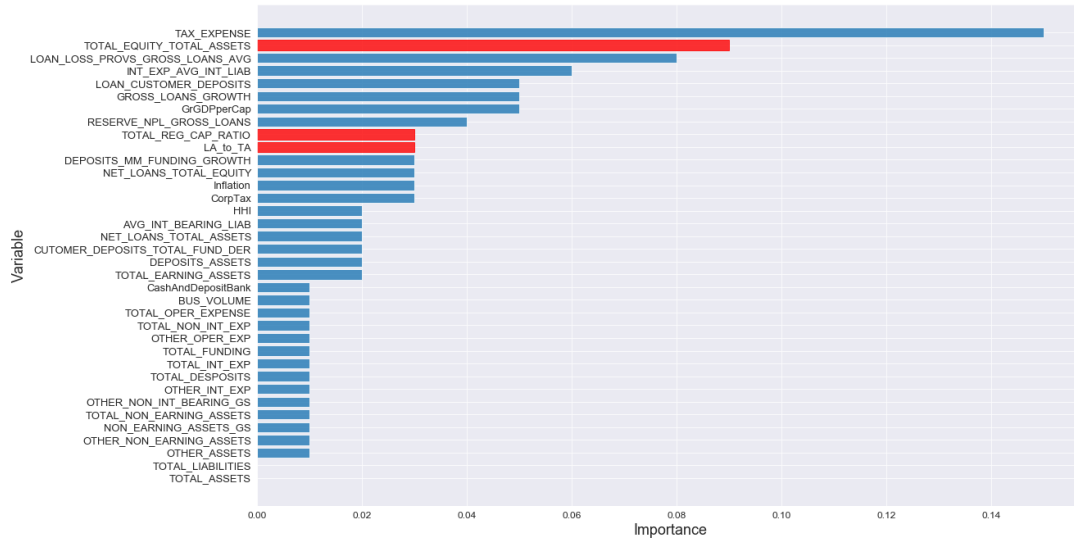
To check the validity of our models, we have a look at the shape of residuals distribution for the two dependent variables both in- and out-of-sample. In all cases, the Jarque-Bera test rejects the null hypothesis of normality. However, all residuals are symmetrically distributed around their mean, which is always close to zero. In addition, distributions all display low standard deviations for ROAA. Important outliers seem to drive higher standard deviations in the case of ROAE. Figure B.15 in Appendix B displays residuals distribution for RF models both in- and out-of-sample. Table B.5 in Appendix B presents several statistics concerning the residuals.

## 6.2. When $\frac{E}{TA}$ increases banks' performance (ROAA)...

Since RF regressions outperform the other methods when ROAA is the dependent variable, we report only the results of these regressions in the main text. The remaining results can be found in Appendix C. Figure 3 presents the importance of each explanatory variable at predicting ROAA. As could have been expected, variables associated with the specificity of banks' business model prove to be great predictors of ROAA. Recall that banks rely on short-term deposits to fund long-term investments, especially through loans. It is therefore natural that variables associated with an increase in the amount of loans granted – for instance loan loss provisions, loans to customer deposits, gross loans etc. – count among the main predictors of ROAA. Similarly, it is not surprising that variables impacting directly the cost of the resources thanks to which banks fund their activity – for instance interest expense or growth rate of deposits to money market funding – also rank among the main predictors of ROAA. Variables describing the macrodynamics of the economy – i.e., GDP growth and

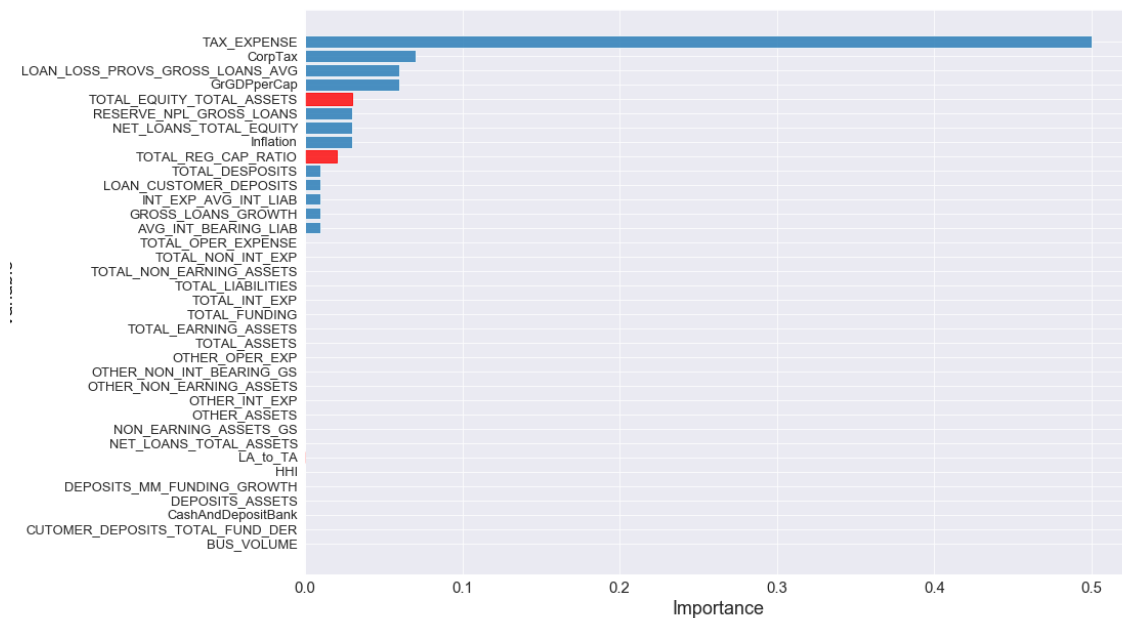
inflation – are also significant predictors of ROAA. This was expected since they are largely correlated with the demand for credit coming from the real economy. Concerning the three variables we are interested in, we notice that  $\frac{E}{TA}$  is the second best predictor of ROAA, just behind tax expense, which is a trivial determinant of ROAA. TCR and  $\frac{LA}{TA}$  are associated with a weaker predictive power, but still remain significant determinants of ROAA.

Figure 3: In-sample predictive power of each variable on ROAA



Source: Authors' calculations from FitchConnect data.

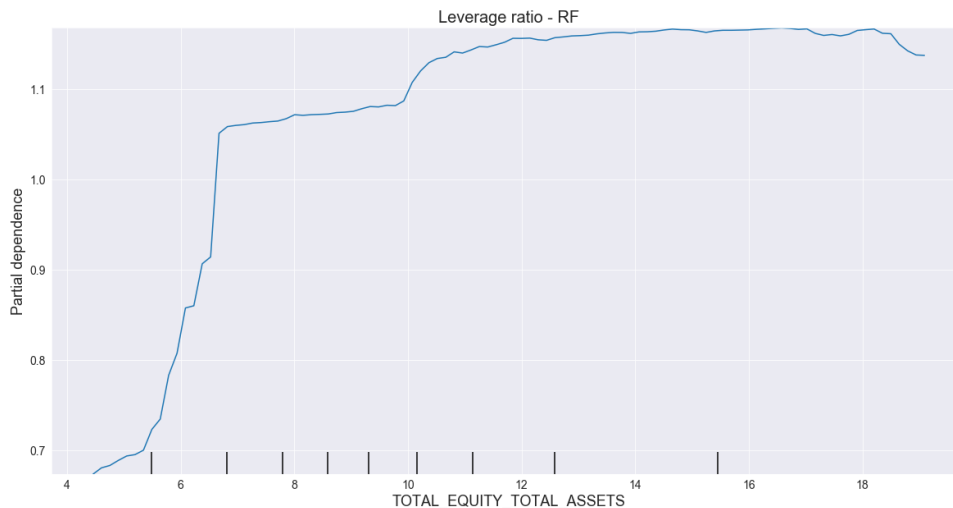
Figure 4: Out-of-sample predictive power of each variable on ROAA



Source: Authors' calculations from FitchConnect data.

Figure 4 displays the percentage of the out-of-sample fit due to each explanatory variable. We notice that  $\frac{E}{TA}$  and TCR remain significant determinants of ROAA. Variables describing the specificity of the business model of banks – i.e., variables associated with the credit activity of banks – naturally remain strong predictors of ROAA. To determine in which direction goes the effect of the explanatory variables on ROAA, we use PDPs. ALEs are reported in Appendix D.1 and serve as robustness checks. The marginal impact of  $\frac{E}{TA}$  on ROAA is displayed in Figure 5.

Figure 5: Marginal impact of  $\frac{E}{TA}$  on ROAA



Source: Authors' calculations from FitchConnect data. The distribution of  $\frac{E}{TA}$  is reported on the x-axis (deciles).

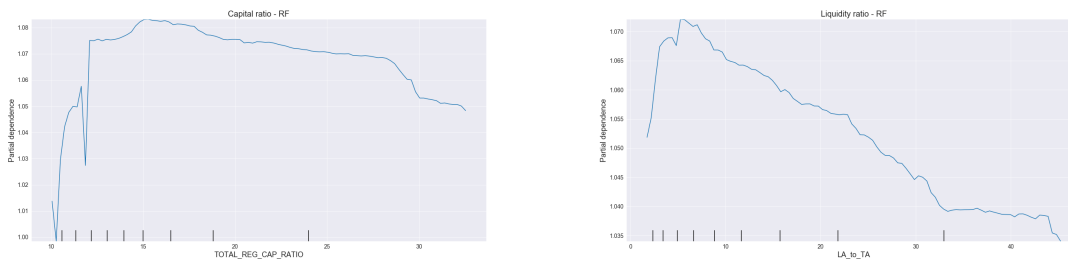
We notice that  $\frac{E}{TA}$  has a marginal positive impact on ROAA. More precisely, the impact of  $\frac{E}{TA}$  on ROAA looks S-shaped: it is first weakly positive, then strongly positive and, for the highest values of  $\frac{E}{TA}$ , weakly positive again. The statistical significance of this impact is inferred from Figure 3: the strong predictive power of variable  $\frac{E}{TA}$  on ROAA can indeed be translated into a significant impact of the former on the latter.<sup>12</sup> Robustness of this result has been ensured through cross-validation. No matter the initial value of the ratio  $\frac{E}{TA}$ , increasing equity requirements would therefore have a positive impact on the performance of banks as measured by the ROAA. Higher values of  $\frac{E}{TA}$  are associated with lower probabilities of default and therefore with lower funding costs. Banks whose funding structure strongly relies on equity thus face low funding costs, which allow them to perform well. In sum, looking at the impact of the ratio  $\frac{E}{TA}$  on banks' performance does not support the view according to which banking regulation hampers banks' activities.

Let us consider the impact of TCR and  $\frac{LA}{TA}$  to assess whether this view can however be defended. Recall that the latter serves as a proxy for the LCR. Results are displayed in Figure 6.

<sup>12</sup>We additionally ran a fixed-effect regression on a sub-sample of 350 banks (sample had to be reduced to ensure cylindricity): the impact of  $\frac{E}{TA}$  on ROAA was indeed positive and significant at 1%.



Figure 6: Marginal impacts of TCR and  $\frac{LA}{TA}$  on ROAA

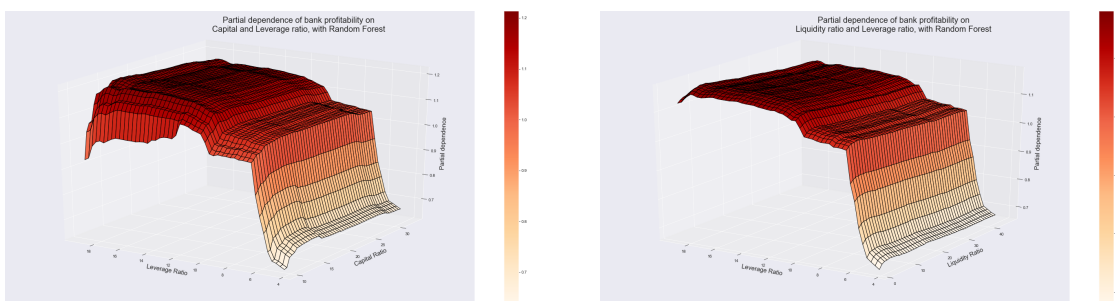


Source: Authors' calculations from FitchConnect data. The left plot displays the PDP for TCR, while the right plot displays the PDP for  $\frac{LA}{TA}$ . The distributions of TCR and  $\frac{LA}{TA}$  are reported on the x-axes (deciles).

TCR has a non-linear impact on ROAA. Below a threshold value of approximately 16%, TCR has a positive impact on ROAA, while the effect is weakly negative for values of TCR lying above this threshold value. The idea according to which capital regulation could eventually have a negative impact on banks' performance seems to hold only for high capital ratios.<sup>13</sup> Similarly, the marginal impact of  $\frac{LA}{TA}$  on ROAA is positive below a certain threshold and negative above it. The effects of TCR and  $\frac{LA}{TA}$  on ROAA are however weaker than the impact of  $\frac{E}{TA}$  on ROAA: Figure 3 indeed shows that the predictive power of the former variables is weaker than that of the latter.

It is noticeable that the impact of TCR and that of  $\frac{E}{TA}$  on ROAA do not go the same direction: the impact of TCR is first positive then negative, while that of  $\frac{E}{TA}$  is always positive. That is quite surprising since those two variables both account for banks' capitalization. Recall however that TCR is defined as the ratio of regulatory capital (which includes both Tier 1 capital and Tier 2) over risk-weighted assets, while  $\frac{E}{TA}$  is the simplest capital ratio possible: the ratio of equity over total assets. What is thus to be learned from the results presented so far is that a simple capital regulation (based on a simple definition of capital and risk) never impedes banks' performance, while the sophisticated Basel III ratio does so for capital ratios above a certain threshold.

Figure 7: Two-way PDPs (ROAA)



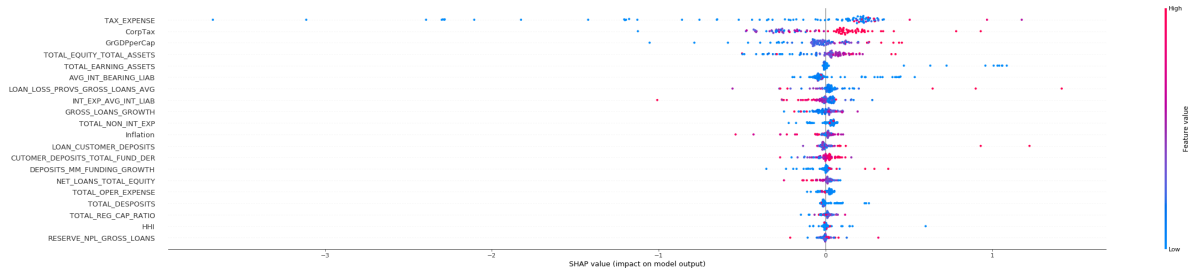
Source: Authors' calculations from FitchConnect data. The left plot displays the marginal impact of the interaction between  $\frac{E}{TA}$  and TCR on ROAA, while the right plot shows the marginal impact of the interaction between  $\frac{E}{TA}$  and  $\frac{LA}{TA}$  on ROAA.

<sup>13</sup>16% is indeed higher than the sum of all capital requirements combined.

Figure 7 presents the effect of the interaction between  $\frac{E}{TA}$  and the other two variables (TCR and  $\frac{LA}{TA}$ ) on ROAA. As for the marginal impact of each of these variables, we resort to PDPs. As was expected from Figure 3, the impact of  $\frac{E}{TA}$  on ROAA is far stronger than the effects of  $\frac{LA}{TA}$  and TCR. Having for instance a look at the left plot, we indeed notice that the impact of the interaction between  $\frac{E}{TA}$  and TCR on ROAA is entirely driven by  $\frac{E}{TA}$ . The same goes for the interaction between  $\frac{E}{TA}$  and  $\frac{LA}{TA}$ .

Another way to estimate the contribution of the variables at predicting ROAA is to resort to the Shapley value. 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. Shapley values are displayed in Figure 8.

Figure 8: Shapley value of each feature (ROAA)



Source: Authors' calculations from FitchConnect data. The lines display the Shapley value of each feature: a negative value implies that the feature negatively contributes to the predicted value of ROAA, a positive value implies that the feature positively contributes to the predicted value of ROAA.

Shapley values allow to have another insight on the results presented so far. For most of its values,  $\frac{E}{TA}$  has indeed a strong positive impact on ROAA: Shapley values mostly lie to the right of the 0. More specifically, most values that are associated with a negative contribution are small values of  $\frac{E}{TA}$ , medium and high values of this ratio being associated with a positive impact on ROAA. This twofold result is consistent with what has been learned from the PDP (Figure 5): the impact of  $\frac{E}{TA}$  on ROAA is most of the time positive and it is expected to be stronger for medium values of  $\frac{E}{TA}$  (the PDP looks S-shaped). The impact of TCR on ROAA is less clear, but seems rather positive for small values of TCR.

To summarize,  $\frac{E}{TA}$  has a strong and positive effect on ROAA. On the contrary, TCR and  $\frac{LA}{TA}$  have a weaker impact on ROAA and are thus less convincing determinants of banks' performance. The so-called negative impact of capital regulation on banks' activities is therefore likely to be very small when capital regulation is implemented through sophisticated ratios (such as TCR) and non-existent when it is designed as a much simpler ratio (such as  $\frac{E}{TA}$ ).

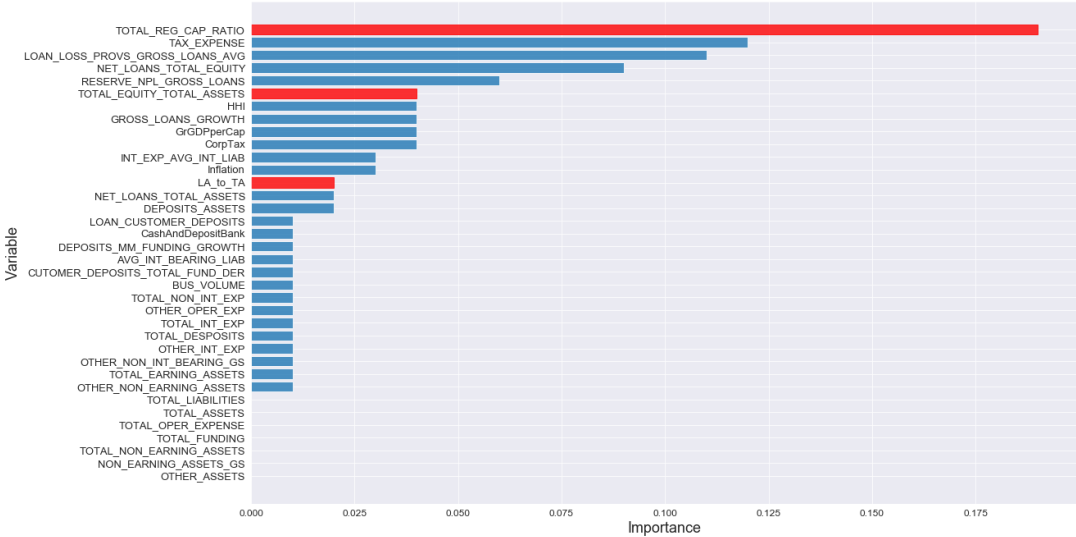
### 6.3. ... but has a negative marginal impact on the shareholder value (ROAE)

In the previous section, we presented results supporting the idea that stronger equity requirements could have a positive impact on banks' performance when measured as the ROAA. This is in contradiction with the common wisdom, as put forward by the banking industry to oppose any strengthening

of capital regulation. In this section, we resort to another measure of banks' performance (the ROAE) to try to make sense of this paradox. When ROAE is the dependent variable, recall that Table 1 in section 6.1 shows that RF regressions outperform other models in-sample, but are outperformed out-of-sample by ANN and SVR. In this case, results drawn from ANN and SVR are therefore to be considered with attention. These results are presented and interpreted in Appendix C. We present here the results of RF regressions.

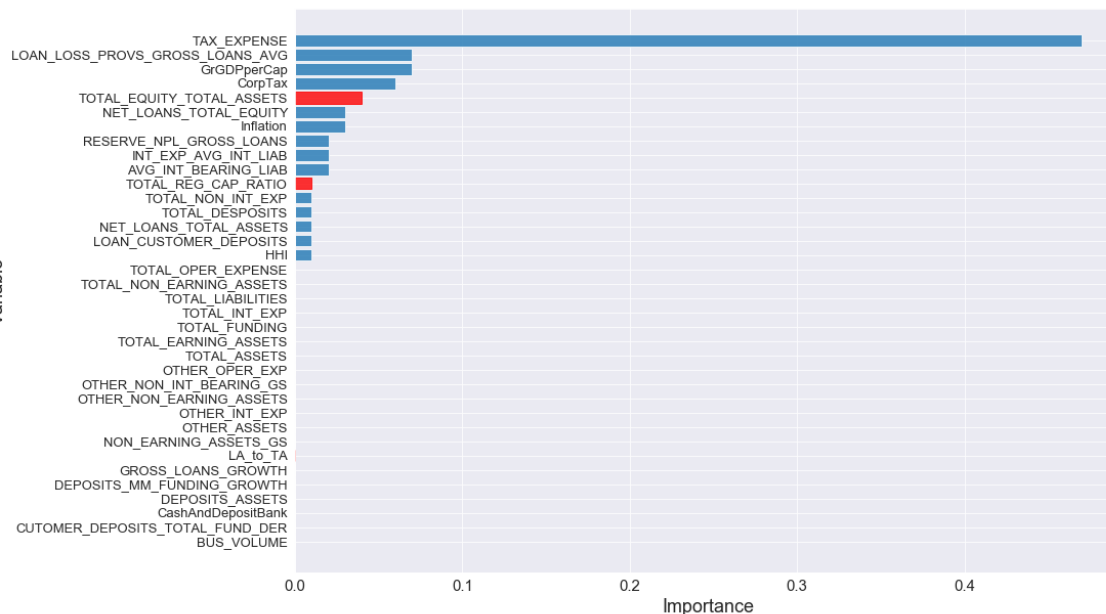
Figure 9 displays the predictive power of the explanatory variables on ROAE in the RF regression. TCR is the strongest determinant of ROAE.  $\frac{E}{TA}$  is a weaker but still strong determinant of ROAE since it ranks sixth among all predictors. The impact of  $\frac{LA}{TA}$  is even weaker. Figure 10 displays the percentage of out-of-sample fit that can be attributed to each explanatory variable. In particular, we notice that  $\frac{E}{TA}$  remains a strong predictor of ROAE, while the predictive power of TCR is weaker. Turning to the marginal impact of  $\frac{E}{TA}$  on ROAE (Figure 11), we notice that increasing the value of this ratio is expected to have a positive impact on ROAE for values of this ratio below approximately 8%. For values of  $\frac{E}{TA}$  above this threshold, the relationship is the other way round. In other words, up to a certain point, increasing the value of the ratio  $\frac{E}{TA}$  has a positive impact on the shareholder value. However, above this point, stronger equity requirements would have a negative impact on the shareholder value. This may be the reason why banks oppose so fiercely every attempt to strongly increase equity requirements.

Figure 9: In-sample predictive power of each variable on ROAE



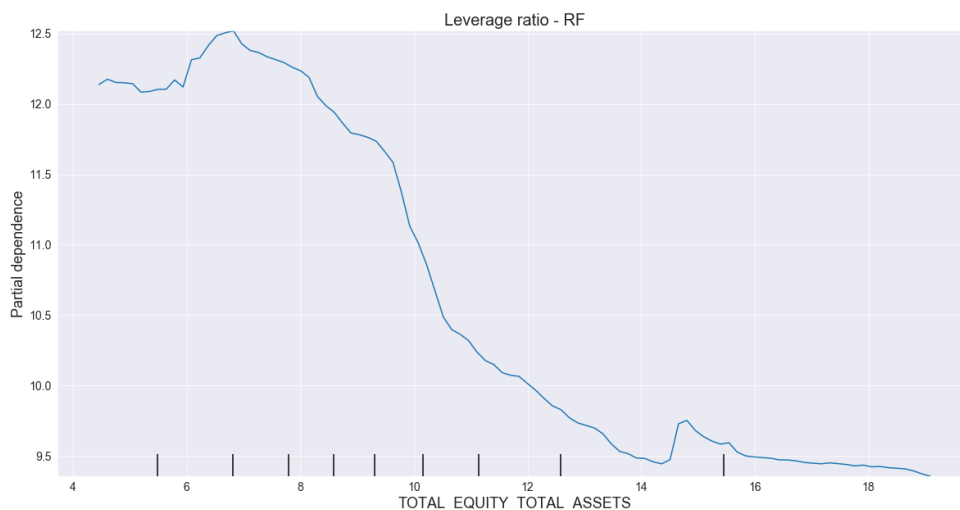
Source: Authors' calculations from FitchConnect data.

Figure 10: Out-of-sample predictive power of each variable on ROAE



Source: Authors' calculations from FitchConnect data.

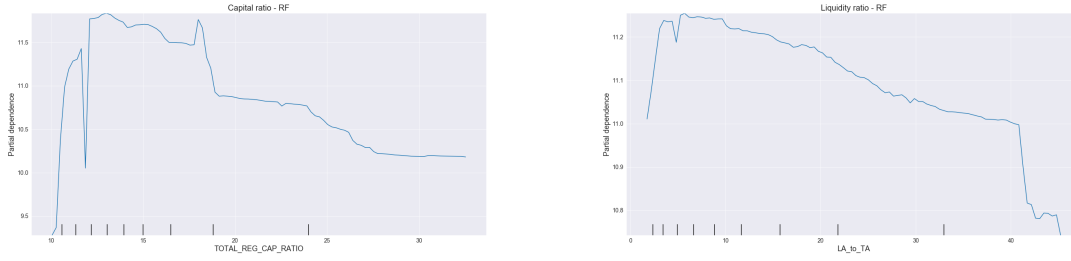
Figure 11: Marginal impact of  $\frac{E}{TA}$  on ROAE



Source: Authors' calculations from FitchConnect data. The figure displays the PDP of  $\frac{E}{TA}$ . The distribution of  $\frac{E}{TA}$  is reported on the x-axis (deciles).

The marginal impacts of TCR and  $\frac{LA}{TA}$  on ROAE are roughly the same as those on ROAA. Results are shown in Figure 12. TCR has a non-linear effect on ROAE: a positive impact below a threshold value and a negative or null impact above it. The same goes for the impact of  $\frac{LA}{TA}$  on ROAE. However, given the predictive power of each variable as presented in Figure 9, the impact of TCR on ROAE is expected to be stronger than those of  $\frac{LA}{TA}$  and  $\frac{E}{TA}$ .

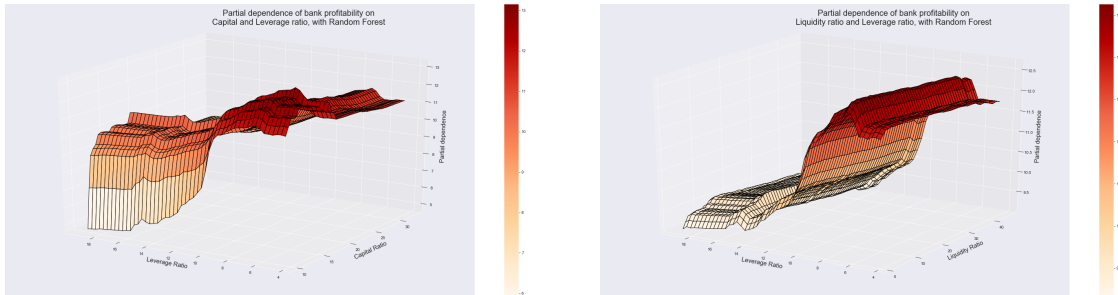
Figure 12: Marginal impacts of TCR and  $\frac{L}{TA}$  on ROAE



Source: Authors' calculations from FitchConnect data. The left plot displays the PDP for TCR, while the right plot shows the PDP for  $\frac{L}{TA}$ . The distributions of TCR and  $\frac{L}{TA}$  are reported on the x-axes (deciles).

Let us have a look at the impact on ROAE of the interaction between  $\frac{E}{TA}$  and TCR, on the one hand, and  $\frac{E}{TA}$  and  $\frac{L}{TA}$  on the other (Figure 13). We notice (left plot) that the impact of TCR on ROAE weakly dominates that of  $\frac{E}{TA}$ . On the contrary, the impact of this latter variable strongly dominates that of  $\frac{L}{TA}$ . From the plot on the left, we notice that shareholders are better off when the ratio  $\frac{E}{TA}$  is very low. In that case, even strong increases in the value of TCR do not impact much the ROAE. If banks want to maximize their ROAE, their best interest is therefore to lobby for the implementation of a complex capital ratio instead of a simple leverage ratio: the lowest the latter, the better banks are. This is precisely the current shape of capital regulation: it is mostly designed around a complex capital ratio (which negative impact on ROAE is limited), while the regulatory leverage ratio is set at a very low level.

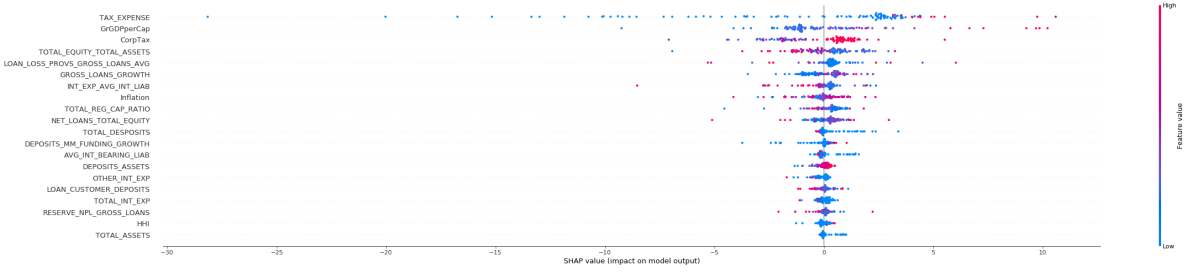
Figure 13: Two-way PDPs (ROAE)



Source: Authors' calculations from FitchConnect data. The left plot displays the impact of the interaction between  $\frac{E}{TA}$  and TCR on ROAE, while the right plot shows the impact of the interaction between  $\frac{E}{TA}$  and  $\frac{L}{TA}$  on ROAE.

Here also, Shapley values provide another insight on these results. They are displayed in Figure 14. Shapley values confirm the diagnosis established concerning the impact of  $\frac{E}{TA}$  on ROAE: small values of  $\frac{E}{TA}$  contribute positively to the prediction of ROAE (they lie to the right of the 0), while large values are associated with a negative impact on ROAE (they lie to the left of the 0). In other words, equity requirements may have a positive impact on the ROAE up to a certain threshold. Above this threshold, equity requirements are expected to have a negative impact on the ROAE.

Figure 14: Shapley value of each feature (ROAE)



Source: Authors’ calculations from FitchConnect data. The lines display the Shapley value of each feature: a negative value implies that the feature negatively contributes to the predicted value of ROAE, a positive value implies that the feature positively contributes to the predicted value of ROAE.

To conclude,  $\frac{E}{TA}$  has a negative impact on ROAE when it is above a threshold of approximately 8%. Implementing stronger equity requirements would therefore have an adverse impact on the shareholder value of banks. On the contrary, TCR has a strong non-linear effect on ROAE: a positive one below 16%, and a negative or null one above this threshold. If banks look at maximizing their shareholder value, they are, therefore, incentivized to oppose strong equity requirements to prefer instead much complex capital ratios. This is precisely how current capital regulation is implemented. Having a look at the impact of regulatory variables on ROAE, we therefore manage to formulate an explanation to make sense of the paradox mentioned at the beginning of this section. Despite the positive impact of the ratio equity over total assets on banks’ performance when measured as the ROAA, it is likely that banks oppose strong equity requirements (i.e., above 8%) since they negatively affect their shareholder value.

## 7. Robustness

### 7.1. ALEs

As mentioned in section 5.2, the existence of a strong linear correlation between some of the explanatory variables may have biased the computation of the PDPs that have been used so far to determine the nature of the impact of these variables on ROAA and ROAE. To check the robustness of our results, we thus compute ALEs. ALEs are displayed in Appendix D.1. In particular, Figure D.19 displays the ALEs for RF regressions. ALEs confirm the results drawn from the interpretation of PDPs:  $\frac{E}{TA}$  has a S-shaped effect on ROAA (plot (a) in Figure D.19) and impacts positively ROAE up to a threshold of approximately 8% above which the impact becomes negative (plot (b) in Figure D.19). Results concerning TCR and  $\frac{L}{TA}$  are also consistent with what has been concluded from the reading of PDPs. ALEs are also provided for ANN (Figure D.20) and for SVR (Figure D.21). Recall that in those cases, we focus mostly on the results when ROAE is the dependent variable (see Appendix C.2 and Appendix C.3). In that case, we notice that, consistently with the PDPs, ALEs suggest that the impact of  $\frac{E}{TA}$  on ROAE is most of the time expected to be negative.

## 7.2. Classification

As another robustness check, we rephrase our research question so as to transform it in a classification problem. To do so, we split our sample into two groups: one comprising banks for which the value of the dependent variable (ROAA and ROAE) is below its median value (group 0), and another comprising banks with a value of the dependent variable above its median value (group 1). In this setup, we look at the impact of the explanatory variables on the probability for banks to go from group 0 to group 1 – or the probability to stay in this latter for banks that are already in it. Classification is done through ANN, RF, Support Vector Classification (SVC) and Logit. As in section 6.1, we first compare the performance of the different models. Performance is here measured as the proportion of banks that models manage to correctly classify. Table 2 presents the performance of the models. Two things are worth noticing. First, all models perform well both in- and out-of-sample. Second, RF always outperforms the other models.

Table 2: The performance of the models (classification)

(a) ROAA as the dependent variable

Sample	Models			
	Logit	RF	ANN	Linear SVC
In-sample	0.65	0.97	0.83	0.74
Out-of-sample	0.64	0.84	0.75	0.72

(b) ROAE as the dependent variable

Sample	Models			
	Logit	RF	ANN	Linear SVC
In-sample	0.69	0.98	0.83	0.72
Out-of-sample	0.69	0.86	0.79	0.70

Source: Authors' calculations. Tables show the proportion of banks correctly classified by each model.

Results are presented for the two models that perform the best, namely RF and ANN. Results are reported in Appendix D.2. Having a look at the results of RF (Figure D.22) we observe that the results obtained from classification are the same as those obtained from regression.  $\frac{E}{TA}$  is indeed a strong predictor of ROAA (plot (a) in Figure D.22) and its impact on ROAA is positive (plot (c) in Figure D.22): when  $\frac{E}{TA}$  increases, the probability that a bank goes from group 0 to group 1 increases. In addition, we notice that in line with what has been said so far, the PDP for  $\frac{E}{TA}$  (plot (c) in Figure D.22) is S-shaped. When ROAE is the dependent variable,  $\frac{E}{TA}$  remains a strong predictor (plot (b) in Figure D.22) and an increase in  $\frac{E}{TA}$  positively impacts the probability with which banks go from group 0 to group 1 up to a threshold value of 8% (plot (d) in Figure D.22). Above this threshold, an increase in  $\frac{E}{TA}$  negatively impacts the probability that banks go from group 0 to group 1. Results drawn from ANN (Figure D.23) confirm those of RF.  $\frac{E}{TA}$  is indeed a strong predictor of both ROAA (plot (a) in Figure D.23) and ROAE (plot (b) in Figure D.23). In addition, the impact of an increase in  $\frac{E}{TA}$  on the probability that banks go from group 0 to group 1 is positive when ROAA is the dependent variable (plot (c) in Figure D.23) and negative when ROAE is the dependent variable (plot (d) in Figure D.23).

### 7.3. Bank-specific and country-specific fixed effects

As a robustness check, we control for bank-specific and country-specific fixed effects. More specifically, we introduce country dummies and "systemicity" dummies.<sup>14</sup> Doing so for RF regressions, we notice that dummies do not appear as determinants of either of the dependent variables and that the performance (Table 3) of the model remains unchanged.

Table 3: The performance of RF regressions (fixed effects)

Sample	Dependent variable	
	ROAA	ROAE
In-sample	0.79	0.82
Out-of-sample	0.32	0.18

Source: Authors' calculations. The table shows the in- and out-of-sample coefficients of determination ( $R^2$ ) for RF when ROAA and ROAE are the dependent variables and bank-specific and country-specific fixed effects are taken into account.

## 8. Methodological contribution and discussion

There exists a gap between what is going on in the financial industry where ML methods are gaining ground and academic finance where only few papers resort to these methods. Indeed, while ML methods are being used to implement trading strategies (Krauss et al., 2017) or as "challenger" models in credit scoring (Dumitrescu et al., 2021; Lessmann et al., 2015), academic finance tends to disregard ML methods in favor of standard econometric approaches (Krauss et al., 2017).

This gap can be explained by the lack of interpretability of learning approaches that are often considered as black boxes. This is especially the case as far as RF models are concerned. As a consequence, even if RF models outperform standard approaches in terms of predictive power, they are often considered as inadequate for disciplines such as economics where causal relationships between variables need to be clearly established (Aria et al., 2021).

In line with these two comments, the methodological contribution of the paper is twofold. First, we resort to ML to deal with an essential question for academics, practitioners and regulators. Doing so, we contribute to introducing the use of ML into economics and academic finance and show, in line with previous studies (Dumitrescu et al., 2021; Götze et al., 2020), that ensemble learning methods such as RF outperform other methods (lasso, logistic regressions, ANN and SVM). This result can be explained because RF models allow dealing with non-linearities and therefore provide some insights on threshold effects and interactions between variables, which are both essential to take into consideration when dealing with balance sheet data. Second, to overcome the lack of interpretability of RF we resort to various interpretation methods: PDP, ALE and Shapley Value. Doing so, we contribute to the literature on interpretable ML (Aria et al., 2021; Molnar, 2020) and provide evidence that the

---

<sup>14</sup>Banks listed either as global systemically important (G-SIBs) or as domestic systemically important (D-SIBs) are considered as systemic.



relationships between the equity ratio, the ROAA and the ROAE are not linear but subject to tipping points. Thanks to RF, we were able to identify those tipping points, which would not have been possible resorting only to standard econometric approaches.

Therefore, this paper calls for introducing RF (and more generally ML) into the toolbox of economics and finance researchers. Indeed, even if those methods are often considered as black boxes – which they can be when blindly implemented – and therefore regarded with caution by academics in finance and economics, there now exist several solutions to interpret results so as to shed light on economics and financial issues.

## 9. Conclusion

Banking regulation currently faces multiple challenges. To deal with them, it is very tempting to define one rule per problem to solve. This is what has been done in the recent years: the solvency risk is dealt with through risk-based capital ratios, the liquidity risk through liquidity ratios, and the cost associated with banks' failures through bail-in standards. Banking regulation is thus made of multiple complex rules whose implementation is often arguable. In this paper, we argue that another view on banking regulation could be adopted. Instead of the current multiplication of complex rules, we suggest implementing strong equity requirements. Such requirements would be simpler and more transparent and efficient than current rules.

Strong equity requirements are however opposed by the banking industry, arguing that they would impede banks' activities too much. Resorting to various ML techniques (RF, ANN, SVR and Lasso), we show that the equity ratio (equity over total assets) has, on the contrary, a positive impact on banks' performance when measured as the ROAA. Far from impeding banks' performance, equity requirements could instead foster it. However, when the equity ratio is greater than 8%, it displays a negative impact on the ROAE. In other words, increasing equity requirements above 8% of total assets would have a negative impact on the shareholder value. In sum, the cost associated with equity requirements is not a social cost (a reduction in banks' performance), but a private cost mostly supported by shareholders. Given the current level of the regulatory equity ratio (i.e., 3%), there is therefore no economic reason to oppose to stronger equity requirements. This is the economic message of this paper.

From a methodological perspective, we offer evidence that ML techniques can successfully be applied to banks' balance sheet data. Starting with a great number of explanatory variables (i.e., 36), we manage to disentangle which prove the most significant predictors of the dependent variables thanks to these techniques. In addition, thanks to several interpretation methods (PDP, ALE, Shapley value), we manage to precisely specify the complex and non-linear relationships between some variables to obtain results that are of particular interest for banking regulation. This paper therefore contributes to the literature on interpretable ML by showing how ML techniques can be used to complement the conclusions drawn from standard econometric models. Concerning the specific research question tackled by the paper, RF models allow to exhibit the shape of the relationship between the equity ratio and two performance measures, which makes it possible to finely answer the question of capital regulation.

One weakness of our study is that it does not fully allow to conclude that a world where a unique and simple leverage ratio would be implemented would be safer and more efficient than the actual world where complex rules are in place. Achieving such a result requires to build a counterfactual world where banks are indeed only subject to a unique and simple leverage ratio and compare how things go in this world with what happens in the real world. Building such counterfactual is a challenge for future works. However, our results still allow to conclude that *ceteris paribus* a strong increase in equity requirements would certainly improve the efficiency of banking regulation without impeding banks' performance, which remains an essential result as far as banking regulation is concerned.

## References

- V. V. Acharya, D. Gale, and T. Yorulmazer. [Rollover risk and market freezes](#). *Journal of Finance*, 66(4):1177–1209, 2011. 7
- 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. 4, 5, 6, 7
- E. I. Altman. [Financial ratios, discriminant 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
- M. Aria, C. Cuccurullo, and A. Gnasso. [A comparison among interpretative proposals for Random Forests](#). *Machine Learning with Applications*, 6:100094, Dec. 2021. ISSN 26668270. doi: 10.1016/j.mlwa.2021.100094. URL <https://linkinghub.elsevier.com/retrieve/pii/S2666827021000475>. 4, 22
- M. Asmild and M. Zhu. Controlling for the use of extreme weights in bank efficiency assessments during the financial crisis. *European Journal of Operational Research*, 251(3):999–1015, June 2016. ISSN 03772217. doi: 10.1016/j.ejor.2015.12.021. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221715011455>. 3
- O. Badunenko, S. C. Kumbhakar, and A. Lozano-Vivas. Achieving a sustainable cost-efficient business model in banking: The case of European commercial banks. *European Journal of Operational Research*, page S0377221720310857, Dec. 2020. ISSN 03772217. doi: 10.1016/j.ejor.2020.12.039. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221720310857>. 4
- A. Baker. [The new political economy of the macroprudential ideational shift](#). *New Political Economy*, 18(1):112–139, 2013. 5
- M. Baker and J. Wurgler. [Do Strict Capital Requirements Raise the Cost of Capital? Bank Regulation, Capital Structure, and the Low-Risk Anomaly](#). *The American Economic Review*, 105(5):315–320, 2015. 4

- A. Berger. [The Relationship between Capital and Earnings in Banking](#). *Journal of Money, Credit and Banking*, 27(2):432–56, 1995. 4, 9
- A. Berger, R. J. Herring, and G. Szego. [The Role of Capital in Financial Institutions](#). *Center for Financial Institutions Working Papers*, 1995. 9
- A. N. Berger and C. Bouwman. [Bank Liquidity Creation](#). *Review of Financial Studies*, 22(9):3779–3837, 2009. 4
- F. Boissay, C. Cantú, S. Claessens, and A. Villegas. Impact of Financial Regulations: Insights from an Online Repository of Studies. *BIS Quarterly Review*, Mar. 2019. 10
- 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>. 5
- P. Bolton, S. Cecchetti, J.-P. Danthine, and X. Vives. [Sound At Last? Assessing a Decade of Financial Regulation](#). CEPR Press, 2019. 6
- P. Bolton, M. Després, L. A. Pereira da Silva, F. Samama, and R. Svartzman. [The green swan – Central banking and financial stability in the age of climate change](#). *Bank for International Settlements and Bank of France Working Paper*, 2020. 6
- E. Bordeleau and C. Graham. [The Impact of Liquidity on Bank Profitability](#). *Bank of Canada*, (10-38), 2010. 4
- 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, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. 8
- L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. 1984. 8, 9
- M. Brunnermeier. [Deciphering the liquidity and credit crunch 2007-2008](#). *Journal of Economic Perspectives*, 23(1):77–100, 2009. 1
- 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, 2016. 7
- J. M. Corcuera, J. De Spiegeleer, J. Fajardo, H. Jönsson, W. Schoutens, and A. Valdivia. [Close form pricing formulas for coupon cancellable cocos](#). *Journal of Banking & Finance*, 42(1):339–351, 2014. 5
- I. Distinguin, C. Roulet, and A. Tarazi. [Bank Regulatory Capital and Liquidity: Evidence from U.S. and European Publicly Traded Banks](#). *Journal of Banking & Finance*, 37, 03 2013. 9

- E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, page S0377221721005695, June 2021. ISSN 03772217. doi: 10.1016/j.ejor.2021.06.053. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221721005695>. 22
- M. D. Fethi and F. Pasiouras. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, 204(2): 189–198, July 2010. ISSN 03772217. doi: 10.1016/j.ejor.2009.08.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221709005438>. 4
- J. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 11 2000. doi: 10.1214/aos/1013203451. 9
- L. Gambacorta and H. S. Shin. Why bank capital matters for monetary policy? *Journal of Financial Intermediation*, 35(B):17–29, July 2018. 4
- J. Goddard, H. Liu, P. Molyneux, and J. O. Wilson. Do Bank Profits Converge? *European Financial Management*, 19(2):345–365, 2013. 4
- C. Goodhart. Problems of monetary management: The U.K. experience. *Papers in monetary economics Reserve Bank of Australia*, 1975. 6
- T. Götze, M. Gürtler, and E. Witowski. Improving CAT bond pricing models via machine learning. *Journal of Asset Management*, 21(5):428–446, Sept. 2020. ISSN 1470-8272, 1479-179X. doi: 10.1057/s41260-020-00167-0. URL <https://link.springer.com/10.1057/s41260-020-00167-0>. 22
- 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
- C.-H. Ho and C.-J. Lin. Large-scale linear support vector regression. *J. Mach. Learn. Res.*, 13(1): 3323–3348, Nov. 2012. ISSN 1532-4435. 9
- G. Iannotta, G. Nocera, and A. Sironi. Ownership structure, risk and performance in the European banking industry. *Journal of Banking & Finance*, 31(7):2127 – 2149, 2007. View Online. 4
- O. Jordà, B. Richter, M. Schularick, and A. M. Taylor. Bank Capital Redux: Solvency, Liquidity, and Crisis. *The Review of Economic Studies*, 88(1):260–286, Feb. 2021. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdaa040. URL <https://academic.oup.com/restud/article/88/1/260/5889963>. 3
- C. Krauss, X. A. Do, and N. Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2):689–702, June 2017. ISSN 03772217. doi: 10.1016/j.ejor.2016.10.031. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221716308657>. 22
- G. Le Quang. Mind the conversion risk: a theoretical assessment of contingent convertible bonds. *EconomiX Working Paper*, 2019. 6

- C.-C. Lee and M.-F. Hsieh. [The impact of bank capital on profitability and risk in Asian banking.](#) *Journal of International Money and Finance*, 32:251 – 281, 2013. 4
- S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas. [Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research.](#) *European Journal of Operational Research*, 247(1):124–136, Nov. 2015. ISSN 03772217. doi: 10.1016/j.ejor.2015.05.030. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221715004208>. 22
- 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>. 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. 3, 4
- W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, Dec. 1943. ISSN 0007-4985, 1522-9602. doi: 10.1007/BF02478259. URL <http://link.springer.com/10.1007/BF02478259>. 8
- 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. 4, 8, 9, 22
- P. Molyneux and J. Thornton. [Determinants of European bank profitability: A note.](#) *Journal of Banking & Finance*, 16(6):1173–1178, December 1992. 4
- S. Morris and H. S. Shin. [Illiquidity component of credit risk – the 2015 Lawrence R. Klein Lecture.](#) *International Economic Review*, 57(4):1135–1148, 2016. 7
- M. Osborne, A.-M. Fuertes, and A. Milne. Capital and profitability in banking : Evidence from us banks. 2012. Working Paper. 9
- A. Persaud. [Why bail-in securities are fool’s gold.](#) *Peterson Institute for International Economics Policy Briefs*, PB14-23, 2014. 5
- A. Petropoulos, V. Siakoulis, E. Stavroulakis, and N. E. Vlachogiannakis. Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3):1092–1113, July 2020. ISSN 01692070. doi: 10.1016/j.ijforecast.2019.11.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169207019302663>. 1, 2, 3, 4
- 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

- K. Tanaka, T. Kinkyō, and S. Hamori. Random forests-based early warning system for bank failures. *Economics Letters*, 148:118–121, Nov. 2016. ISSN 01651765. doi: 10.1016/j.econlet.2016.09.024. URL <https://linkinghub.elsevier.com/retrieve/pii/S0165176516303834>. 2, 3, 4
- R. Tibshirani. [Regression Shrinkage and Selection via the Lasso](#). *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246. 8
- V. T. Tran, C.-T. Lin, and H. Nguyen. [Liquidity creation, regulatory capital, and bank profitability](#). *International Review of Financial Analysis*, 48(C):98–109, 2016. 9
- 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>. 3
- T. Xu, K. Hu, and S. Das, Udaibir. [Bank Profitability and Financial Stability](#). *IMF Working Paper*, 2019. 9

## Appendix A. Data sources and definitions

Table A.4: Data sources and definitions

Data	Definition	Source
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
RESERVE NPL GROSS LOANS	Ratio of volume of NPL to gross loans. It gives a measure of credit risks took by a bank.	FitchConnect
LOAN CUSTOMER DE-POSITS	Loan to customer deposit accounts, which can be withdrawn on demand or short notice.	FitchConnect
LOAN LOSS PROVISION	Provision made by a bank to hedge against loan losses.	FitchConnect
LOAN LOSS PROVISION GROSS LOAN AVG	Ratio of loan loss provision to gross loans.	FitchConnect
DEPOSITS MM FUNDING GROWTH	Growth rate of deposits to money market funding.	FitchConnect
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 LIABILITIES	Liabilities of each bank.	FitchConnect
INT EXP AVG INT LIAB	Ratio of total interest expense / average interest-bearing liabilities.	FitchConnect
AVG INT BEARING LIAB	Average interest-bearing liabilities	FitchConnect
LIQUIDITY ASSETS GS	Liquid assets detained by the bank	FitchConnect
NET LOANS TOTAL EQ-UIITY	Ratio of net loans to total equity.	FitchConnect
NET LOANS TOTAL AS-SETS	Ratio of net loans to total assets.	FitchConnect
GROSS LOANS GROWTH	Growth rate of gross loans.	FitchConnect
TOTAL ASSETS	Total assets of the bank. Often used as a size proxy.	FitchConnect
CUSTOMER DEPOSITS TOTAL FUND DER	Ratio of customer deposits to total fund.	FitchConnect
DEPOSITS ASSETS	Money placed into banking institutions for safekeeping.	FitchConnect

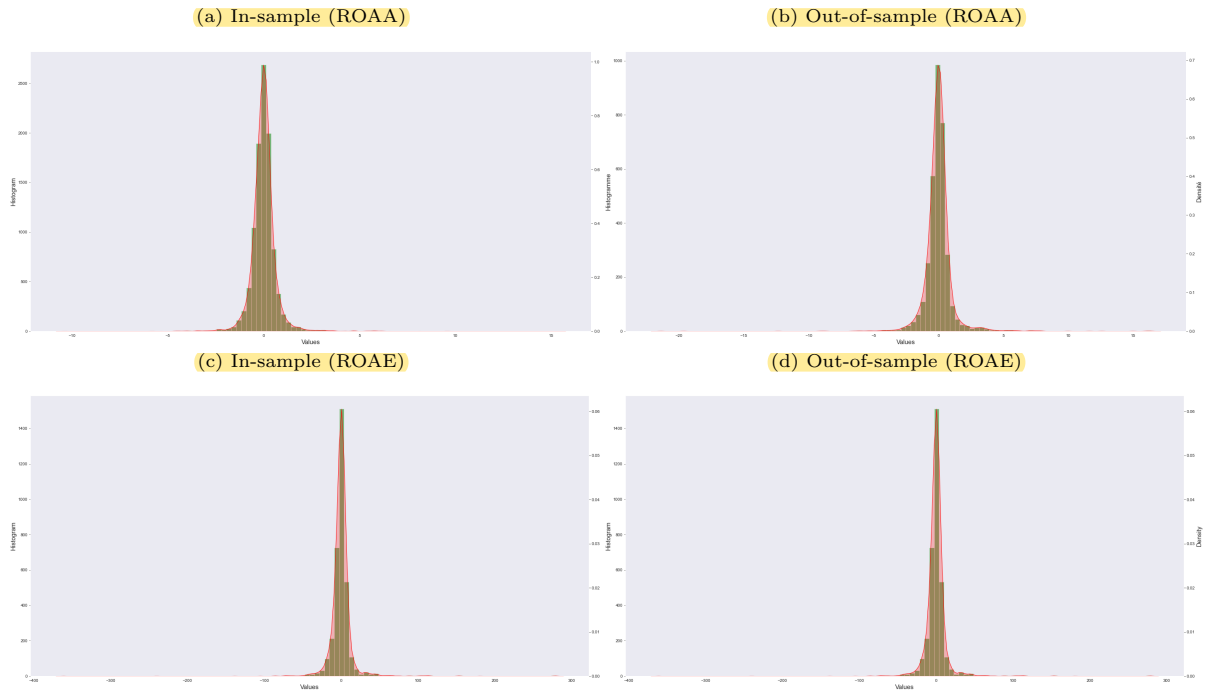
Table A.4: (continued)

BUS 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	FitchConnect
TOTAL OPER EXPENSE	Operating costs include administration costs such as staff costs.	FitchConnect
TOTAL NON INT EXP	Operating expense that is classified separately from interest expense and provision for credit losses.	FitchConnect
OTHER OPER EXP	Operating expenses.	FitchConnect
TOTAL FUNDING	Total Deposits, Money Market and Short-term Funding + Total Long Term Funding + Derivatives + Trading Liabilities	FitchConnect
TOTAL INT EXP	Interests on expenses costs.	FitchConnect
TOTAL DEPOSITS	Total deposits.	FitchConnect
TAX EXPENSE	Expense for current and deferred tax for the period.	FitchConnect
OTHER INT EXP	Interest expenses.	FitchConnect
OTHER NON INT BEARING GS	Non interest-bearing.	FitchConnect
TOTAL NON EARNING ASSETS	All assets that do not generate income.	FitchConnect
NON EARNING ASSETS	Assets that do not generate income.	FitchConnect
TOTAL EARNING ASSETS	All assets that generate income.	FitchConnect
OTHER NON EARNING ASSETS	Other assets that do not generate income.	FitchConnect
OTHER ASSETS	Other assets	FithConnect
HHI	Herfindahl-Hirschman Index. Gives a measure of the market concentration.	FitchConnect
CashAndDepositsBank	Cash and deposits from other banks.	FitchConnect
Inflation	Annual inflation rate.	OECD
GrGDPperCap	Annual GDP growth rate per capita.	World Bank
Corp Tax	Corporate tax rate.	OECD



## Appendix B. Test of the residuals (RF)

Figure B.15: Residuals distribution



Source: Authors' calculations from FitchConnect data.

Table B.5: Statistics on residuals

(a) ROAA as the dependent variable

	Models				
	Lasso (AIC)	Lasso (BIC)	RF	ANN	Linear SVR
In-sample extremes	10.48	10.46	15.09	28.85	28.19
Out-of-sample extremes	10.84	10.76	10.24	28.59	28.84
In-sample mean	0.000	0.000	-0.006	0.001	0.071
Out-of-sample mean	-0.004	-0.005	-0.018	-0.26	0.043
In-sample std	1.48	1.489	0.688	0.897	0.900
Out-of-sample std	1.49	1.496	1.26	0.893	0.898
In-sample Jarque-Bera	0.0	0.0	0.0	0.0	0.0
Out-of-sample Jarque-Bera	0.0	0.0	0.0	0.0	0.0

(b) ROAE as the dependent variable

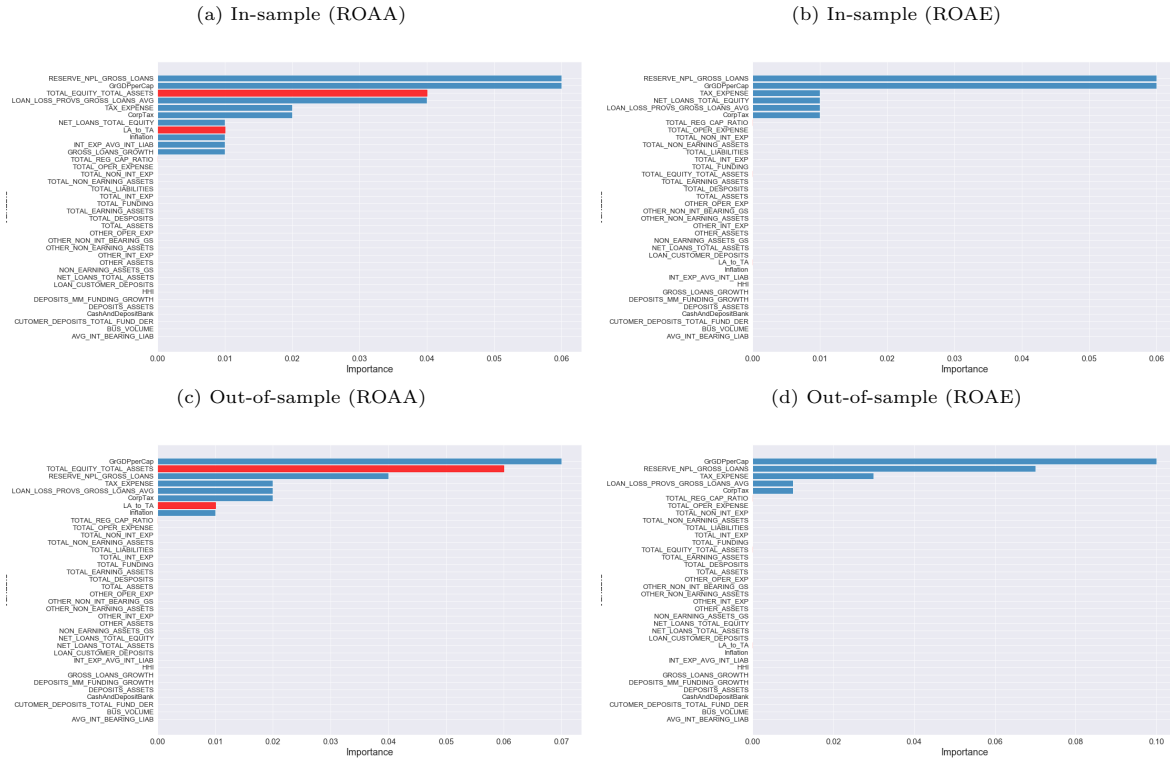
	Models				
	Lasso (AIC)	Lasso (BIC)	RF	ANN	Linear SVR
In-sample extremes	6.74	6.84	10.42	27.78	27.14
Out-of-sample extremes	11.73	11.44	8.23	28.07	27.74
In-sample mean	0.000	0.000	-0.099	0.002	0.055
Out-of-sample mean	-0.284	-0.302	-0.322	-0.029	0.023
In-sample std	22.13	22.19	9.679	0.869	0.873
Out-of-sample std	16.91	16.94	16.46	0.860	0.864
In-sample Jarque-Bera	0.0	0.0	0.0	0.0	0.0
Out-of-sample Jarque-Bera	0.0	0.0	0.0	0.0	0.0

Source: Authors' calculations. "In-sample extremes" and "Out-of-sample extremes" refer to the proportion of observations for which the residual is higher than one standard deviation. The p-value of the Jarque-Bera test is given for "In-sample Jarque-Bera" and "Out-of-sample Jarque-Bera".

## Appendix C. Results for Lasso, ANN and linear SVR

### Appendix C.1. Lasso

Figure C.16: Predictive power of explanatory variables (Lasso)



Source: Authors' calculations from FitchConnect data.

Table C.6: Lasso regressions (ROAA as the dependent variable)

Variable	Models	
	Lasso (AIC)	Lasso (BIC)
$\frac{E}{TA}$	0.0336	0.033
TCR	0.00107	0.0007
$\frac{LA}{TA}$	-0.0044	-0.0041

Source: Authors' calculations. ROAA is the dependent variable. Variables selection is done using both the Akaike information criterion (AIC) and the Bayes information criterion (BIC). A coefficient equal to 0 means that the variable has been excluded.

Table C.7: Lasso regressions (ROAE as the dependent variable)

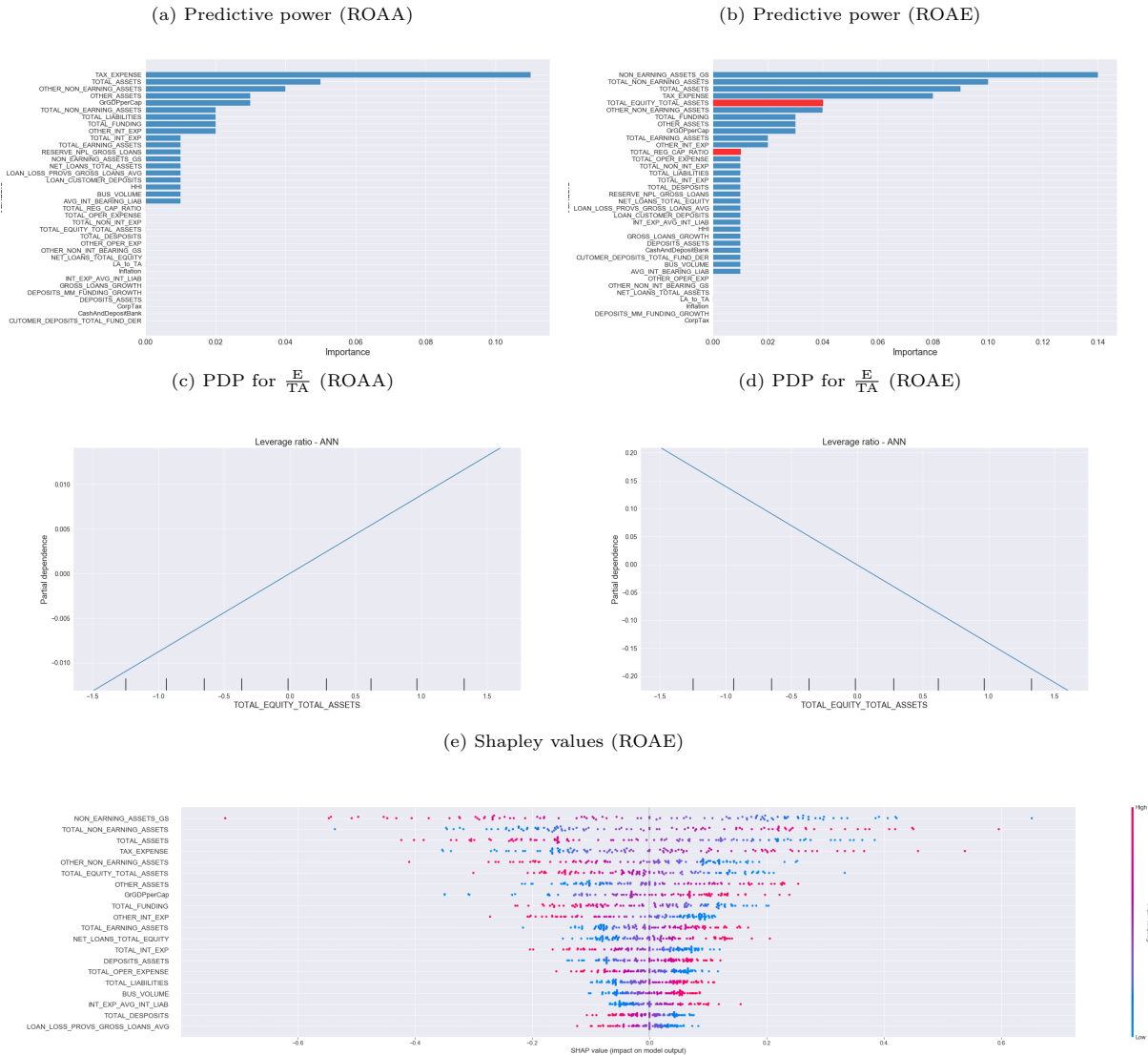
Coefficient	Models	
	Lasso (AIC)	Lasso (BIC)
$\frac{E}{TA}$	-0.0997	-0.0444
TCR	-0.004	0
$\frac{LA}{TA}$	-0.024	-0.0225

Source: Authors' calculations. ROAE is the dependent variable. Variables selection is done using both the Akaike information criterion (AIC) and the Bayes information criterion (BIC). A coefficient equal to 0 means that the variable has been excluded.

Lasso is the model that performs the least both in- and out-of-sample (see section 6.1). The results presented here must therefore be considered with caution. Having a look at the predictive power of explanatory variables (Figure C.16), we notice that  $\frac{E}{TA}$  is identified as a strong determinant of ROAA.  $\frac{LA}{TA}$  is a weaker determinant. The predictive power of TCR on ROAA is however limited. When ROAE is the dependent variable, none of those variables are identified as having a strong impact. Turning to the results of Lasso regression (Tables C.6 and C.7), we notice that our main result holds: the impact of  $\frac{E}{TA}$  on ROAA is indeed positive, while the impact of  $\frac{E}{TA}$  on ROAE is negative.

Appendix C.2. ANN

Figure C.17: Results of ANN regressions



Source: Authors' calculations from FitchConnect data.

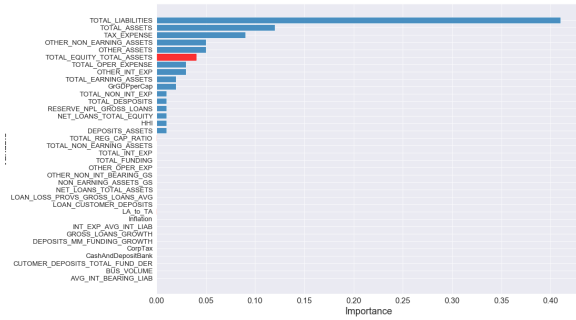
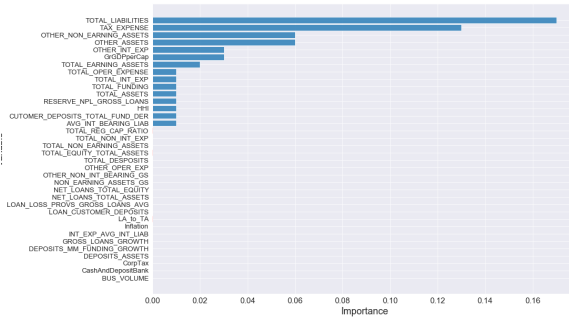
When ROAA is the dependent variable, ANN performs far less than RF. On the contrary, when ROAE is the dependent variable, ANN out-performs RF out-of-sample (see section 6.1). We therefore focus on the situation where ROAE is the dependent variable. In this case, ANN regressions confirm that there exists a significant and negative impact (graphs (b) and (d)) of  $\frac{E}{TA}$  on ROAE. The Shapley value (graph (e)) confirms the results drawn from RF regressions: the impact of  $\frac{E}{TA}$  on ROAE is expected to be especially negative for largest values of  $\frac{E}{TA}$ , which corroborates the idea that strong equity requirements would be detrimental to the shareholder value.

Appendix C.3. Linear SVR

Figure C.18: Results of linear SVR regressions

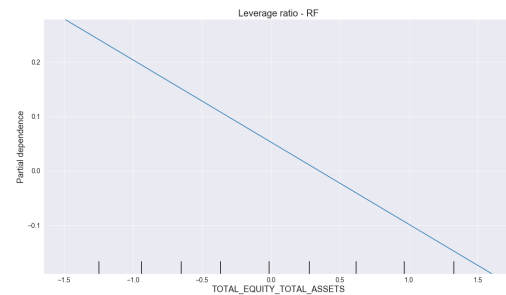
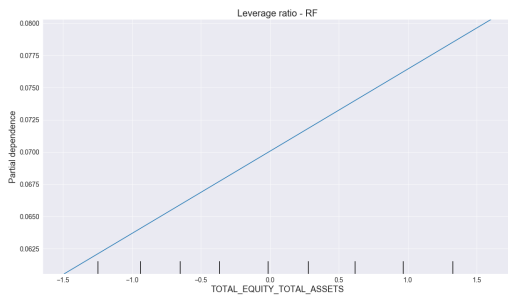
(a) Predictive power (ROAA)

(b) Predictive power (ROAE)



(c) PDP for  $\frac{E}{TA}$  (ROAA)

(d) PDP for  $\frac{E}{TA}$  (ROAE)



(e) Shapley values (ROAE)



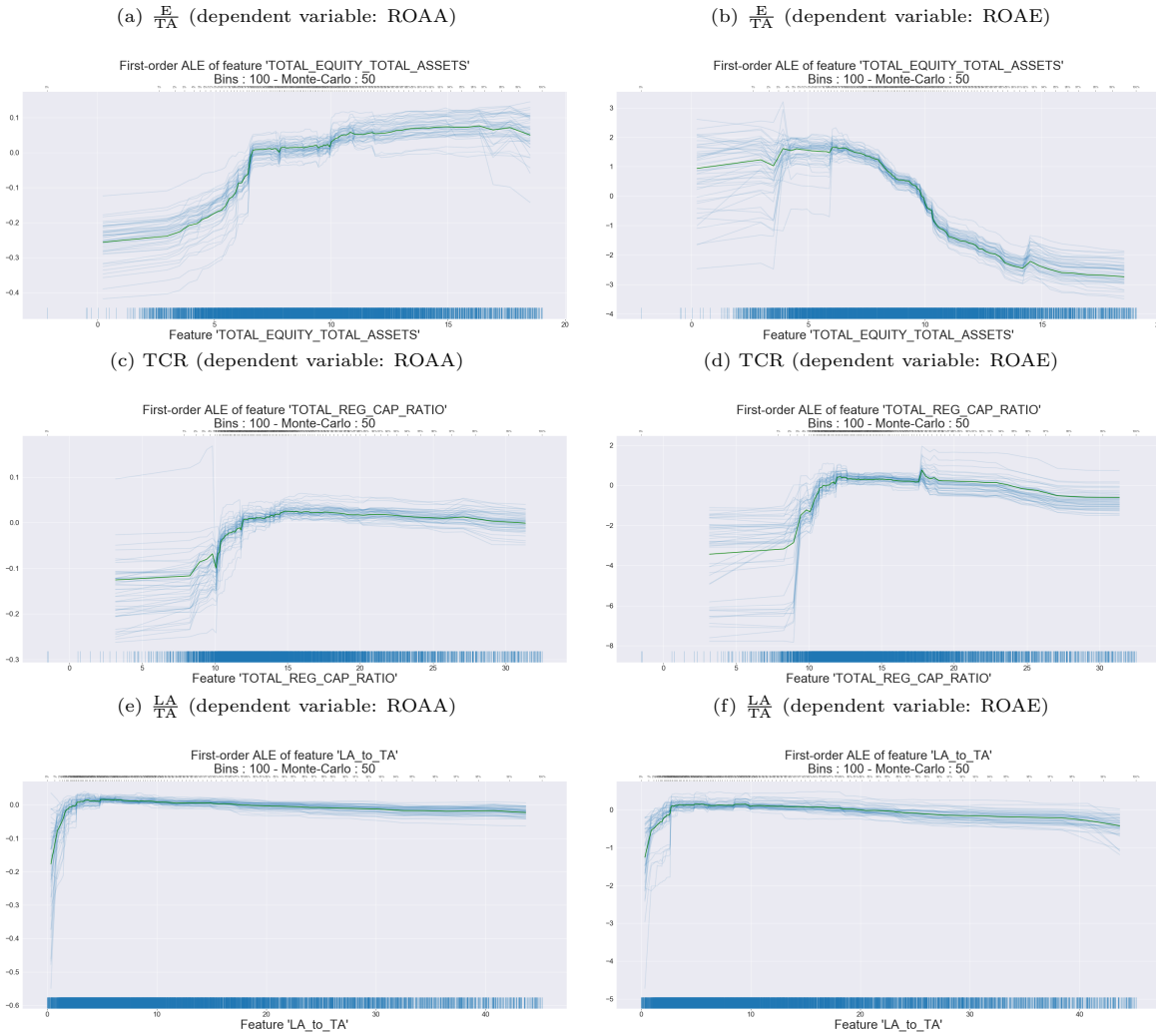
Source: Authors' calculations from FitchConnect data.

As ANN regressions, linear SVR regressions perform less than RF when ROAA is the dependent variable, but outperform RF regressions out-of-sample when ROAE is the dependent variable (see section 6.1). Here again, we thus focus on this last case. SVR confirms the significant (plot (b)) and negative (plot (d)) impact of  $\frac{E}{TA}$  on ROAE. The Shapley value helps once again specifying how  $\frac{E}{TA}$  is expected to determine ROAE: consistently with what has been found thanks to RF and ANN regressions, the Shapley value (plot (e)) implies that  $\frac{E}{TA}$  negatively impacts ROAE especially when the value taken by  $\frac{E}{TA}$  is high.

## Appendix D. Robustness outputs

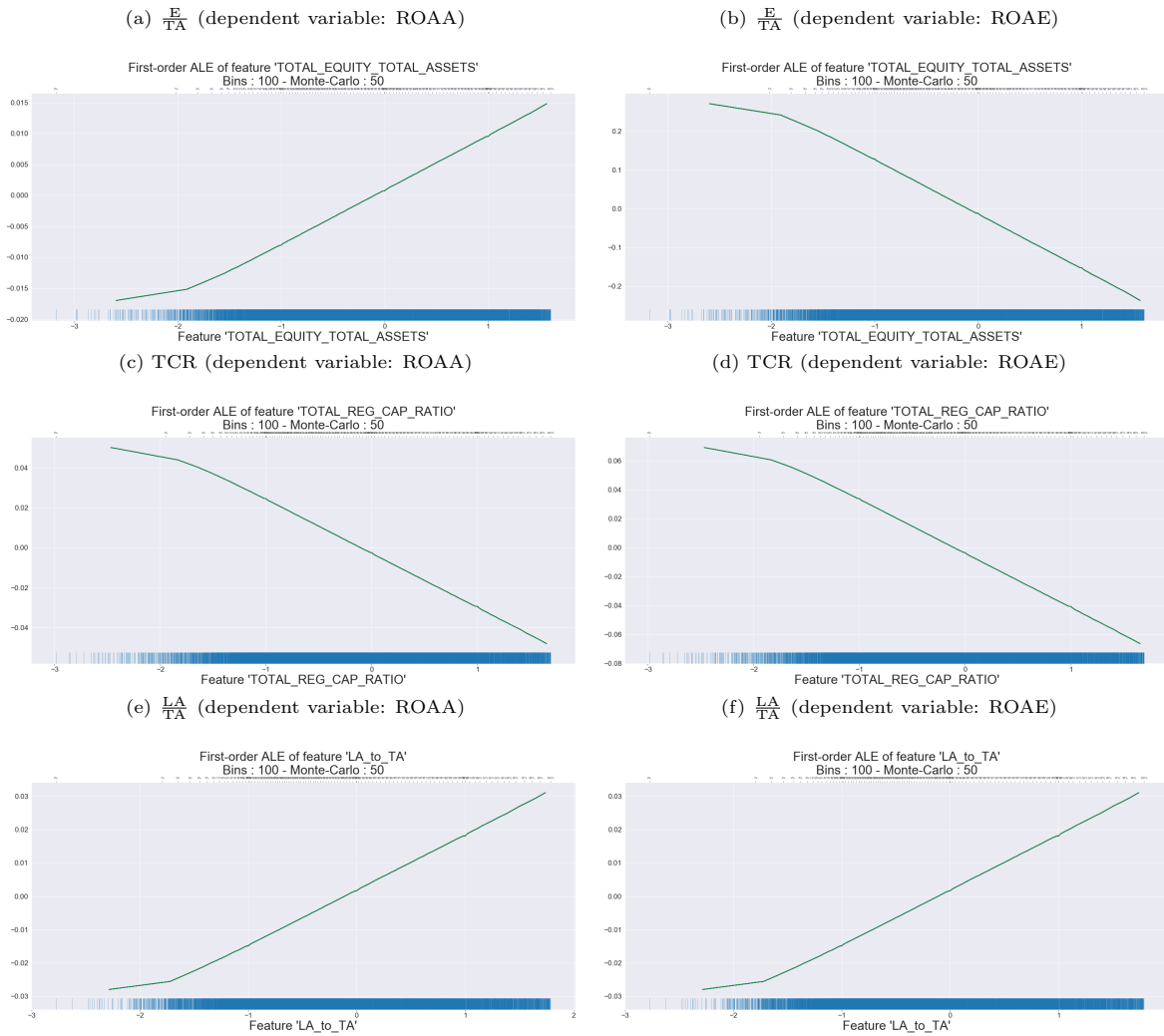
### Appendix D.1. ALEs

Figure D.19: ALEs (RF)



Source: Authors' calculations from FitchConnect data. The distributions of the variables are reported on the x-axes (centiles).

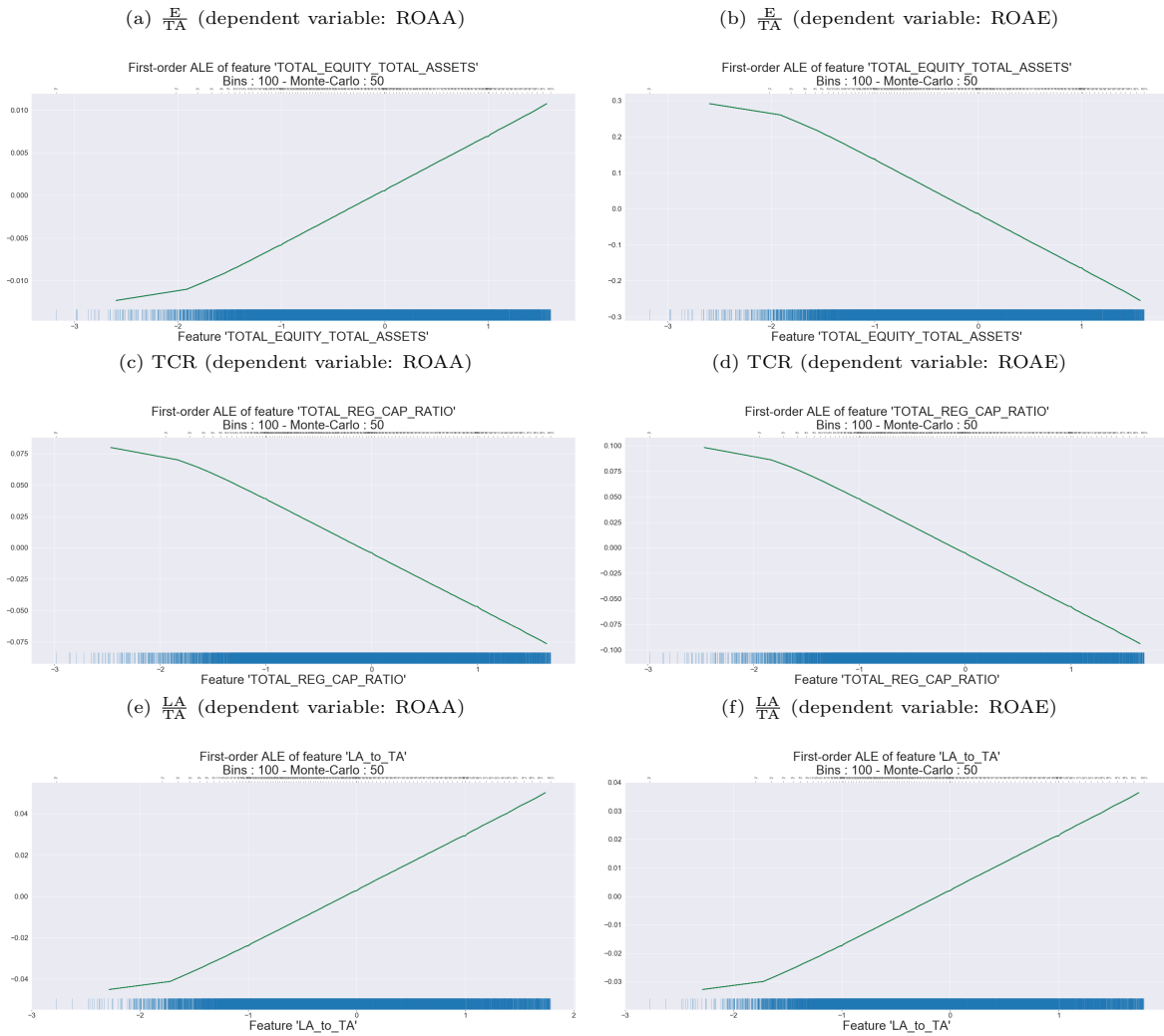
Figure D.20: ALEs (ANN)



Source: Authors' calculations from FitchConnect data. The distributions of the variables are reported on the x-axes (centiles).



Figure D.21: ALEs (Linear SVR)

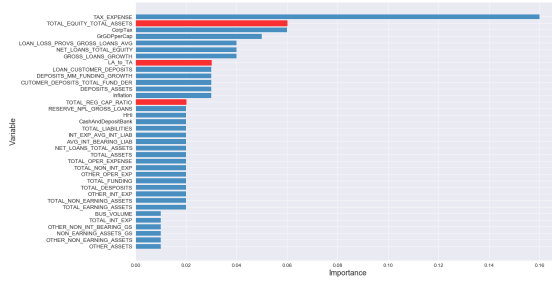


Source: Authors' calculations from FitchConnect data. The distributions of the variables are reported on the x-axes (centiles).

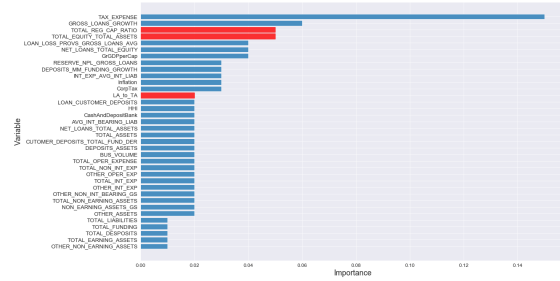
Appendix D.2. Classification

Figure D.22: Classification (RF)

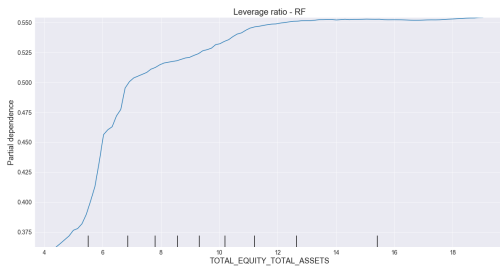
(a) Predictive power (ROAA)



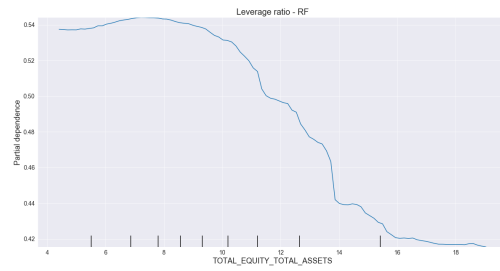
(b) Predictive power (ROAE)



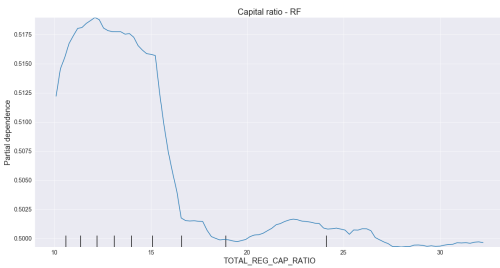
(c) PDP for  $\frac{E}{TA}$  (ROAA)



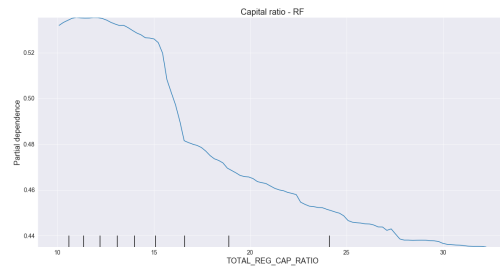
(d) PDP for  $\frac{E}{TA}$  (ROAE)



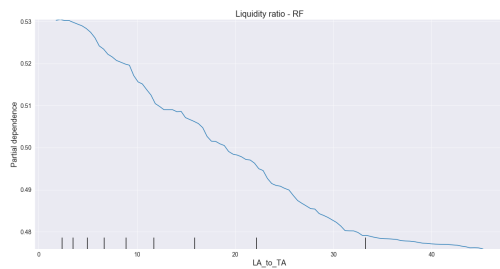
(e) PDP for TCR (ROAA)



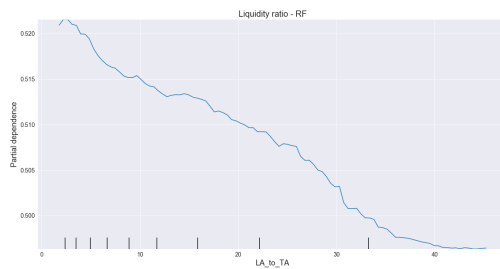
(f) PDP for TCR (ROAE)



(g) PDP for  $\frac{LA}{TA}$  (ROAA)



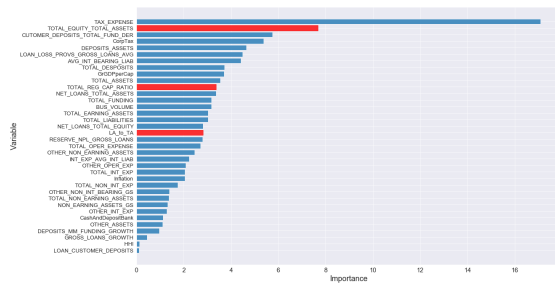
(h) PDP for  $\frac{LA}{TA}$  (ROAE)



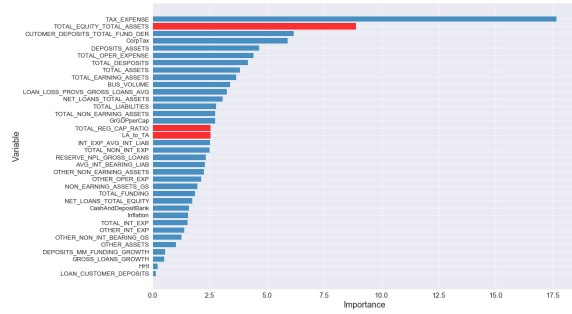
Source: Authors' calculations from FitchConnect data. In PDPs, the distributions of the variables are reported on the x-axes (deciles).

Figure D.23: Classification (ANN)

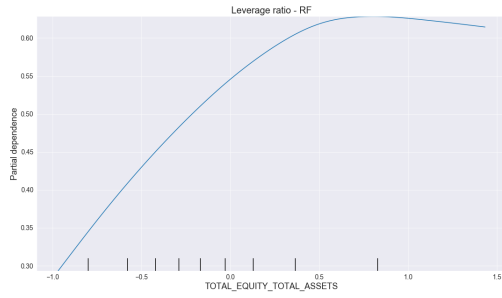
(a) Predictive power (ROAA)



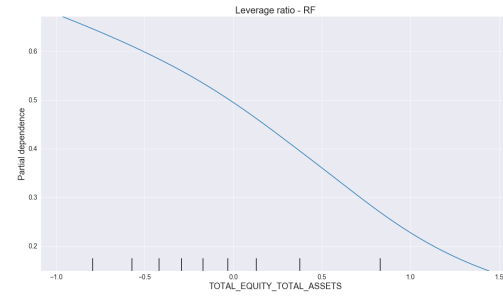
(b) Predictive power (ROAE)



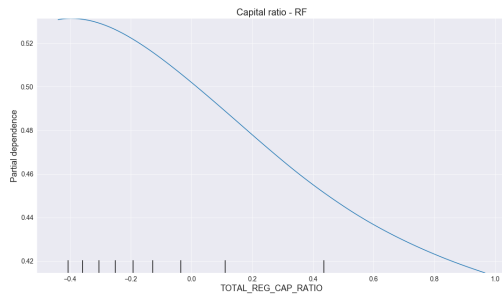
(c) PDP for  $\frac{E}{TA}$  (ROAA)



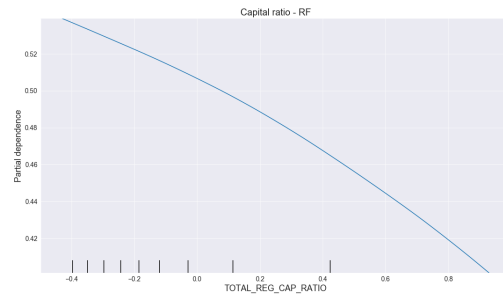
(e) PDP for TCR (ROAA)



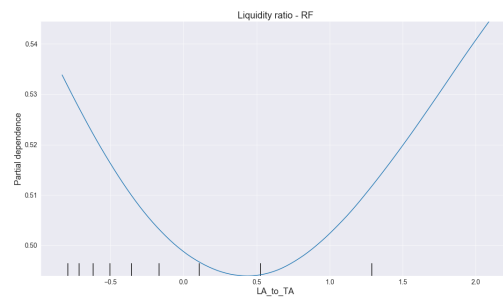
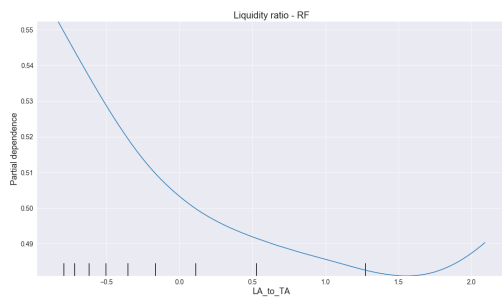
(f) PDP for TCR (ROAE)



(g) PDP for  $\frac{L}{TA}$  (ROAA)



(h) PDP for  $\frac{L}{TA}$  (ROAE)



Source: Authors' calculations from FitchConnect data. In PDPs, the distributions of the variables are reported on the x-axes (deciles).