

---

## Determinants of banks' profitability: Do Basel III liquidity and capital ratios matter?

---

Document de Travail  
Working Paper  
2019-24

Pierre Durand



UMR 7235

EconomiX - UMR7235  
Université Paris Nanterre  
Bâtiment G - Maurice Allais, 200, Avenue de la République  
92001 Nanterre cedex

Email : [secretariat@economix.fr](mailto:secretariat@economix.fr)



# Determinants of banks' profitability: Do Basel III liquidity and capital ratios matter?

Pierre Durand\*

December 3, 2019

## Abstract

In this paper, we investigate the role played by the Total Capital Ratio (TCR) and Liquidity Coverage Ratio (LCR) among determinants of banks' profitability. To this end, using Random Forest regressions and a large dataset of banks' balance sheet variables, we assess the impact and predicting power of Basel III capital and liquidity ratios. Our results confirm the trade-off theory of the capital structure: banks have an optimal capital ratio below which the relation between capital and profitability is positive. On average, this optimum falls between 15% and 20%. Furthermore, we show that LCR has a positive, but weak, effect on profitability. Overall, our findings illustrate the fact that regulatory ratios do not constitute binding conditions for banks' performance.

**Keywords:** Basel III, Capital ratio, Liquidity ratio, Banks' profitability, Random Forest regressions.

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

**Acknowledgements:** I would like to thank all my colleagues at the EconomiX research centre, including Benjamin Egron, Chouaib Jouf, Florian Morvillier and Charles Teuf. I would particularly thank Valérie Mignon for her guidance and advice. I also would like to thank Eloïse for her support. Finally, I would like to address a particular thought to my father, and all my family, without whom none of this would have been possible.

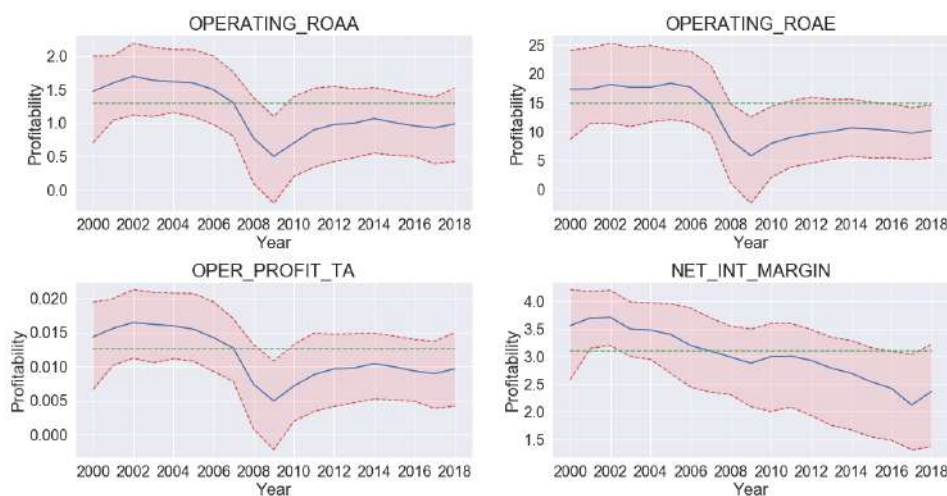
\*EconomiX – CNRS, University of Paris Nanterre

Mail: pierre(.)alexis(.)durand(@)hotmail(.)fr

# 1 Introduction

Banks play a fundamental role in the economy through their role of liquidity creation and risk transformation. In a context of hardening competition with alternative and unintermediated financing channels (the so-called *shadow banking*) and tightening prudential regulation, a crucial interrogation arises: is banks' profitability at risk? This question finds several interests. First, the idea according to which a threat hangs over banks' profitability, consists of an argument for the banking industry when defending its interests in front of regulators. Second, some evidence shows a positive link between banks' profitability and their internal financial stability (Keeley, 1990; Xu et al., 2019).<sup>1</sup> Therefore, if Basel III effectively acts as a constraint on profitability, it could also threaten the very macroeconomic financial stability it seeks to achieve.

Figure 1 – Evolution of profitability variables



Source: Author's calculations from FitchConnect data (550 to 1150 banks per year in 21 developed countries). The figure reports the dynamics of four profitability variables: Return On Average Assets (ROAA), Return On Average Equity (ROAE), Operating Profit to Total Asset (OPTA) and Net Interest Margin (NIM). The blue line represents the median of each variable. The red lines indicate the first and last quartiles. The green line shows the 2007 level of each median.

As shown in Figure 1, profitability seems to have recovered a positive dynamic since the financial crisis. However, performances' median levels remain lower than before 2007. Moreover, profitability as measured by Return On Average Assets (ROAA), Return On Average Equity (ROAE), and Operating Profit to Total Assets (OPTA) displays a more flat than positive dynamic since 2012. At a first sight, it might correspond to a slowdown due to the implementation of Basel III agreements. It is equally likely that this phenomenon is imputable to the weakening of the financial system, the European sovereign debt crisis, or the Zero Lower Bound context. As for the Net Interest Margin (NIM), it decreases since 2000. Thus, the recent regulatory accords cannot be responsible for this overall pattern.

<sup>1</sup>See also the speech by L. de Guindo, Vice-President of the ECB, on May, 1st 2019: "Clearly, bank profitability matters for financial stability". It shows how much an institution as the ECB values banks' profitability.

Indeed, Basel III agreements were signed in 2010. They are transposed and progressively implemented since 2012, and will continue to be applied until the early 2020's with the finalisation process ("Basel IV"). Therefore, knowing if prudential regulation prevents banks from being profitable constitutes a relevant and primordial question to be answered. From now, it should be noted that when we discuss regulatory requirements, we refer to Total Capital Ratio (TCR) and Liquidity Coverage Ratio (LCR), two flagship requirements of Basel III agreements.

Although the existing literature dealing with the role of capital and liquidity requirements in determining profitability provides some evidences on the link between capital and profitability (Berger, 1995; Berger et al., 1995; Gropp and Heider, 2010; Kraus and Litzenberger, 1973; Osborne et al., 2012), it is rather scarce on the relation between liquidity and profit (Berger and Bouwman, 2009; Bordeleau and Graham, 2010; Molyneux and Thornton, 1992), and even more on the connections amid both capital and liquidity, and performance (Distinguin et al., 2013; Tran et al., 2016). Obviously, those two variables (capital and liquidity ratios) are not the only ones to determine banks' profitability. Incidentally, many studies investigate other determinants, among which macroeconomic variables such as the monetary policy (Borio et al., 2017), tax variables (Albertazzi and Gambacorta, 2009; Chia and Whalley, 1999), or the Non-Performing Loans (García-Herrero et al., 2009). However, from what has been observed in the literature, no study seems to establish a clear comparison between regulatory ratios and other determinants.<sup>2</sup>

The objective of this paper is to fill this gap by comparing the importance of those different variables in determining banks' performance. Specifically, we aim to identify among the determinants used in the literature, which one best improves the out-of-sample prediction of the different profitability variables we study. Moreover, we seek to compare regulatory ratios' performance with that of other variables. To this end, a key contribution of our paper is to use random forest (RF) regressions, whose out-of-sample forecasting performance is widely recognized (Hastie et al., 2009). Three reasons motivate this choice of methodology rather than more traditional methods: RFs do not impose the choice of a functional form on the estimated model, they allow us to account for a wide range of explanatory variables (or features), and they incorporate both a procedure for estimating the determinants and their impact on the explained variable (or label). In order to ensure the stability and soundness of our results, we provide several robustness checks: (i) we consider four profitability variables, (ii) all results are compared to the linear model, (iii) all simple RFs' outcomes are compared with the mean of results of 100 RFs ran on random sample selection, and (iv) two samples are used in our empirical strategy, one being smaller than the second but containing Basel III Liquidity Coverage Ratio (LCR).<sup>3</sup> We go further than the existing literature in various ways. We consider a broad set of deter-

---

<sup>2</sup>It should be noted here that some studies, such as Distinguin et al. (2013), include both regulatory ratios and other determinants. However, their approach is not to compare the determining efficiency of those two kinds of variables (regulatory versus non-regulatory).

<sup>3</sup>The NSFR, a fundamental ratio implemented under Basel III, is not incorporated for unavailability reasons.

minants issued from the literature, including regulatory ratios. In particular, we rely on LCR which has not been widely studied due to its recent implementation. From a methodological viewpoint, we consider RF regressions which have not been used so far to investigate bank's profitability. This methodology is particularly relevant since it allows us to detect and model existing non-linearities in the individual effect of each requirement and the interactions between variables. Moreover, its out-of-sample prediction performance is known to be satisfactory compared to more standards methods (Hastie et al., 2009).

Relying on annual data over the 2000-2018 and 2012-2018 periods, our main findings can be summarized as follows. Above all, the Liquidity Coverage Ratio does not appear to be the most critical determinant of banks' performance, which is a quite innovative result in the literature. In all cases, LCR has a positive quasi-linear effect on variables of performance. On the contrary, even if the TCR never takes the first place, it always belongs to the four to five first most important determinants. In most cases, capital ratios exert a non-linear effect on profitability: a positive one for low values and a negative one for values above 15%. This finding confirms the trade-off theory of an existing optimum (Kraus and Litzenberger, 1973). In addition, a non-linear effect is at play when looking at both ratios' impact; such combined impact revealing an accumulation phenomenon. Finally, we show that banks' profitability main determinants can be gathered in four groups: expenses including taxes, provisions for credit risks, assets and equity variables, and loan features.

The remainder of this paper is structured as follows. Section 2 presents the related literature. Data and methodology are discussed in Section 3. Some descriptive statistics are displayed in Section 4. Sections 5 presents the results, and robustness checks are provided in Section 6. We give our conclusions in Section 7.

## 2 Literature review

### 2.1 Definition, measures and accounting relations between regulatory ratios and profitability

We consider four commonly used profitability variables in order to measure banks' performance: 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), OPTA (Xu et al., 2019),<sup>4</sup> and NIM (Albertazzi and Gambacorta, 2009; Xu et al., 2019). Let us now describe these variables.

- ROAA is measured by the ratio of net income to average total assets. It measures the

---

<sup>4</sup>Note that the usual variable here is Operating Profit in level. A size effect being strongly detected in our regressions, we controlled for this bias dividing Operating Profit by Total Assets.

profitability of assets, meaning how a firm uses the resources it owns to generate profit. It refers to the returns on the assets purchased using each unit of money invested. Net Income designates the total revenues (from all the activities of the firm) minus business expenses (costs of sales, administrative expenses, operating expenses,...) and taxes. Assets refer to the resources with economic value owned by the firm and from which a benefit is expected (cash, cash equivalent, account receivable, long term resources, financial assets, intangible assets).

- ROAE is the net income to average shareholders' equity ratio and represents the amount of income created for each monetary unit of shareholders' equity. It measures the return on each unit of money invested. Equity corresponds to the assets minus the liabilities.
- Operating Profits are the profits generated by a firm's core business. This variable measures the business potential profitability and is distinct from net incomes, which represents the revenues of all activities. In order to control for a significant size effect, we divide operating profits by total assets. From an accounting point of view, this procedure means that the new ratio refers to the firms' capacity to use their resources to make the main line of business profitable.
- Net Interest Margin gives the returns on invested funds. It is measured by the difference between the interests received and those paid, divided by the average invested assets. The NIM is supposed to be influenced by macroeconomic variables (interest rates particularly).

As discussed earlier, we consider two regulatory ratios to account for Basel III requirements: total capital ratio and liquidity coverage ratio. The TCR is defined as the bank's capital in relation to its risk-weighted assets (RWA):  $TCR = \frac{\text{Tier 1} + \text{Tier 2}}{\text{RWA}}$ . The ratio's numerator is accountingly close to net incomes since it regroups cash, government securities, interest-earning loans, less loan-loss reserves, and debt. RWA is the sum of a bank's assets weighted by the risk associated with each type of asset. The regulatory prerogatives imply an increase in this ratio following the Basel III implementation. Therefore, each bank should either increase its capital or decrease its RWAs. An increase in the numerator should be associated with a rise in profitability if it comes from a reduction of reserves and debt, or an extension in interests. However, if this increase is based, for instance, on an expansion of cash and government securities, at the expense of higher interest assets, banking performance may then suffer. The antagonism between these two effects creates an ambiguity in the expected impact of the TCR on profitability.

The LCR is defined as the High Quality Liquid Assets (HQLA) in relation to Total Net Cash Flow Amount:  $LCR = \frac{\text{HQLA}}{\text{Total net cash flow amount}}$ . HQLA includes assets with a high potential of quick cash convertibility. It regroups among others, cash and government securities, some loans and deposits, and some securities. Those assets are weighted depending on their ability to be converted in cash. As for the TCR, the liquidity ratio encourages acquisition of traditionally

low interest assets (such as government securities in the context of ZLB). Therefore, one should expect a negative relationship between LCR and profitability. However, this relation could not be true since an increase in HQLA, all things being equal, can also be the sign of a healthy expense of business. In this sense, a rise in the LCR can also refer to an increase in income. Regarding profitability variables and regulatory ratios definitions, no clear accounting relation can be drawn between them. It does not mean that no relationship exists: we will see in the next section which links can appear from an economic perspective.

## 2.2 On the relationship between regulatory ratios and profitability

According to the trade-off theory of the capital structure (Kraus and Litzenberger, 1973), an optimum for the debt to equity ratio exists in describing the best capital structure of a bank. Therefore and from this perspective, if the TCR required by regulators is superior to the optimal capital ratio for a given bank, then an increase in capital for this bank generates costs and jeopardizes its performance. On the contrary, if a bank's optimal capital ratio is lower than the regulatory TCR, no cost is created by an increase in capital. The Modigliani and Miller (1958)'s theorem is more categorical and states that the debt to equity ratio does not affect the firm's average costs. However, many hypotheses compel this theorem, and empirical investigations do not converge towards clear-cut findings: Balling (2015) validates the theorem, while Miller (1995) raises doubts regarding its application to banks, which is confirmed by Goddard et al. (2013).

All put together, the relation linking capital ratio to banking profitability depends on the determinants of the capital ratio optimum, and banks' place in relation to this optimum. Banks' individual characteristics condition the position to the optimal capital ratio. The determination of the optimum itself depends on several elements. Berger (1995) shows that the optimal capital structure corresponds to a situation in which total benefits, including tax expenses, equal costs of debt. Acting on bankruptcy costs, the capital ratio influences the optimum level and, therefore, the relation between capital and profitability. Berger (1995) also finds that the economic cycles can influence the optimum through an impact on bankruptcy costs: during periods of troubles, expected losses increase, as the optimum, making positive the link between capital and profitability, and upside down during prosperous years. This result is later confirmed by Osborne et al. (2012).

As we can see, there is no clear consensus on an immutable (positive or negative) link between capital and profitability. Empirically, the evidences are mitigated: some findings underline positive relations between well-capitalization and profitability (Berger, 1995; Iannotta et al., 2007; Lee and Hsieh, 2013), whereas some others show negative links (Goddard et al., 2013; Baker and Wurgler, 2015).

Regarding the impact of LCR on profitability, the literature is also non-consensual. On the one hand, liquid assets detention decreases maturity mismatch risks and liquidity risks. Thus,

it should reduce default probability and financing costs, and, therefore, support performances (Berger and Bouwman, 2009; Bordeleau and Graham, 2010). Moreover, some evidence shows that the quality of assets supports profitability. It appears when the portion of gross margin provisioned in order to cover poor quality assets is superior to income generated by those assets (DeYoung and Rice, 2004). However, liquid assets generate lower returns and therefore lower revenues. From this point of view, the incentive to increase HQLA tends to weigh on banks' profitability (Goddard et al., 2013; Molyneux and Thornton, 1992). As for the TCR's impact on banks performance, no consensus appears in the literature supporting a positive or negative relation between LCR and profit.

Other determinants of banks' profitability have been highlighted in the literature, such as: market concentration (Bourke, 1989; Dietrich and Wanzenried, 2014; Molyneux and Thornton, 1992), inflation (Bordeleau and Graham, 2010; Bourke, 1989; Dietrich and Wanzenried, 2014; Molyneux and Thornton, 1992), cash and bank deposits (Bourke, 1989; Molyneux and Thornton, 1992), credit risks and loan loss provisions (Dietrich and Wanzenried, 2014; García-Herrero et al., 2009), GDP growth rates (Bordeleau and Graham, 2010; Dietrich and Wanzenried, 2014), etc. This list is far from being exhaustive, and it would be interesting but fastidious to list all of the possible determinants. In this regard, Random Forest regressions appear to be particularly relevant. Indeed, as Random Forest regressions allow for large number of features, we shall take in our empirical approach as many profitability determinants as we can in order to control for any effect that can influence banks' performance on an accounting and balance sheet point of view, as well as on an economic perspective.

## 3 Data and methodology

### 3.1 Data

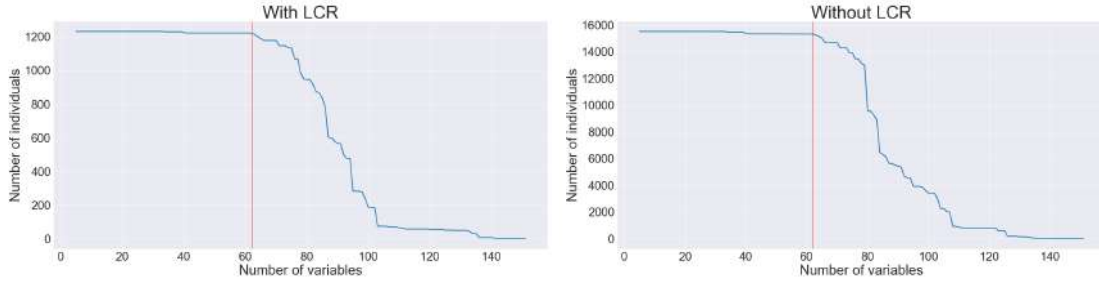
Our two datasets consist of banks' balance sheet variables from the twenty-one following countries: Belgium, Canada, China, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, Poland, Portugal, Singapore, Spain, Sweden, Switzerland, United-Kingdom and United-States. Data are annual and come from FitchConnect.

In order to be able to consider a maximum number of balance sheet variables that could determine profitability, as well as a maximum number of banks, we first constituted a large dataset on which some treatments have to be applied. The first constraint comes from the fact that RF regressions cannot be carried out successfully if the database contains any missing value. Looking into the evolution of the number of available observations depending on the number of variables without any missing value (Figure 2), we kept the following shapes for both databases: 63 variables for 1221 banks in the sample with LCR, and 62 variables for 15310 banks in the sample without LCR. Rearranging our databases removing twin variables (with a very similar



definition and measurement), we kept respectively 46 and 45 variables<sup>5</sup> (among which five identity variables, four profitability labels, and the rest of features, Table 2). Note that as the LCR is specific to Basel III, the dataset that contains it covers the 2012-2018 period only, whereas the other dataset goes from 2000 to 2018.

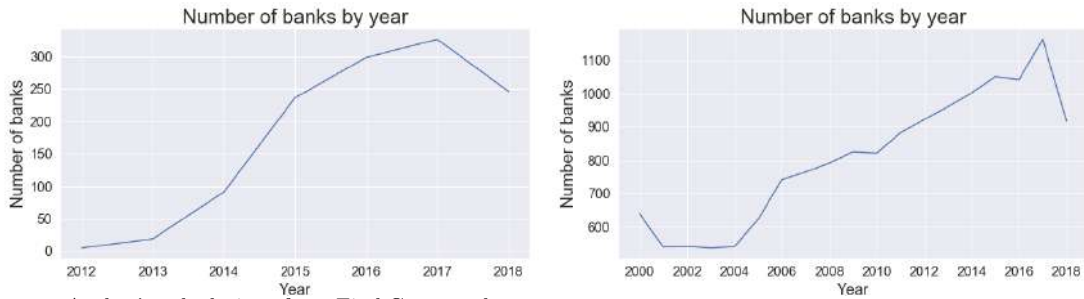
Figure 2 – Datasets’ size by number of variables



Source: Author’s calculations from FitchConnect data.

The critical difference in the number of observations when forcing the dataset to contain the LCR comes from the fact that FitchConnect does not disclose any data for this regulatory ratio regarding US banks. In a robustness procedure (see Section 6) and in order to control for this lack of data, we run our empirical investigation using a proxy variable for LCR. For this third dataset, we have 46 variables for a sample of 4046 banks.

Figure 3 – Number of banks per year



Source: Author’s calculations from FitchConnect data.

Note that we have an unusual data structure. Indeed, our datasets are all balanced in the sense that there are no missing values. They also are unbalanced regarding the number of periods per banks. Figure 3 displays the evolution of the number of banks by year for the two databases we consider. As can be seen, very few LCR data were disclosed between 2012 and 2014. This particular dataset architecture is not an issue regarding RF implementation, and each bank for a given year represents an observation.

<sup>5</sup>We also constrained the dataset to include some accounting and macroeconomic variables identified from the literature.

## 3.2 Random forest regressions

### *Tree-based models and Random Forest regressions*

We use Random Forest regressions in our empirical strategy, due to their high out-of-sample performance (Hastie et al., 2009). Besides, the RF methodology is particularly well designed in identifying determinants from a large set of variables and in assessing their impact on the explained variable. Indeed, this methodology allows us to consider a large sample of explanatory variables. In our case, since the problem of detecting bank profitability's determinants is rather accounting, the possibility of taking into account many balance sheet variables is quite relevant. RF is also capable of capturing non-linearities and interactions between variables.

The Random Forest methodology (Breiman, 2001) is a supervised statistical learning method<sup>6</sup> issued from bootstrap aggregation (or bagging) techniques. Bagging consists of averaging results of multiple repetitions of the same experiment, which is itself characterized by a high variance and a low bias. Boosting methods being more adaptive than bootstrap aggregation approaches, they are generally preferred. However, RF regressions are slightly different from bagging,<sup>7</sup> and their performance is often close to those of boosting (Hastie et al., 2009). Since RF is computationally simpler to implement, it is generally preferred to boosting methods.

The main idea behind the RF method is to average a more or less large number of decision trees. Therefore, in order to understand this technique, we first briefly present the construction of a tree. This method consists of partitioning explanatory variables' space into regions, and then predict an output value through a simple model (like a constant). The  $M$  final regions (or leaves) of the tree  $\{R_m, m \in [1, M]\}$ , are obtained *via* recursive binary partitions. At each split of features space, we choose the variable for which the split gives the best fit of the output variable (or label). Once the tree is built, the label is predicted by a constant value in each final region, that is:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (1)$$

where  $I$  is an indicative function that scores 1 when  $x \in R_m$ ,  $M$  the number of final leaves (final regions), and  $c_m$  the predicted value of the explained variable  $y$ , in region  $m$ . Therefore, this method is a non-parametric estimation of the unknown function  $f$ . This function defines the true model:  $y = f(x) + \epsilon$ , where  $\epsilon$  designates the error term. Based on a criterion minimization of the sum of squares (Hastie et al., 2009), the best  $\hat{c}_m$  is given by the average (*ave*) of  $y_i$  in region  $R_m$ :

$$\hat{c}_m = \text{ave}(y_i | x_i \in R_m) \quad (2)$$

---

<sup>6</sup>The model is trained on a sub-sample of the dataset and then tested on observations that have not been included in the training sample. The supervised aspect of this approach is the fact that we impose to the model the variable it shall predict.

<sup>7</sup>RF method averages responses of trees like bootstrap aggregation does. The difference comes from the fact that trees are built on a randomly selected sample among the training data and features.

To this point, there are still two parameters to define, which are  $j$  and  $s$ , respectively, the splitting variable and point. Those define the two half-planes wrote as:  $R_1(j, s) = \{X|X_j \leq s\}$  and  $R_2 = \{X|X_j > s\}$ . Finding  $j$  and  $s$  is equivalent to solve the problem given by:

$$\min_{(j,s)}[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \quad (3)$$

Solutions for the inner minimizations are given by Equation (2). Replacing  $c_{i,i \in \llbracket 1,2 \rrbracket}$  by  $\hat{c}_m$  in the expression (3), the determination of  $s$  for each feature is deeply simplified. The selection of  $(j, s)$  is made comparing the results of the minimization problem given in (3) for every input variable. Once the two regions are identified, the process is repeated among them. As mentioned, RF, as proposed by [Breiman \(2001\)](#), aggregates the results of several trees that are built using a different bootstrap sample of the data (as bagging methods). This procedure allows considering different split points and various first splitting variables from one tree to another. This ensures results convergence and that the complexity of the relationships involved is captured. In addition, [Breiman \(2001\)](#) suggests to change each tree's structure: the features selected in the process of a tree building are not necessarily the same from a tree to another. This ensures the de-correlation between trees. Therefore, parameters  $(j, s)$  for any given node are not the same between trees.

A second step in building a decision tree is to determine the maximum depth of a tree and the minimum number of observations in every leaves. Indeed shallow trees are likely to have poor prediction performance, and too deep trees might lead to overfitting issues and consequently bad out-of-sample forecasting. Following the same logic, a large number of observations per final region will predict poorly, while too little observations per leaf are also subject to overfitting problems.

Thus the determination of those two parameters (depth and observations per leaf) is crucial and can be done in several ways. In the context of a single tree, [Hastie et al. \(2009\)](#) propose to rely on a cost complexity criterion that should be minimized. This procedure works on the fact that an increase in complexity (measured by the depth of the tree) that leads to overfitting the data and decreases the sum of squares is counterbalanced by an increase in a cost term that depends on the tree's depth. In the context of RF, this approach is quite demanding in terms of calculation: the criterion must be minimized for each tree. Another technique consists in making varying those two parameters in multiple RF regression estimations simultaneously and retain those that maximize the out-of-sample prediction performance.

The last parameter to establish is the number of trees in the forest. There are some debates on the optimal value for this parameter (and the very existence of an optimum). [Hastie et al. \(2009\)](#) suggest that the error of the model generally decreases and converges as the number of trees grows. From this perspective, the right number of trees corresponds to the moment where the error does not decrease below a certain threshold. [Oshiro et al. \(2012\)](#) show that, in some

cases, the convergence of errors is not ensured. To avoid any risks in our empirical investigation, we determine simultaneously all three parameters:  $(T_{max}, O_{min}, M_{trees})$ , respectively the maximum depth of trees, the minimum number of observations per final region, and the number of trees. To do so, we compare RF regressions out-of-sample scores making those three parameters vary.

*RF interpretation: variables importance and Quantitative Input Influence (QII)*

As the number of features increases, visualization becomes more complex. However, it remains central since it often appears that numerous variables are irrelevant. This part is managed by the calculation of predictors' relative importance (Breiman et al., 1984; Hastie et al., 2009). Moreover, some techniques known as quantitative input influence make it possible to visualize features' impact on labels. Precisely, we use both Partial Dependence Plots (Friedman, 2000; Hastie et al., 2009) and Accumulated Local Effects (Datta et al., 2016).

In order to assess variables' importance, we rely on a generalization of Breiman et al. (1984)'s measure of relevance for a single tree:

$$\mathcal{I}_j^2 = \frac{1}{K} \sum_{k=1}^K \sum_{t=1}^{No-1} i_t^2 I(v(t) = j) \quad (4)$$

where  $\mathcal{I}_j^2$  is the importance of the predictor variable  $X_j$ ,  $K$  represents the number of trees in the forest,  $No$  indicates the number of nodes in the tree,  $v(t)$  is the variable chosen at node  $t$  to split the space, and  $i_t^2$  refers to the improvement gave by the splitting variable, in squared error risk compared to a constant adjustment across all regions. This measure attributes a score to all features giving their determining power on labels.

A Partial Dependence Plot (PDP) is then built to provide a summary of the output dependence on the joint values of the inputs (Friedman, 2000; Hastie et al., 2009). Considering a subset of  $l < p$  inputs  $X_{S, S^c \{1, 2, \dots, p\}}$  of  $X^T = (X_1, \dots, X_p)$ , such that  $f(X) = f(X_S, X_{S^c})$ ,<sup>8</sup> the partial dependence of  $f$  to  $X_S$  is given by:

$$f_S(X_S) = E_{X_{S^c}} f(X_S, X_{S^c}) \quad (5)$$

Note that Equation (5) defines a measure of  $X_S$  effect on  $f(X)$  after accounting  $X_{S^c}$  effect. To calculate this impact in practice, we process as follows. We first assess Individual Conditional Effect (ICE), meaning the partial dependence of  $f(X)$  on  $X_S$  when considering values of  $X_{S^c}$  for a given individual  $i$ :

$$ICE_i = \{\hat{f}(x_S^k, X_{i, S^c}), x_S^k \in [X_S^{min}, X_S^{max}]\} \quad (6)$$

---

<sup>8</sup> $S^c$  being the complementary of  $S$ :  $S \cup S^c = \{1, 2, \dots, p\}$ .

where  $ICE_i$  is the ICE for the  $i$ -th individual,  $X_{i,S^c}$  refers to values of  $X_{S^c}$  of this individual, and  $x_S^k$  are the values of  $X_S$  that vary from its minimum to its maximum with a step  $k$ . This provides a set of points representing a plot of partial dependence of the explained label on the variables included in  $S$  for the  $i$ -th individual. In a second step, we average those plots for all the individuals, and we obtain the PDP.

One of the most important issues in PDPs is that it assumes independence between the predictor for which the partial dependence is computed and the other one. Besides, making  $x_S^k$  vary across all the distribution of  $X_S$  creates a risk to overfit regions with almost no data. In order to overcome this issue, we rely on Accumulated Local Effect (ALE).

ALE (Datta et al., 2016) also proposes to calculate the marginal effect of  $X_S$ . The main differences with PDP can be summarized as follows: ALE is unbiased even when features are correlated, it marginalizes over probable combinations of features, and it is faster to compute. Technically, ALE bases its calculation on existing data intervals for explanatory variables. Moreover, ALE averages the changes of predictions, not the predictions itself. Another significant difference with PDP is that ALE accumulates the local gradients over the range of features  $S$ , giving the effect of those on the predicted variable. Finally, ALE method is centred so that the average effect is zero.

In practice, ALE for one given feature is computed, dividing it into many intervals, and computing the differences in the predictions.<sup>9</sup> First, the uncentred effect is calculated:

$$\hat{f}_j(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i: x_j^{(i)} \in N_j(k)} [f(z_{k,j}, x_{\setminus j}^{(i)}) - f(z_{k-1,j}, x_{\setminus j}^{(i)})] \quad (7)$$

where  $\hat{f}_j(x)$  is the uncentred effect of the variable  $j$ .  $f(z_{k,j}, x_{\setminus j}^{(i)})$  gives the prediction given by the model, and considers the  $i$ -th individual for features values excepted  $x_j$  that takes the value  $z_{k,j}$ . The  $z$  are the values taken by the variable  $X_j$  that has been distributed on a grid defined by a given step. The internal sum adds up the impacts of all individuals within an interval ( $i : x_j^{(i)} \in N_j(k)$ ) that appears as a neighbourhood. This sum is weighted by the number of individuals  $n_j(k)$  present in the  $k$ -th neighbourhood. Finally, we sum the average effect over all intervals.

Second, we center in order to obtain a null main effect:

$$\hat{f}_j(x) = \hat{f}_j(x) - \frac{1}{n} \sum_{i=1}^n \hat{f}_j(x^{(i)}) \quad (8)$$

$\hat{f}_j(x)$  is interpreted as the main impact of the explanatory variable compared to the average prediction of the data.

---

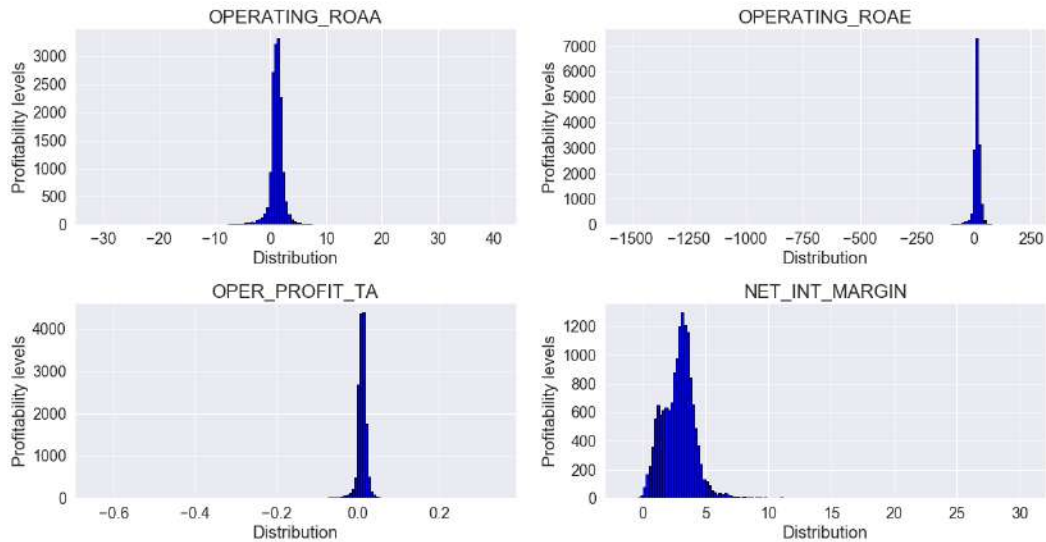
<sup>9</sup>This approximates the local gradients and allows us to compute ALE using Random Forest regressions.

From a practical point of view, some comments should be made. We use the sickit-learn package for RF regressions that includes PDPs calculation. ALE being a far more recent technique, it has not been widely developed yet. As for now, and to the best of our knowledge, only the ALE-python package proposes Accumulated Local Effect in python; we thus rely on this package in our empirical approach.<sup>10</sup> In the representation, we generally cut the extreme values of the variable we look at; indeed, the deviation from a value to another increasing as we move towards the distribution queues. As a result, the QII curves stretch for high values of the input and the additional information loses in relevance.

## 4 Descriptive statistics

Before presenting the results of RF regressions, we display some descriptive statistics on the explained variables' distribution and the features correlations.

Figure 4 – Profitability variables' distribution



Source: Author's calculations from FitchConnect data.

Figure 4 shows the distribution of our four profitability variables. As can be seen, the output variables are highly concentrated around their respective average. All four variables also display outliers,<sup>11</sup> but RF regressions are robust to the presence such observations. Indeed, each tree is built on a new randomly selected sample of features and data. Therefore, among the entire

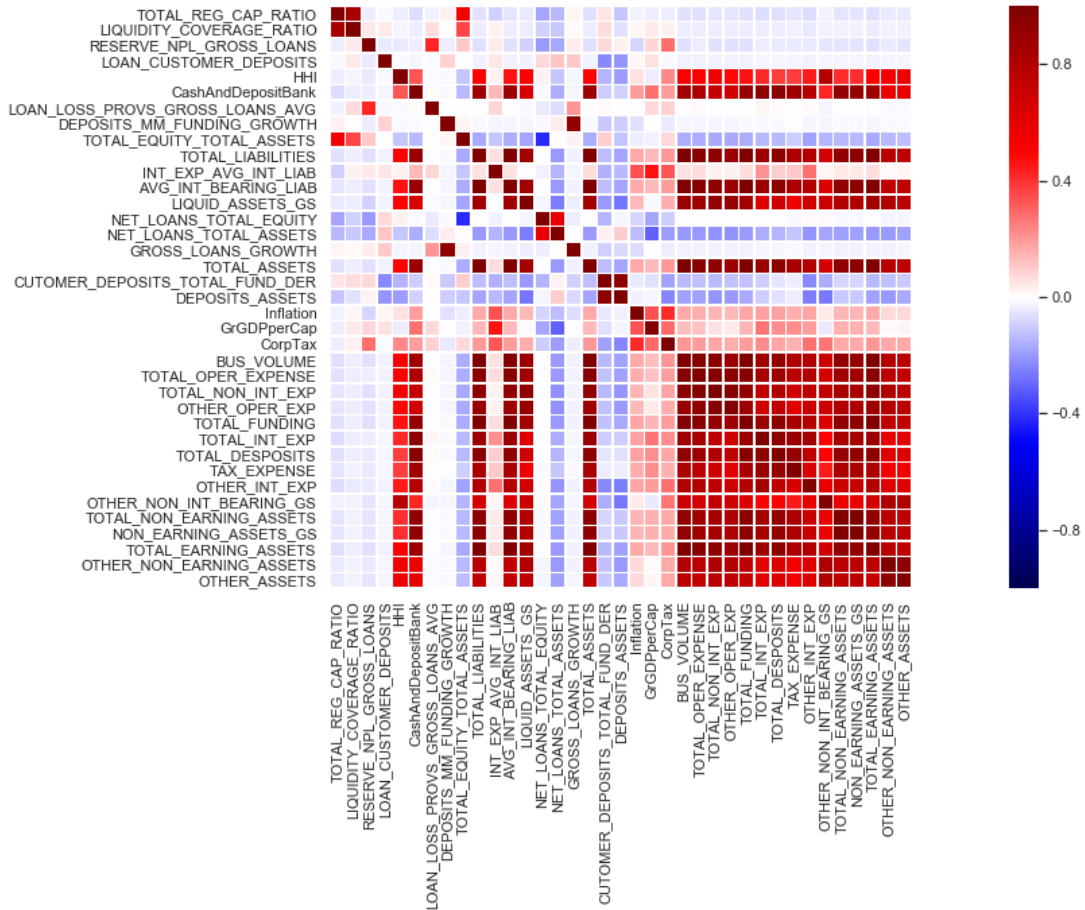
<sup>10</sup>Note that this package includes Monte-Carlo simulations. Therefore, we provide the robustness checks consisting in assessing the average results over a 100 RF, only on PDPs.

<sup>11</sup>Those outliers are not apparent being too few.

forest, outliers are taken into account.<sup>12</sup>

As mentioned earlier, Partial Dependence Plots' calculation can be biased when features are correlated. As shown by Figure 5 which displays the linear correlations between all explanatory variables, high correlation exists for some features. Indeed, the sample of variables going from Business Volume to Other Assets displays high correlation levels with numerous other inputs. The use of ALE in order to support PDPs' results is relevant.

Figure 5 – Linear correlation between features



Source: Author's calculations from FitchConnect data.

<sup>12</sup>To ensure that no bias is created by those high values, we constrained the sample selection of the robustness check by averaging results over 100 RF regressions. To do so, we forced the model to select data from all deciles. We controlled the sampling comparing the new and old distributions' first two moments.

## 5 Results: Random forest with ROAA

Let us consider in this section only the results for RF regressions on ROAA, in the sample including LCR. The results averaging outcomes over 100 RFs, for the three other profitability variables, in the whole sample and with the LCR proxy are all presented in Section 6 dedicated to robustness checks. We fix RF parameters and make vary the number of trees, the minimum observations per leave, and the maximum number of nodes/ depth of trees simultaneously.<sup>13</sup> We compare the set of parameters that maximize the out-of-sample  $R^2$ : 200 trees with a minimum of 10 observations per final region, and a maximum of 12 nodes.

Whatever the profitability variable we explain, our empirical approach is based on four steps. In the first one, we look at the model's quality, relying on its  $R^2$  in- and out-of-sample. Once the usefulness of the model is confirmed, we focus in a second step, on the most important variables. Precisely, our goal is to compare the predicting power of regulatory ratios with those of other determinants. The third stage consists in the calculation of most important variables' PDP and ALE. Finally, we interpret the results. This approach in four steps is also followed in all robustness checks.

To check the model's quality, we look at the model's performance in and out-of-sample, and compare those two scores to those issued from linear regressions (OLS).<sup>14</sup> Table 1 displays the results for those two regressions on ROAA.

Table 1 – Model's quality: RF versus OLS

Sample	Models	
	RF	OLS
In-sample	0.85	0.25
Out-of-sample	0.43	-0.32

Source: Author's calculations. The table shows the coefficient of determination ( $R^2$ ) scores.

As can be seen, RF outperforms OLS in predicting in and out-of-sample.<sup>15</sup> Even if OLS is undoubtedly not well suitable to estimate our model, it remains relevant to compare its performance to that of RF: it confirms the usefulness of relying on a non-linear model and constitutes a benchmark from which we can establish RF's predicting power. We still propose an alternative method (regularized least squares) in section 6.

Now that the relevance in the use of RF regressions is confirmed, we look into the classification of the most important variables. The calculation of the predicting power of each feature,

<sup>13</sup>We estimated multiple RF making the number of trees vary from 10 to 1000, the depth from 4 to 15 and the maximum number of observation from 8 to 12.

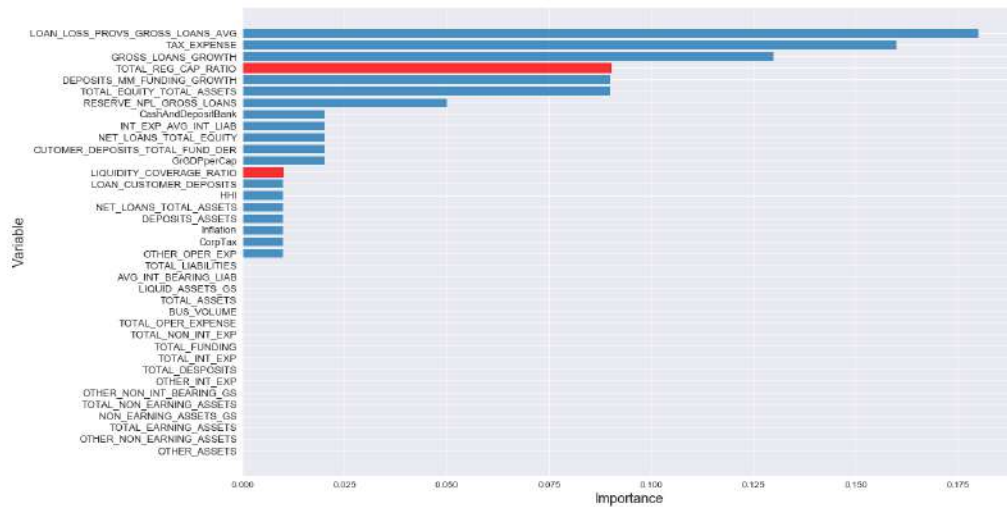
<sup>14</sup>OLS are run using only the 6 first important variables underlined with RF and the two regulatory ratios.

<sup>15</sup>The negative  $R^2$  for out-of-sample OLS means that the linear model performs worse than a constant. It is certainly due to the sample selection procedure. The data contained in the test set are sufficiently different from the data in the learning set so that the linear model predicts poorer than a constant.



as described in Section 3.2, allows us to order explanatory variables. Figure 6 shows the importance of each input. We can observe that TCR has relatively high importance, taking the fourth place among all variables. Including TCR in the features improves the model predictions by 8%. Looking into the partial dependence and accumulated local effect of this ratio to ROAA is therefore relevant. However, the LCR only takes the thirteenth place and enhances the predicting power of the model only by 1%. PDP and ALE of LCR present less interest: we still can calculate the impact of liquidity ratio on ROAA, but we should expect a weak effect.

Figure 6 – ROAA - Variables importance



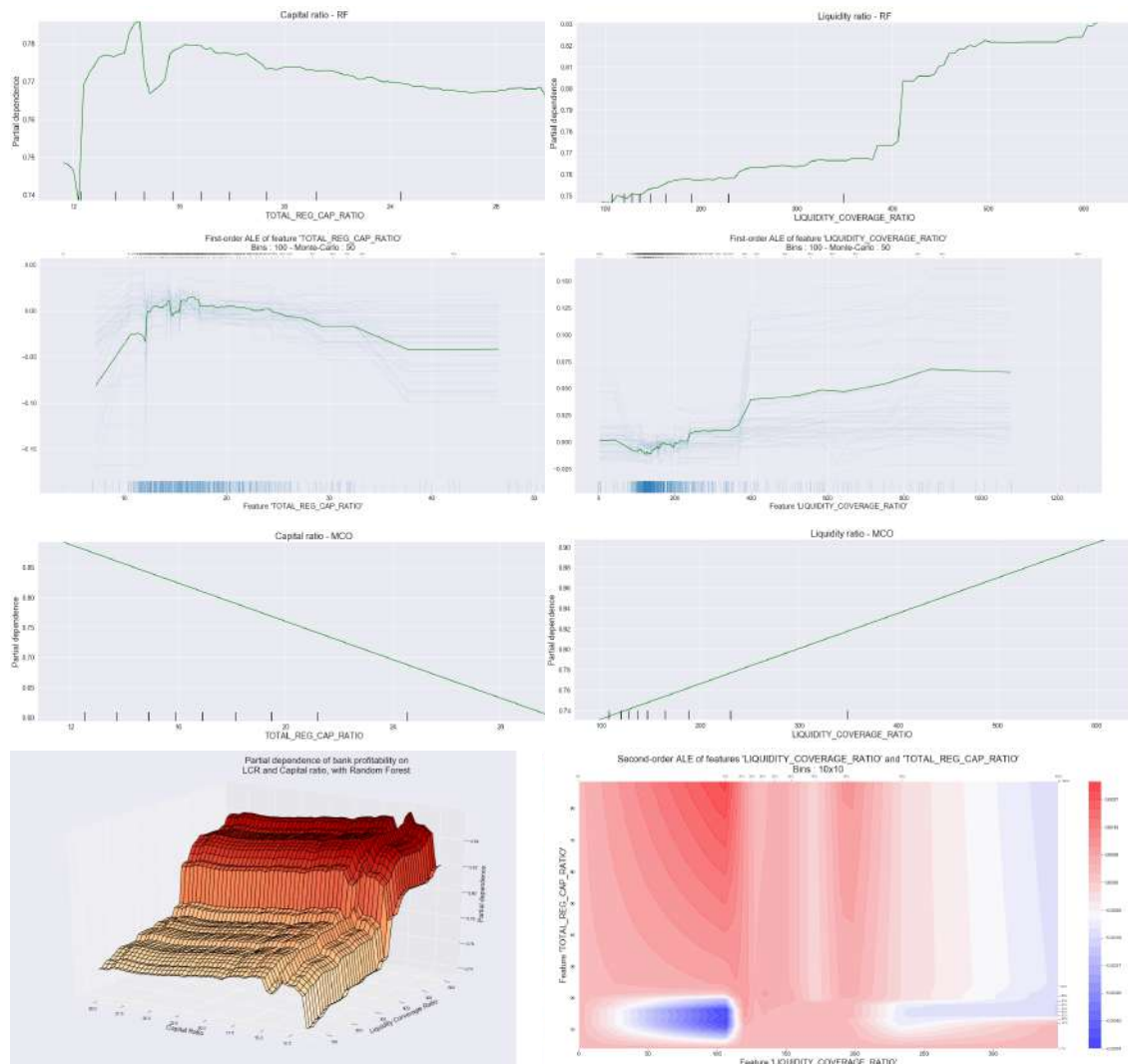
Source: Author's calculations from FitchConnect data. In red the two regulatory ratios.

The third step of our analysis concerns the calculation of features' marginal effect on ROAA. Precisely we plot the one-way and two-ways PDP and ALE of TCR and LCR in relation to ROAA. Then we also look into partial dependence and local effect of the first six most important variables. The results are compared with those of the linear model.

Figure 7 shows the results of the two regulatory ratios' PDP and ALE (one and two-ways) from the RF regression on ROAA, and the PDP of the linear model (a line whose slope is defined by the OLS parameter estimated). As can be seen, the marginal effect of capital on profitability is non-linear: a positive relationship links this regulatory ratio to ROAA for values below 15 to 20%, and a negative effect of TCR on banks' performance is captured for values higher than 20%. This result is obtained with both PDP and ALE. The linear model remains irrelevant in that it displays an overall negative relationship between TCR and ROAA. The non-linearity detected by RF strongly supports [Kraus and Litzenberger \(1973\)](#)'s trade-off theory of the capital structure: on average, a capital ratio optimum exists and is comprised between 15 and 20%. As

for the LCR, RF highlights a linear or quasi-linear<sup>16</sup> relationship between the liquidity Basel III ratio and banks' profitability. This finding endorses the fact that detaining High Quality Liquid Assets goes with healthier activity and low costs funding. Though, what should be kept in mind regarding LCR's impact is that it has weak power on the prediction outcomes of the RF regressions. In other words, LCR does not appear to be a powerful determinant of profitability.

Figure 7 – ROAA - Regulatory ratios' marginal impact



Source: Author's calculations from FitchConnect data. The first two plots display PDPs of TCR and LCR. The two following ones show ALE of TCR and LCR. The fifth and sixth plots represent OLS results. The last line depicts interactions between capital and liquidity ratios in their effect on ROAA. Distribution of explanatory variables can be seen in axes (deciles and percentiles are displayed for, respectively, PDP and ALE).

<sup>16</sup>It seems there would be a threshold above which appears a new and higher regime of profitability. However, looking at the LCR distribution, this relation is captured for too few observations to be considered as reliable.

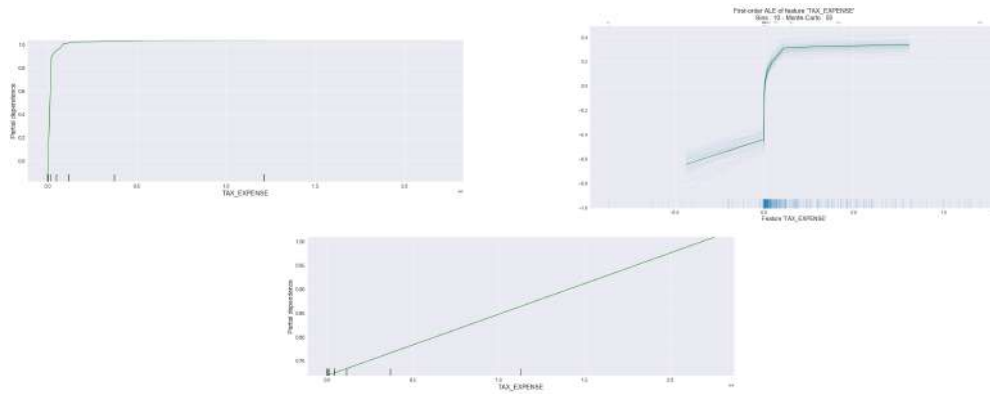
The two-ways PDP and ALE show the cumulative effect of TCR and LCR on ROAA, depending on each-other's level. Indeed, it does not appear that ratios' impact structure is changed when taking into account both of them. However, it is clearly shown that the overall impact changes when moving towards higher values of both ratios. Therefore, an optimal combination appears: on average, when taking reasonable values (between second and ninth deciles of both distributions) for the two Basel III requirements, a bank finds interest in converging around 20% of TCR and the minimum LCR required, namely 100%. However, here again, the interpretation of those two-ways plots should be taken carefully in the sense that LCR gives weak importance in the model's prediction. Finally, some comments on scales should be made. Note that the intervals (y-axis) in which PDPs and ALEs vary are small showing that, even regarding TCR, the determining power may be sure, it is not highly significant.

The fourth step of our analysis must also be conducted on the most important variables displayed in Figure 6, excluding TCR that has been already studied. Figure 8 shows Random Forest regressions' PDP, ALE and OLS PDP for the six first determinants of ROAA. While each variable would deserve particular attention, our main interest here is to establish some categories of powerful determinants of banks' profitability. Thus, we will have to cross-checked our results with those provided in Section 6. Still, let us make some comments. A strong positive relationship links tax expenses to ROAA. As we shall see hereafter, we attribute this finding to an inverse causal effect: the more profitable a bank is, the higher its tax expenses. Loan loss reserves appear to have a negative impact on ROAA. This result is quite intuitive since those reserves represent a cost/unused capital for the bank. The total equity to total assets ratio, which is close to the regulatory leverage ratio, displays a positive impact on profitability. This result underlines the idea according to which high regulatory ratios improve banks' ability to generate interests. This can be explained by a decrease in funding costs. The same logic seems to link the reserve NPL to gross loans ratio, to bank performance. Unsurprisingly, gross loan growth is positively associated with ROAA. Finally, deposits and money market funds have a positive effect on performance. On average, those instruments appear to have positive returns. From those findings, determinants can be classified among four categories: expenses, provisions for credit risks, assets and equity variables, and loan features.

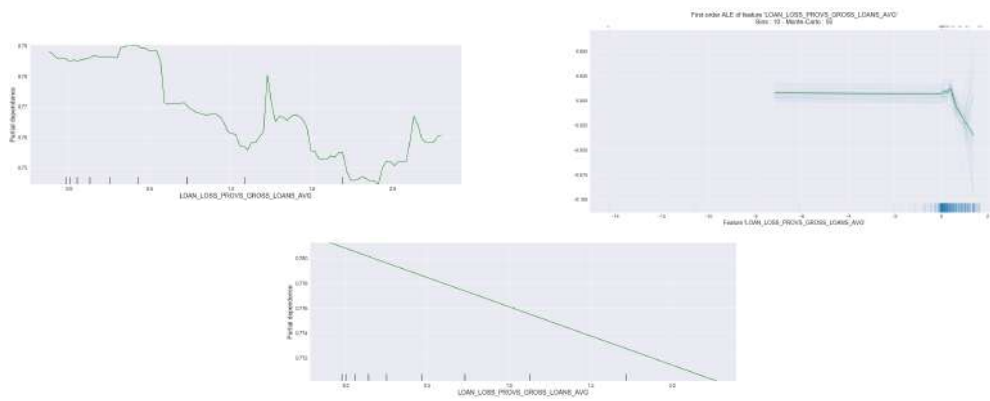
We can summarize our main results as follows. TCR is a strong determinant of profitability though its marginal impact seems low. The trade-off theory of an optimum for the capital ratio is confirmed, and takes a value between 15 and 20%. LCR is shown to be a weak determinant of banks' profitability. Accumulated effects appear when looking at both capital and liquidity impact. Let us now conduct numerous robustness checks to ensure the stability and reliability of our results. The goal of the next section is also to validate the four categories of profitability determinants.

Figure 8 – ROAA - 6 first most important variables' impact

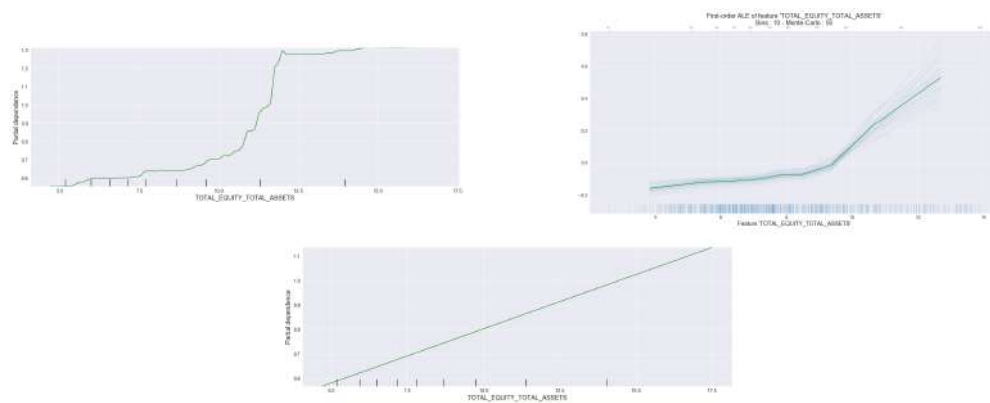
(a) Tax expenses



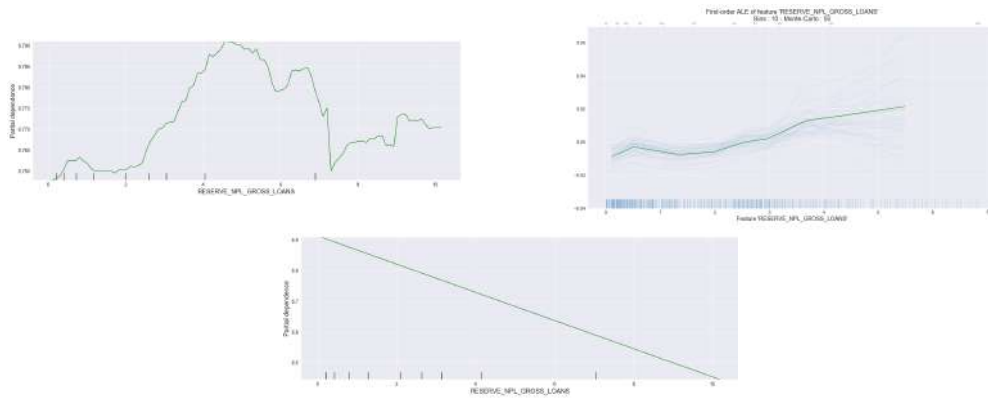
(b) Loan loss provisions to gross loans



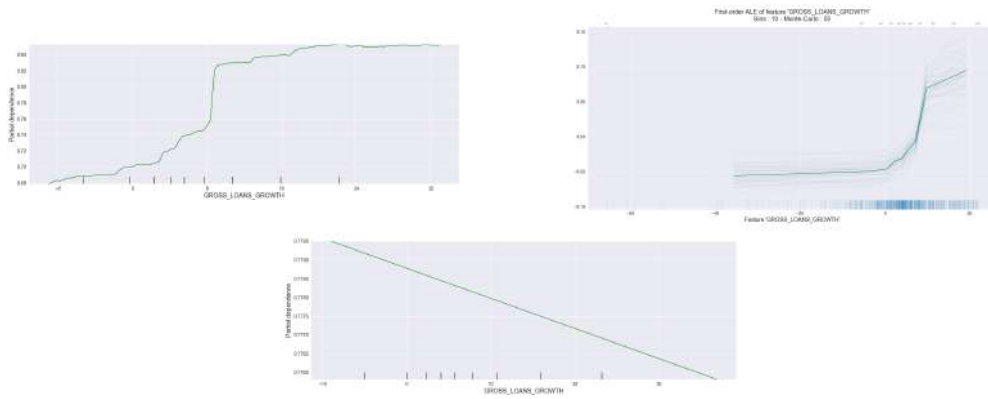
(c) Total equity to total assets



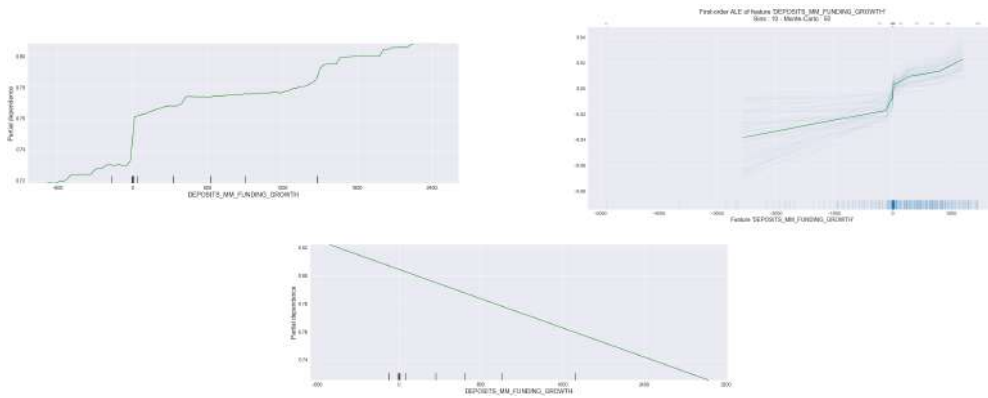
(d) Reserve NPL to gross loans



(e) Gross loan growth



(f) Deposits, money market funding growth



Source: Author's calculations from FitchConnect data. For each of the first six most important variables: RF's PDP on top the left, RF's ALE on the top right, and PDP of OLS at the bottom.

## 6 Robustness

### 6.1 Results for other profitability variables

To assess the robustness of our findings to the profitability measure, we use three other bank profitability variables (ROAE, OPTA, and NIM), and apply the same empirical approach as for ROAA. The results of the corresponding RF regressions are displayed in Appendix B.1.

All three models show high in-sample prediction performance, and higher out-of-sample  $R^2$  than those of linear regressions in-sample (Table 3). The out-of sample explicative power of RF regressions goes from 41% to 81%. These results corroborate those obtained with ROAA, and have sufficient prediction power to be trusted. Regarding the most important variables (Figure 9), and the place took by the two Basel III ratios among these, the results also confirm those of the regression with ROAA: capital ratio takes a prominent place, except in the model with NIM, and LCR never appears as a powerful predictor of the model. For both ROAE and OPTA, an optimum level of capital is visible around 15% (Figure 10b): the effect of TCR being positive before and becoming null or negative beyond the optimum. In the three models, LCR shows positive impact on profitability (Figure 12b). The models with ROAE and OPTA confirm the accumulated effect of increasing simultaneously capital and liquidity. Finally, we find some stability in the set of the most important variables. Among the predictors that regularly appear in the first most important variables, we can find the loan loss provisions to gross loans ratio, the tax expenses, the gross loans growth, deposits and money market funding growth, and the NPL reserves to gross loans ratio. Overall, except for the case of capital in the model with NIM as the explained variable, all results are corroborated.

### 6.2 Averaging results of 100 regressions taking into account outliers

The second robustness test we conduct concerns the training data sample selection. Indeed, RF regressions are trained on randomly selected data among the full sample, and then tested on the remaining observations. In order to ensure that our results are not conditioned to a particular selection of the data, we average results over 100 RF. For each RF regression, we re-sample randomly the training data set, and constrain the sampling to select data from every decile. This procedure allows us to take into account the full sample, including outliers. We ensure the proper distribution of data by checking the first two moments of the selected data. The results of those regressions are shown in Appendix B.2.<sup>17</sup>

On average, over 100 regressions, all four models have higher out-of-sample predicting power (from 43% to 76%) than OLS in-sample forecasting performance (from 17% to 45%, Table 4). Among the first most important variables (Figure 13), TCR always occupies the sixth place or better (except for the model with NIM). An optimum for the capital ratio is detected around

---

<sup>17</sup>The computational cost being too high, results of ALEs and two-ways QII are not displayed.

15% (Figure 14a). In all models, LCR improves weakly the predicting power of the forest. Its effect remains strictly positive (Figure 14b). The main determinants of profitability variables are stable compared to what we have shown earlier. Among those, we can find: tax expenses, the loan loss provisions to gross loans ratio, the gross loan growth, the NPL reserves to gross loans ratio, and the deposits and money market funds growth. Once more, most of our previous results are confirmed.

### 6.3 Results with LCR proxy

Our aim in this section is to verify that LCR is not an important determinant of banks' profitability. As mentioned earlier, our database does not include this regulatory ratio for US banks. Therefore, we take the highest correlated variable with LCR that permits to have a large number of observations, and we use it as a proxy for LCR. We retained the total derivative liabilities at fair value. This variable is linearly correlated at 95% with LCR and allows us to consider a sample of more than 4 000 observations. Outputs for this robustness test can be found in Appendix B.3.

In this model, the out-of-sample quality remains high (63%, Table 5). TCR takes the fourth place as the most important variables, while our LCR proxy takes the twenty-ninth (Figure 15). From this perspective, the results regarding LCR in previous regressions are corroborated.

### 6.4 Results in the large sample over 50 regressions

Here, we test the validity of our results regarding the capital ratio. We aim to confirm the existence of an optimum below which the relation between capital and profitability is positive and upside-down for higher values of capital. To do so, we run RF regressions on our four profitability variables using a dataset of more than 15 000 observations.<sup>18</sup> To ensure the stability of our results, we average outcomes of 50 RF regressions. The results of those regressions are displayed in Appendix B.4.

As shown, models' quality remains high (Table 6). The optimum of the capital ratio is verified for all four models, and varies from 13% to 20% (Figure 17). Though, it appears that TCR takes a lower place among the most important variables (Figure 16). The first determinants of profitability are close to those obtained in previous regressions.

### 6.5 Results of regularized least squares

In this last robustness exercise, we propose a more suitable linear method in addition to basic OLS: regularized least squares. This approach consists in adding a regularisation term to the error function to be minimized. In doing so, we control for over-fitting. Moreover, by gradually

<sup>18</sup>Note that this robustness check also allows us to confirm our results on the capital ratio optimum with US banks.

increasing the term of regularization, the model is forced to keep only the most important variables.

We use the Ridge Regression over all variables in the dataset with the LCR and ROAA as dependent variable. We then estimated multiple models making vary the regularisation parameter and resampling randomly the training and test sets. Doing so we control for both outlier values and the regularisation parameter value. Outlier playing a key role in determining the bias of least squares procedure, we also normalized all data.

Results show positive but low  $R^2$ , in and out-sample, confirming that a linear model is not well fitted in our case. <sup>19</sup>

## 7 Conclusion

This paper analyses the determinants of banks' profitability. Specifically, since the implementation of Basel III new regulatory requirements is under debate in the context of intensified competition, we investigate the capital (TCR) and liquidity (LCR) ratios' predicting power of banks' performance. Moreover, we establish a comparison between those two variables and other potential determinants of banks' profitability.

Relying on a large dataset of more than 1 000 observations and containing both regulatory ratios, we do not find evidence of strong determining power of LCR, although its marginal effect is positive. On the contrary, we show that TCR is a key determinant of banks' profitability. Furthermore, we corroborate the trade-off theory of the capital structure, which states the existence of an optimal capital ratio. Conducting strong robustness checks, we estimate this optimum to be, on average, between 15% and 20%. We also show that a positive accumulation effect lies under the combined impact of capital and liquidity. Finally, we show that profitability's determinants can be classified into four groups: tax variables, provisions for credit risks (NPL reserves, loan loss provision...), assets and equity variables (TCR, leverage ratio...), and loans variables (gross loan growth, net loans...). The validity of our results is ensured by the use of Random Forest regressions that are known for their ability to perform out-of-sample and capture complex relationships between variables. The reliability of our findings has been successfully tested through numerous robustness checks.

Tackling the determining place of regulatory ratios in banks' profitability, our paper has important policy implications. Indeed, we assess that the optimal capital ratio falls between 15% and 20%, which is higher than Basel III requirements. This implies that the current minimum regulatory capital ratio is not binding for most banks. We also show that LCR has a weak and positive effect on profitability. Therefore, the actual regulatory requirements do not prevent banks from being profitable. The optimal capital ratio is specific to each bank. From this

---

<sup>19</sup>Results are not displayed in this paper but remain available at the reader's request.



perspective, all banks which have a TCR below its optimum still have the flexibility to change their capital and liquidity structure in order to be compliant with Basel III and increase their profitability. In other words, regulatory accords find efficiency by ensuring banks' internal financial stability without compelling their business.

The present study can be extended in several ways. From a methodological point of view, a full dataset including LCR data for US banks should undoubtedly be worthwhile. Moreover, as we show the existence of non-linearities in the relation between capital and profitability, the use of a dynamic Panel Smooth Transition Regression (PSTR) would be particularly relevant in order to account for any temporal effect. This matter finds interest in investigating Berger (1995)'s hypotheses according to which the sign of the relation between capital and banks' performance depends on the cycle. From an economic point of view, further research should be done to investigate the impact of regulatory ratios on banks' profitability. Specifically, it will soon be possible to study the effect of Net Stable Funding Ratio (NSFR).

## References

- U. Albertazzi and L. Gambacorta. Bank profitability and the business cycle. *Journal of Financial Stability*, 5(4):393 – 409, 2009. ISSN 1572-3089. 2, 3
- 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. 5
- M. Balling. Modigliani-miller, basel 3 and crd 4. *SUERF Policy Notes*, (2):1–8, 2015. 5
- A. Berger. The Relationship between Capital and Earnings in Banking. *Journal of Money, Credit and Banking*, 27(2):432–56, 1995. 2, 3, 5, 23
- A. Berger, R. J. Herring, and G. Szego. The Role of Capital in Financial Institutions. *Center for Financial Institutions Working Papers*, 1995. 2, 3
- A. N. Berger and C. Bouwman. Bank Liquidity Creation. *Review of Financial Studies*, 22(9): 3779–3837, 2009. 2, 6
- E. Bordeleau and C. Graham. The Impact of Liquidity on Bank Profitability. *Bank of Canada*, (10-38), 2010. 2, 6
- C. Borio, L. Gambacorta, and B. Hofmann. The influence of monetary policy on bank profitability. *International Finance*, 20:48–63, 04 2017. 2
- P. Bourke. Concentration and other determinants of bank profitability in Europe, North America and Australia. *Journal of Banking & Finance*, 13(1):65 – 79, 1989. ISSN 0378-4266. 6

- L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. 8, 9
- L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. 1984. 10
- N. C. Chia and J. Whalley. The Tax Treatment of Financial Intermediation. *Journal of Money, Credit and Banking*, 31(4):704–19, 1999. 2
- A. Datta, S. Sen, and Y. Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. pages 598–617, 05 2016. doi: 10.1109/SP.2016.42. 10, 11
- R. DeYoung and T. Rice. Non Interest Income and Financial Performance at US Commercial Banks. *Financial Review*, 39:101 – 127, 02 2004. 6
- A. Dietrich and G. Wanzenried. The determinants of commercial banking profitability in low-, middle-, and high-income countries. *The Quarterly Review of Economics and Finance*, 54(3): 337 – 354, 2014. ISSN 1062-9769. 6
- 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. 2, 3
- J. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 11 2000. doi: 10.1214/aos/1013203451. 10
- A. García-Herrero, S. Gavilá, and D. Santabárbara. What explains the low profitability of Chinese banks? *Journal of Banking & Finance*, 33(11):2080–2092, November 2009. 2, 6
- J. Goddard, H. Liu, P. Molyneux, and J. O. Wilson. Do Bank Profits Converge? *European Financial Management*, 19(2):345–365, 2013. 5, 6
- R. Gropp and F. Heider. The Determinants of Bank Capital Structure. *Review of Finance*, 14 (4):587–622, 03 2010. 2
- 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. 2, 3, 8, 9, 10
- 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. 5
- M. C. Keeley. Deposit Insurance, Risk, and Market Power in Banking. *American Economic Review*, 80(5):1183–1200, December 1990. 1

- A. Kraus and R. H. Litzenger. A State-Preference Model of Optimal Financial Leverage. *Journal of Finance*, 28(4):911–22, 1973. 2, 3, 5, 15
- 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. 5
- M. Miller. Do the M & M propositions apply to banks? *Journal of Banking & Finance*, 19(3-4):483–489, 1995. 5
- F. Modigliani and M. H. Miller. The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, 48(3):261–297, 1958. ISSN 00028282. 5
- P. Molyneux and J. Thornton. Determinants of European bank profitability: A note. *Journal of Banking & Finance*, 16(6):1173–1178, December 1992. 2, 6
- M. Osborne, A.-M. Fuertes, and A. Milne. Capital and profitability in banking : Evidence from us banks. 2012. Working Paper. 2, 3, 5
- T. M. Oshiro, P. S. Perez, and J. A. Baranauskas. How Many Trees in a Random Forest? In P. Perner, editor, *Machine Learning and Data Mining in Pattern Recognition*, pages 154–168, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31537-4. 9
- 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. 2, 3
- T. Xu, K. Hu, and S. Das, Udaibir. Bank Profitability and Financial Stability. *IMF Working Paper*, 2019. 1, 3

## A Data sources and definitions

Table 2 – 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
LIQUIDITY COVER-AGE RATIO	Short term liquidity ratio as define under Basel III agreements. Ratio of High Quality Liquid Assets to cash flow under stress.	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

Table 2 – (continued)

LOAN CUSTOMER DE- POSITS	Loan to customer deposit accounts, which can be withdrawn on demand or short notice.	FitchConnect
LOAN LOSS PROVI- SION	Provision made by a bank to hedge against loan losses.	FitchConnect
LOAN LOSS PROVI- SION GROSS LOAN AVG	Ratio of loan loss provision to gross loans.	FitchConnect
DEPOSITS MM FUND- ING GROWTH	Growth rate of deposits to money market funding.	FitchConnect
TOTAL EQUITY TO- 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 EQUITY	Ratio of net loans to total equity.	FitchConnect
NET LOANS TOTAL ASSETS	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
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

Table 2 – (continued)

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.	
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

## B Robustness outputs

### B.1 Results for other profitability variables

Table 3 – Model’s quality: RF versus OLS - ROAE, OPTA, NIM

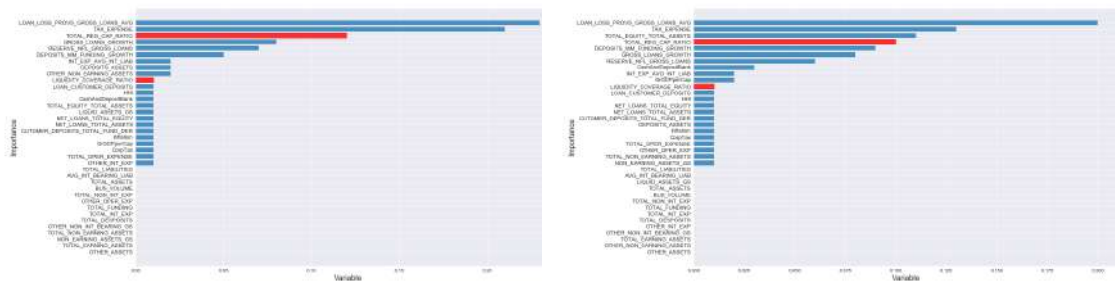
Sample	Models					
	ROAE		OPTA		NIM	
	RF	OLS	RF	OLS	RF	OLS
In-sample	0.81	0.18	0.84	0.21	0.93	0.49
Out-of-sample	0.66	-0.19	0.41	-1.65	0.81	0.37

Source: Author’s calculations. The table shows the coefficient of determination ( $R^2$ ) scores.

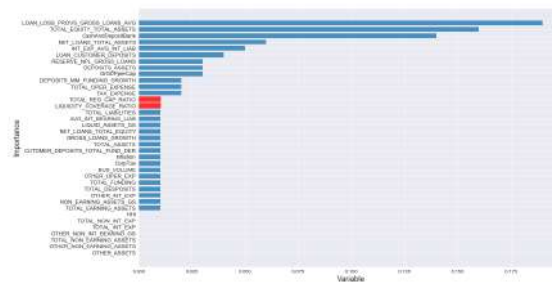
Figure 9 – Variables’ Importance - ROAE, OPTA, NIM

(a) ROAE

(b) OPTA



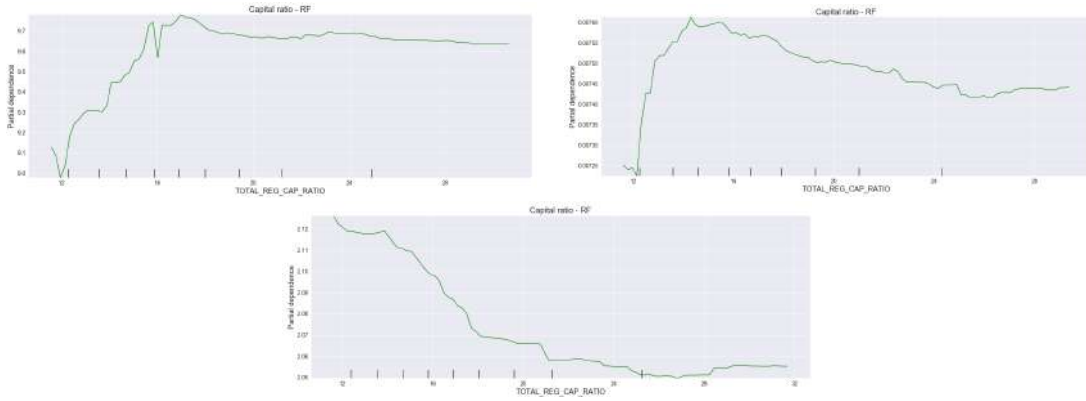
(c) NIM



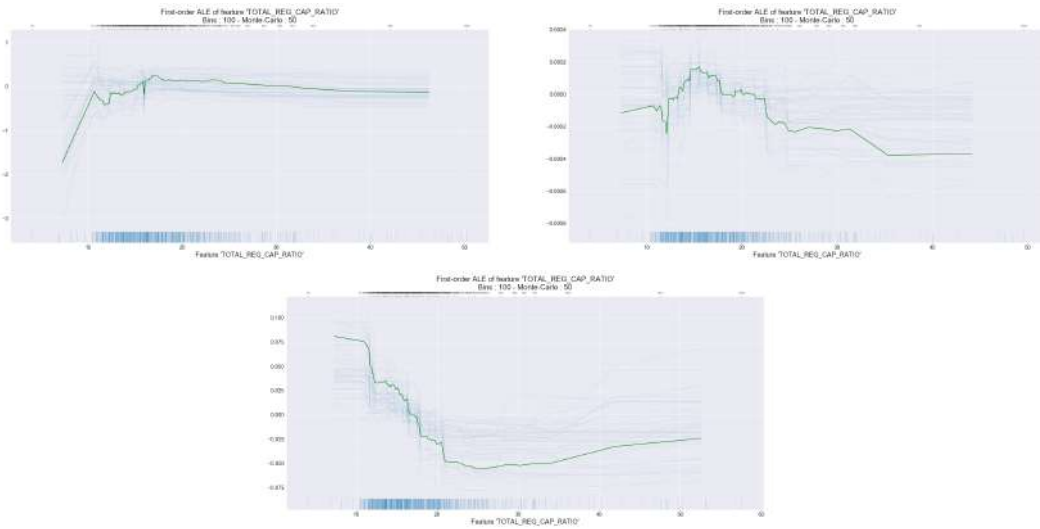
Source: Author’s calculations from FitchConnect data.

Figure 10 – PDP and ALE of TCR - ROAE, OPTA, NIM

(a) Partial Dependence Plots



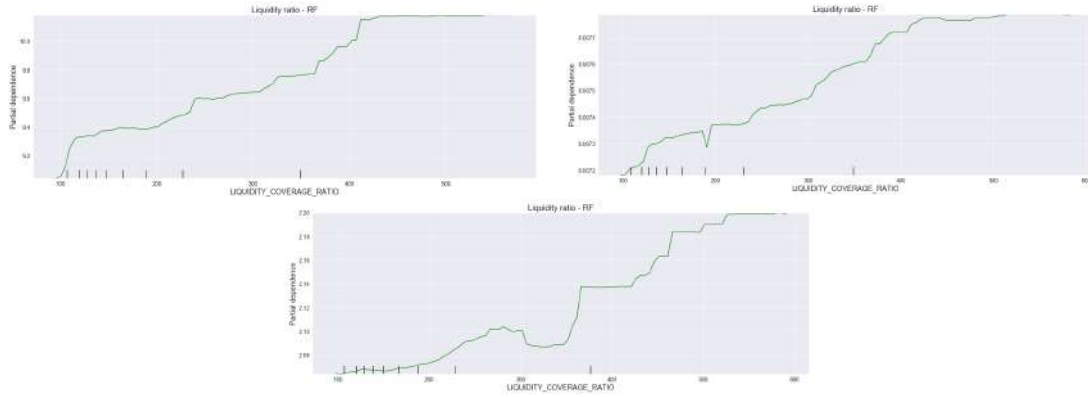
(b) Accumulated Local Effect



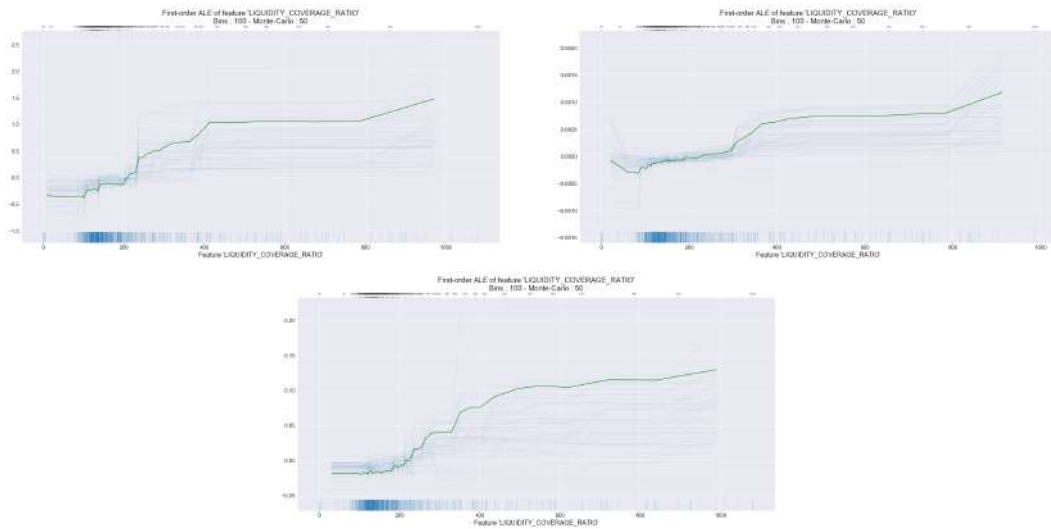
Source: Author's calculations from FitchConnect data.

Figure 11 – PDP and ALE of LCR - ROAE, OPTA, NIM

(a) Partial Dependence Plots



(b) Accumulated Local Effect

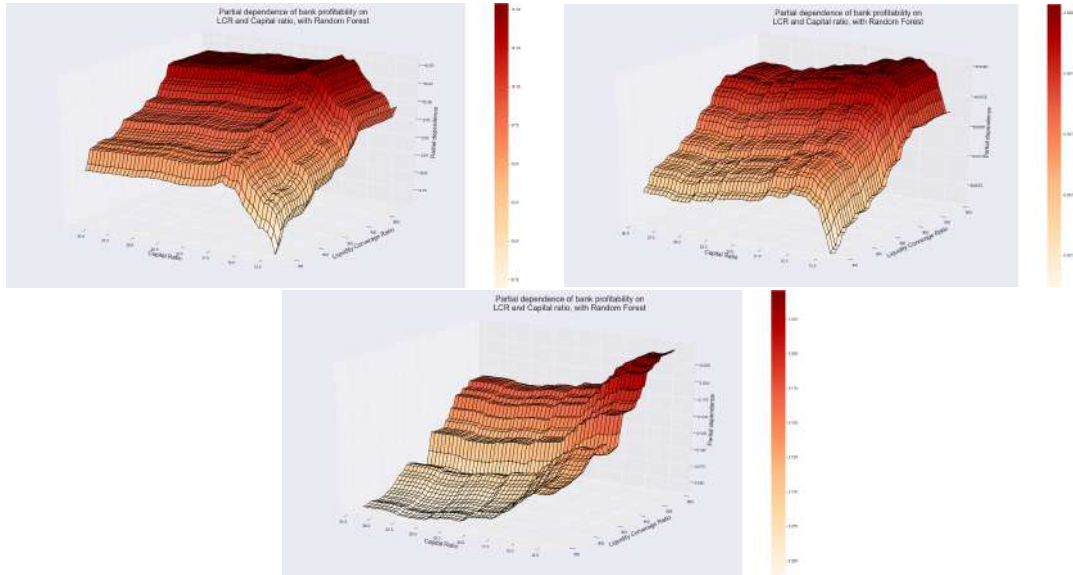


Source: Author's calculations from FitchConnect data.

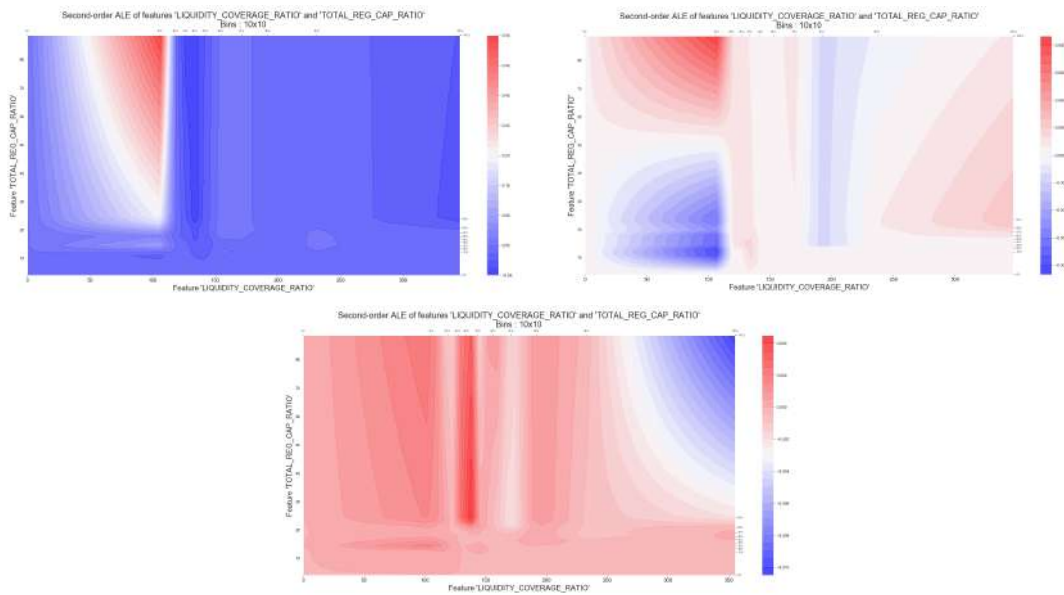


Figure 12 – Two ways PDP and ALE - ROAE, OPTA, NIM

(a) Partial Dependence Plots



(b) Accumulated Local Effect



Source: Author's calculations from FitchConnect data.

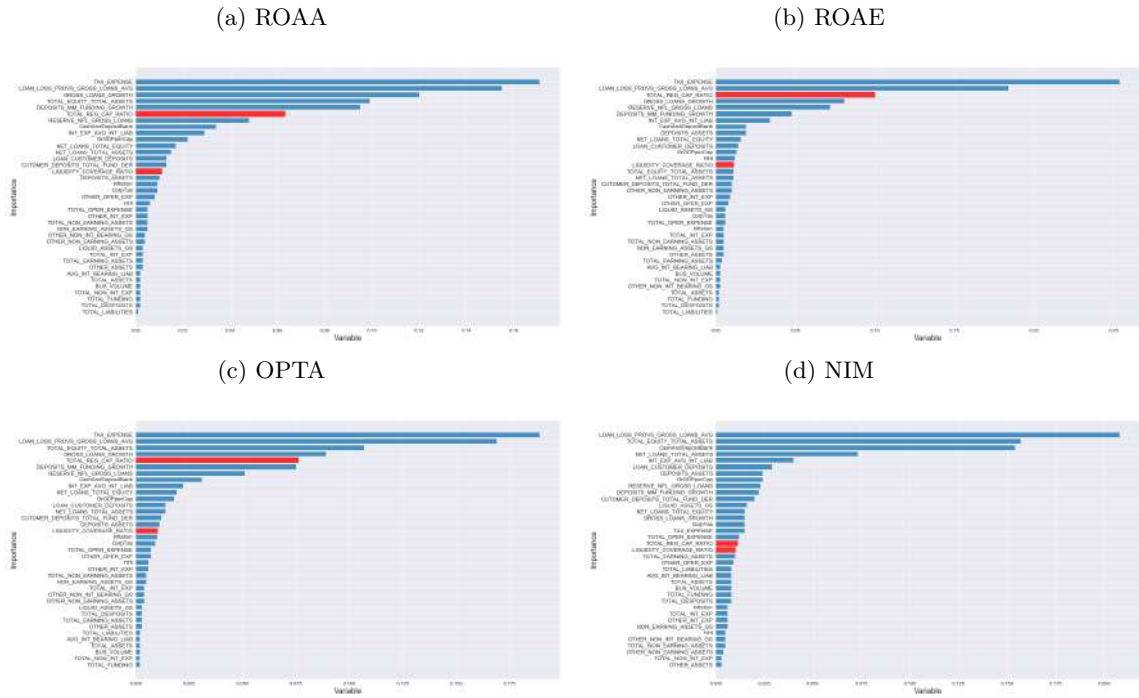
## B.2 Averaging results of 100 regressions taking into account for outliers

Table 4 – Model’s quality over 100 RF

Sample	Models							
	ROAA		ROAE		OPTA		NIM	
	RF	OLS	RF	OLS	RF	OLS	RF	OLS
In-sample	0.86	0.23	0.84	0.19	0.85	0.17	0.93	0.45
Out-of-sample	0.49	-0.41	0.43	-1.22	0.45	-0.68	0.76	-0.43

Source: Author’s calculations. The table shows the coefficient of determination ( $R^2$ ) scores.

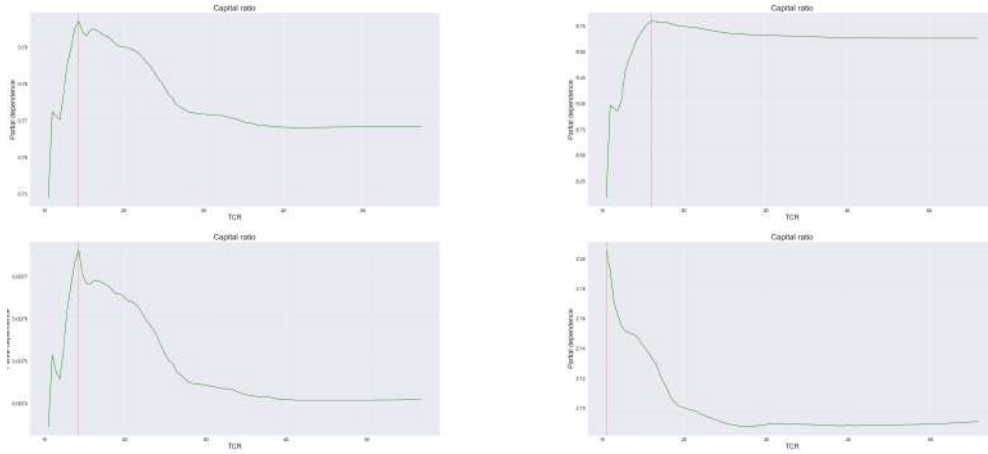
Figure 13 – Variables’ Importance over 100 RF



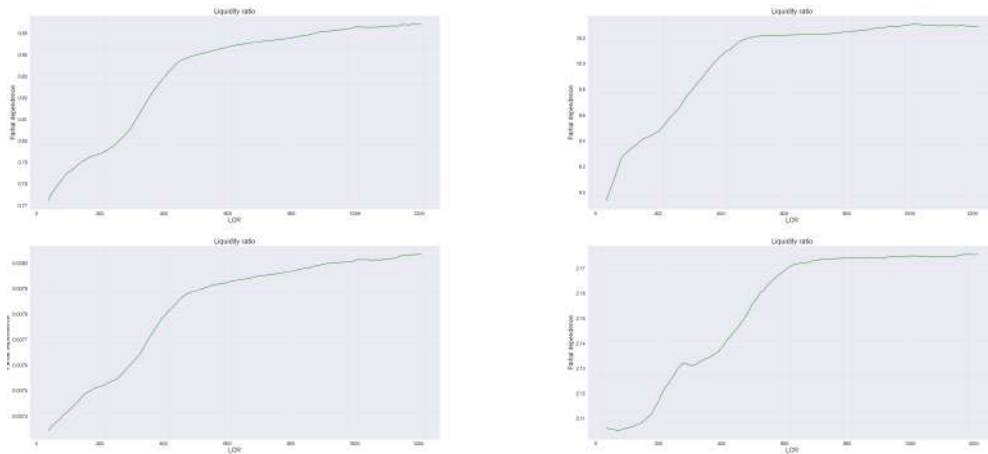
Source: Author’s calculations from FitchConnect data. In red the two regulatory ratios.

Figure 14 – One way PDP over 100 RF - ROAA, ROAE, OPTA, NIM

(a) Partial Dependence Plots - TCR



(b) Partial Dependence Plots - LCR



Source: Author's calculations from FitchConnect data. In red the level of the optimum TCR.

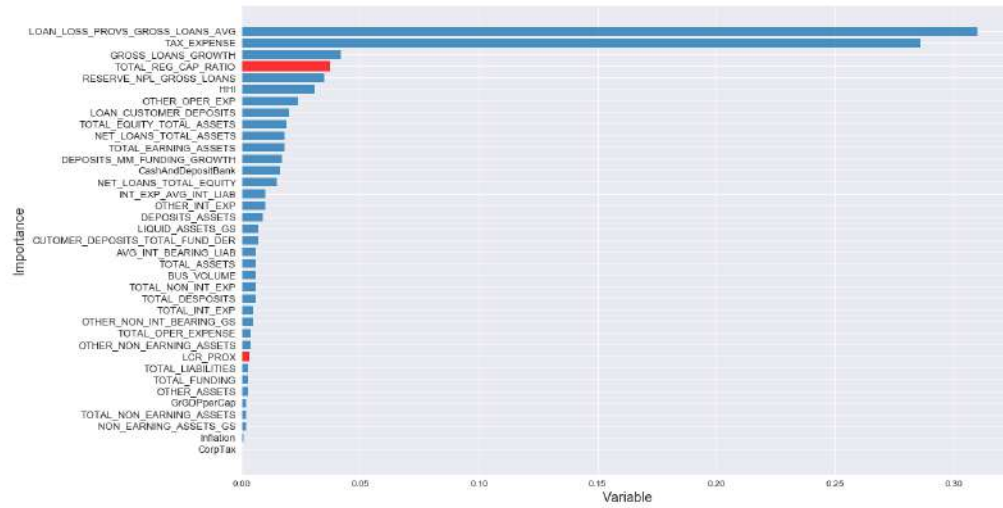
### B.3 Results with LCR proxy

Table 5 – Model’s quality with LCR proxy over 100 RF - ROAA

Sample	Models	
	RF	OLS
In-sample	0.86	0.32
Out-of-sample	0.63	0.24

Source: Author’s calculations. The table shows the coefficient of determination ( $R^2$ ) scores.

Figure 15 – Model with LCR proxy - ROAA - Variables importance



Source: Author’s calculations from FitchConnect data. In red the two regulatory ratios.

## B.4 Results in large sample over 50 regressions

Table 6 – Model’s quality in large sample over 50 RF

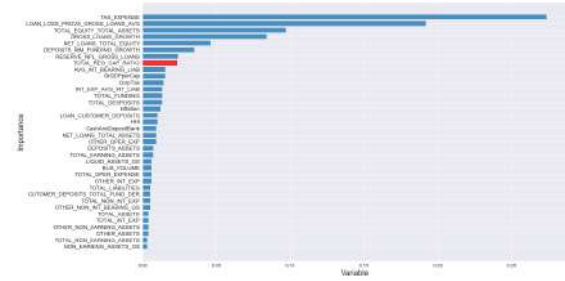
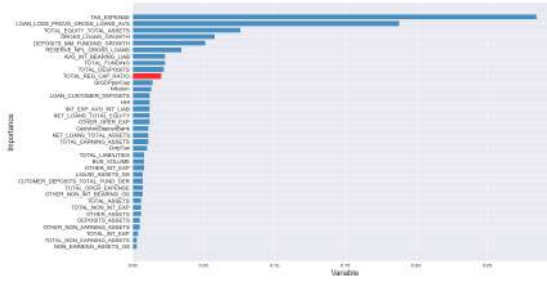
Sample	Models							
	ROAE		ROAE		OPTA		NIM	
	RF	OLS	RF	OLS	RF	OLS	RF	OLS
In-sample	0.78	0.1	0.71	0.08	0.77	0.1	0.93	0.45
Out-of-sample	0.48	0.06	0.04	0.05	0.38	0.06	0.76	-0.43

Source: Author’s calculations. The table shows the coefficient of determination ( $R^2$ ) scores.

Figure 16 – Variables’ Importance over 50 RF

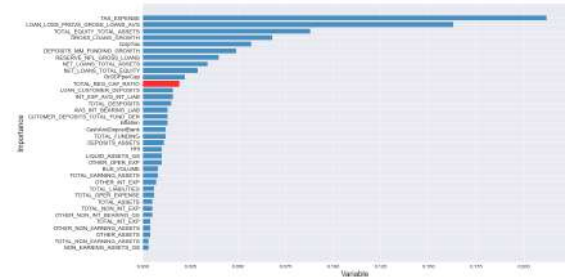
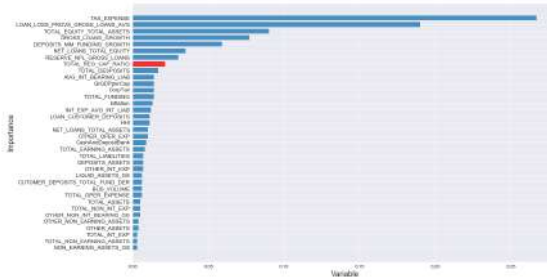
(a) ROAA

(b) ROAE



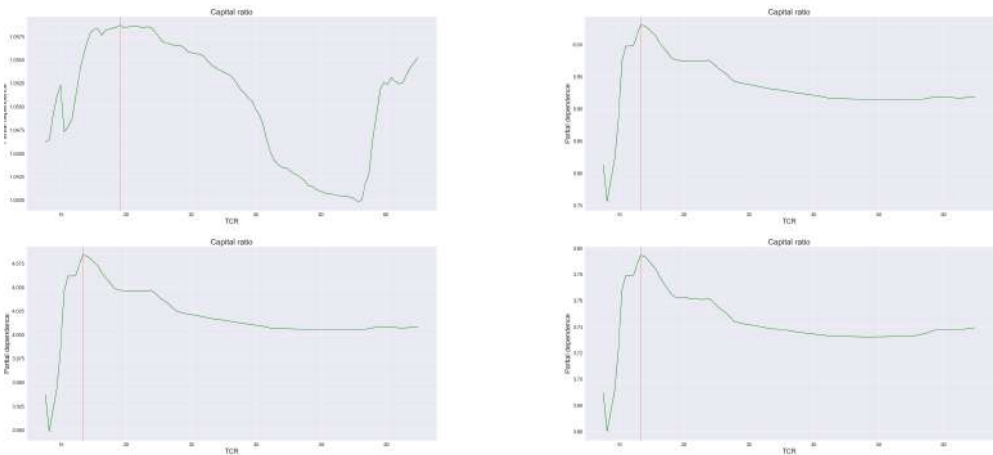
(c) OPTA

(d) NIM



Source: Author’s calculations from FitchConnect data. In red the two regulatory ratios.

Figure 17 – TCR one way PDP over 50 RF over large sample - ROAA, ROAE, OPTA, NIM



Source: Author’s calculations from FitchConnect data. In red the level of the optimum TCR.