Faculty of Economics and Business, Amsterdam School of Economics
Thesis MSc Econometrics: Financial Econometrics

Trading behaviour of European banking groups

Influence of capital changes analyzed econometrically with machine learning

Puja Chandrikasingh

Student number:	11059842
Date of final version:	August 15, 2019
Supervisor:	Dr. M. J. van der Leij
Second reader:	Prof. dr. I. P. P. van Lelyveld

University of Amsterdam

DeNederlandscheBank
EUROSYSTEEM

Statement of Originality

This document is written by Puja Chandrikasingh who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it. The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

The views expressed in this paper are those of the author, and not necessarily those of De Nederlandsche Bank or its committees. All errors are mine.

Abstract

This thesis considers to what extent a bank adjusts its holdings when its capital changes. In order to analyze this, three hypotheses are tested: the *flight-to-safety* hypothesis, the *moral suasion* hypothesis and the *regulatory-arbitrage* hypothesis. Furthermore, the effects are estimated by an econometric regression and two machine learning methods, namely a Neural Network and XGBoost. The machine learning methods are compared to the econometric regression by constructing OLS-like coefficients through the Shapley framework and using a novel framework based on numerical differentiation. Our results are in favour of the flight-to-safety hypothesis, which means that a bank shifts to safer securities when its capital decreases. For the other two hypotheses we consistently find insignificant effects in all our models. Hence, when the capital of a bank decreases, it does not change domestic securities differently from foreign securities, nor government securities differently from non-government securities.

Contents

Chapter 1

Introduction

Ever since the financial crisis of 2008, an increasing amount of attention has been paid to banks. As the main financial intermediary, the banking sector plays a central role in the economy and is crucial for the stability of financial systems (De Jonghe, 2010). However, as banks shift their activities from interest income to non-traditional banking activities, this stability is under pressure. A solution to this erosion of the financial stability could be more regulation on the banking sector.

A popular class of regulations on this matter is the tightening of the requirements on bank capital. This is studied by Martynova (2015), who has provided a survey on the effect of bank capital requirements on economic growth. She argues that there is still a lot of uncertainty in the effectiveness of capital requirements, as there is no consensus that more stringent capital requirements reduce banks' risk-taking. Besides the risk-component, Drechsler et al. (2016) argue that weakly capitalized banks in distressed countries might be pressured to buy their home countries sovereign debt.

Hence, this thesis expands on this matter by studying how banks adjust their trading behaviour when their capital changes. Besides analyzing the risk component, measured by the height of the yield to maturity, we also take a look into shifts in foreign or domestic securities and government and non-government securities. We propose three hypotheses concerning the effect of these variables on the holdings of a security, namely the *flight-to-safety* hypothesis, the *moral suasion* hypothesis and the *regulatory arbitrage* hypothesis.

When the capital of a bank decreases, the first hypothesis suggests that the bank will balance towards safer securities, which we define as securities with a lower yield. The second hypothesis suggests that in the case of a decrease in the capital of a bank, the holdings in domestic government securities increases, while the last hypothesis suggests that the holdings in high yield government securities increases in this case.

Van Lelyveld et al. (2019) preform a similar analysis in their upcoming paper. However, whereas they use the Security Holdings Statistics (SHS) aggregated at sector level, we use the SHS data at bank level. The first provides a general overview, while the latter says more about how individual banks act. Moreover, in this thesis we use two machine learning methods, Neural

Networks and XGBoost, in combination with an econometric model, while Van Lelyveld et al. (2019) only apply the latter.

In order to compare the results of the three different estimation techniques, we use a Shapley framework to construct coefficients for the machine learning methods that can be interpreted in the same manner as the econometric regression coefficients. Moreover, we introduce a novel method to determine the signs of the coefficients from the Shapley framework. The proposed method can also estimate coefficients for machine learning methods, hence it could essentially replace Shapley framework. Hence, this thesis contributes not only to understanding the trading behaviour of banks, but also to the literature that addresses the issue of interpretability in machine learning models.

Using the three models, we have found some evidence in favour of the first hypothesis of the phenomena of flight-to-safety, which means that when the capital of a bank decreases it rebalances towards safer securities, i.e. securities with a lower yield-to-maturity. We do not find significant evidence in favour nor against the moral suasion and regulatory arbitrage hypotheses, as the results of the different models are insignificant and in contradiction with each other. Hence, the way a RGB adjusts its holdings for both domestic and foreign securities, does not differ significantly. Furthermore, we conclude that the way a RGB adjusts its holdings for both government and non-government securities also does not differ significantly.

The remainder of this thesis is organized as follows. Chapter 2 provides a literature review. Thereafter, the data is described in Chapter 3. Chapter 4 explains the methodology. Chapter 5 provides a summary of the obtained results. Chapter 6 considers the issue of endogeneity. Finally, Chapter 7 presents the conclusion and discussion.

Chapter 2

Theoretical Framework

This chapter elaborates on the economic theory regarding the motives a bank has to invest in a specific security. This results in the three hypotheses we are trying to confirm or disprove in this thesis. Lastly, the findings of related work are presented.

2.1 Economic theory

In order to analyze to what extent a bank adjusts its trading behaviour when its capital changes, we first need to identify the motives a bank has to invest in a specific security. Hence, the first part of this section elaborates on possible motives. Subsequently, the focus shifts to the channels through which a change in capital could influence the trading behaviour of a bank. This results in three hypotheses, namely the *flight to safety* hypothesis, the *moral suasion* hypothesis and the *regulatory arbitrage* hypothesis.

2.1.1 Motives to invest in a specific security

This section discusses the reasons that might motivate a bank to hold a specific security. For this purpose, the capital of the bank, the riskiness of the security and whether the security is issued in the same country as the bank is established, are of particular importance.

First, the capital of the bank is considered. This variable does not influence the decision of a bank to hold a security directly, but it does so indirectly through the *doom loop*, which is a feedback loop between sovereign and banking insolvency. Farhi and Tirole (2017) explain the doom loop as follows. They state that banks hold on to sovereign bonds until they need the resources to invest in other projects or cover other positions. In case of an economic crisis, the fall of the price of public debt effects the net worth of banks negatively, which in turn leads to a negative impact on the bank's capital. This might lead to the bank requiring a bailout, which in turn leads to the sovereign of the bank financing the bailout by issuing additional debt. As this entails an increase in the total stock of the debt, the issuing of the debt decreases the market price of debt, which hurts the banks balance sheet even further. This could, for example, lead to selling securities to increase capital. On the other end, it could lead to buying more (risky) securities in a gamble to earn high profits.

Acharya et al. (2014) analyze this doom loop using bank- and country-level data for the period of 2007 to 2011. Their findings are twofold. Firstly, their findings suggest that a bailout announcement is already enough to lower the market price of government bonds, as entities expect bond dilution from further debt issuance. Secondly, they have found that nations, which have a banking sector that is relatively more distressed, spend more money on the recapitalization of the banking sector. Thus, the bailout of banks transfers the risk of default from the banking sector to the government, which in turn triggers the increase in sovereign credit risk. This in turn can lead to the weakening of the financial sector by decreasing the value of the sovereign's bond holdings and government guarantees, as the sovereign has less room to increase taxes to finance possible additional bailouts.

Second, the riskiness of a security, on its own and in combination with the banks capital, plays a role in the decision to invest in a certain security. Basel III requires banks to hold a specific percentage of their outstanding assets in cash. However, this is only required for the risk-weighted assets (RWA), which are all assets held by a bank weighted by their credit risk, as not all assets carry the same risk. Hence, due to this prudential regulations, a bank might be interested in investing in government securities, since these securities have a zero risk-weight and hence do not contribute to the RWA (Acharya and Steffen, 2015).

Third, besides classifying a security into government and non-government, one could also make a distinction between high- and low-yield securities as a measurement of riskiness. Acharya and Steffen (2015) state that banks show a degree of carry trade behaviour, as they have found that banks invested in deteriorating but high yield GIIPS [1] bonds and financed parts of these purchases with proceeds from selling lower yield sovereign debt. They did this until the returns of the deteriorating GIIPS bonds adversely affected the banks balance sheet. This resulted in a flight to safety, which meant that the banks invested more into longer-term core European government bonds.

In combination with the banks capital, they find that for large banks and banks with low capital ratios and high risk-weighted assets, the carry trade behaviour is stronger. Moreover, if they restrict their analysis to GIIPS banks, they find clear evidence for home bias and *moral suasion* by domestic sovereigns. The main reason for the moral suasion is the fact that sovereigns can pressure banks operating in their territory to buy more of the sovereign's debt, which is needed to make up for the weak demand and reduce the sovereign bond yields. Banks that are in distress, are reliant on regulatory forbearance, hence they are probably more susceptible to political pressures.

Besides the home bias implied by the moral suasion, home bias could also occur when banks are increasing their domestic holdings for the reason of being less able to evaluate foreign securities and hence viewing them as riskier assets (Giannetti and Laeven, 2012). Moreover, Giannetti and Laeven (2012) find that the home bias increases in the presence of adverse

[1] GIIPS stand for the peripheral countries Greece, Italy, Ireland, Portugal and Spain.

economic shocks affecting the net wealth of international lenders, at least in the international allocation of syndicated loans.

Fourth, the bank's ownership structure could also influence its motives. This, however, is in contradiction with the Modigliani-Miller theorem, which implies that the volume and structure of a bank's activities do not depend on the funding structure of the banks themselves (Miller, 1995 in Buch and Prieto, 2014). However, Laeven and Levine (2009), on the other hand, argue that the impact of capital regulations on risk taken on by banks depends on each bank's ownership structure. In their paper, they have found that banks take on greater risks if the owners have a more powerful position in determining the decision making process, as the owners would like to maximize the profits. Hence, the above discussed effects can differ per bank. This is especially the case for the effects of the riskiness of a security.

Concluding, the motives to invest in a specific security are the capital of the bank itself, the riskiness of the security and whether the security is issued in the same country as the bank is established. The riskiness of the security involves the yield-to-maturity and the sector that issues the security, as government securities are in general safer than corporate securities due to bankruptcy. Of course, there are more variables that influence the decision of a bank whether to invest in a certain security. However, as we aim to analyze the effects of capital instead of constructing the best prediction, we only need the discussed variables as those interact with the amount of capital while other variables do not.

2.1.2 Effect of a change in capital

This section expands on the effect that a change in capital has on the holdings of a bank. The same variables that are part of the motives for a bank to invest in a security also seem to influence the effect of a change in capital.

As stated earlier, a change in bank capital can follow from either an external event, such as a crisis, or an internal event, such as new capital requirements. A bank can respond to these changes in different ways. The most important ways are summarized by Martynova (2015), who has provided a survey on the consequences of more stringent capital requirements. A summary of this survey follows.

Firstly, higher capital requirements can lead to banks cutting their lending. This can be done in combination with a flight to quality, which means that the credit supply to the riskiest borrowers is decreased most. One expects that the same flight to safety is taken in the context of security holdings.

Secondly, higher capital requirements can lead to banks having to increase their equity. This leads to a decline in lending and an increase in the riskiness of outstanding bank loans, which in turn may lead to a reduction in its financial stability. The explanation for this is as follows. The lending declines, as a bank will raise its lending rates in order to keep the same return on equity (ROE), which would otherwise decrease due to higher capital requirements (King, 2010). Increasing the lending rates in turn will lead to lower quality borrowers, who are willing to pay a higher price for a loan. Naturally, this will lead to an increase in bank loan risk (Stiglitz and Weiss, 1981).

Thirdly, higher capital requirements can lead to either a reduction in systematic risk-taking, as there is less money to take on risks with (Martinez-Miera and Suarez, 2014), or an increase in risk-taking, as the shareholder value decreases due to reduced future profits. However, as Martynova (2015) concludes that most empirical evidence suggest that higher bank capital leads to lower riskiness of bank assets, we expect to find similar results in this thesis.

At last, we consider the effects of capital changes on the kind of risk a bank takes on, besides the high- and low-yield classification discussed above. This is analyzed by studying the effect of capital changes on buying domestic or foreign securities and governmental and non-governmental securities. Drechsler et al. (2016) have shown that weakly capitalized banks in distressed countries purchase more home country distressed-sovereign debt. They argue that those banks are most reliant on regulatory approval, which makes them susceptible to pressure from the government to buy domestic sovereign debt to help fund the government. Thus, it is expected that when the capital decreases, a bank purchases more domestic governmental securities. This is especially the case when the country is in distress.

Moreover, the benefit that governmental bonds have over non-governmental bonds, i.e. zero risk-weight, might become more valuable when the capital of a bank decreases. Acharya and Steffen (2015) measure the capital with the Tier 1 capital and Tier 1 capital ratio, which are respectively the bank's core capital[2] and the Tier 1 capital divided by the RWA. They argue that a bank with low Tier 1 capital ratio shift its portfolios into the highest-yielding governmental assets, as those have zero risk weights. A bank does this in an attempt to increase its short-term return and meet the regulatory capital requirements without having to issue economic capital (regulatory capital arbitrage).

[2]The core capital includes equity capital (inclusive of instruments that cannot be redeemed at the option of the holder) and disclosed reserves. Hence, Tier 1 capital is essentially the most perfect form of a bank's capital.

2.1.3 Hypotheses

From the theory described above, we define the following hypotheses in order to determine to what extent a bank adjusts its holdings when its own capital changes. From the discussion above we know that the opposite of the stated hypotheses is also possible. Moreover, it could be that there is not an effect at all.

H1. Flight-to-safety

A bank rebalances its holdings towards safer assets, hence with lower yield, in order to avoid failure, when its capital decreases.

H2. Moral suasion

A bank rebalances its holdings towards domestic sovereign securities when its capital decreases.

H3. Regulatory arbitrage/economize on capital requirements

A bank rebalances by purchasing riskier sovereign debt, while selling higher risk weighted assets (which might be fundamentally safer, such as foreign corporate bonds), when its capital decreases.

The hypotheses do not exclude each other. For example, the moral suasion argument combined with flight-to-safety argument states that the shift into government securities should be more pronounced for domestic government securities.

2.2 Related work

In this section, the existing literature on the topic is considered. There have been several studies regarding the effects of capital on lending and the behaviour of bank trading in economic downturns. One of these studies is done by Abbassi et al. (2016), who found that during crises banks with higher trading expertise increase their investments in securities. They found that this is especially the case when the securities are accompanied with a large price drop. In order to possibly profit from these trading opportunities, they withdraw funds from lending.

Moreover, Timmer (2018) found that banks buy securities when their returns have been high and sell securities when their returns have been low. This pro-cyclical investment behaviour is found to be more present in banks, which are relatively less capitalized. Furthermore, Drechsler et al. (2016) have shown that low capitalized banks have more lender of last resort (LOLR) loans[3] and riskier collateral, as they buy more risky assets, such as distressed sovereign debt.

Lastly, this thesis is most similar to the upcoming paper of Van Lelyveld et al. (2019), as they investigate the effects of distress on the asset choices of EU banking systems. At the moment of writing, the paper of Van Lelyveld et al. (2019) is still classified, hence we cannot elaborate on their findings. There are a few key differences between their paper and this thesis. Firstly, an

[3]Banks can get LOLR loans from the central bank as last resort before bankruptcy.

important characteristic of their paper is that they look at the total banking sector on country level. This might, however, lead to the effects found in each individual bank cancelling each other out in the calculation of the total effect, making the aggregated effect zero. Hence, this thesis will focus on bank specific effects to avoid the possibility of this problem. Secondly, they use an econometric model to estimate the effects, while in this thesis two machine learning techniques are used besides the econometric model. The benefit of machine learning methods is that they can capture a non-linear underlying process. Moreover, there are less to no assumptions that have to be satisfied in order to use them properly.

Overall, the papers, besides Van Lelyveld et al. (2019), focused on explaining the trading behaviour of banks based on variables other than a bank's capital and found that that the effect of that variable changed when the capital changed. This thesis elaborates on the effect of a change in capital on the holdings of a bank using both econometric and machine learning models. Before discussing the methodology, however, we analyze the characteristics of the used data.

Chapter 3

Data

This chapter is made up of two sections. Firstly, the origin of the data is described. Secondly, basic characteristics of the data are presented in the form of descriptive statistics.

3.1 Data description

The data set, used in this thesis, consists of two parts. The first part is made up of the ESCB Securities Holdings Statistics on individual bank level (SHS-G), while the second part consists of intra-bank data, retrieved from the Statistical Data Warehouse of the European Central Bank. The SHS-G data set is available from the fourth quarter of 2013 (2013Q4) and contains information on the holdings of the 24-26[1] largest banking groups in the Euro area, referred to as the Reporting Banking Groups (RBGs).[2] As from 2018Q2 on, the number of reporting entities and the reported quantities have changed, hence we use quarterly data from 2013Q4 up to 2018Q1. The data contains information on two types of securities: debt and equity. Both Timmer (2018) and Van Lelyveld et al. (2019) restrict their analysis to debt securities. Similar to the latter, this thesis focuses on long-term debt securities. The reason for choosing long-term instead of short-term debt, is the fact that short-term debt might be to volatile to use, as with using short-term debt the maturity is often contained in the interval between the current quarter and the subsequent quarter. This is problematic, as the position in the short-term bond will appear to be fully closed when the bond matures. As this change in the position is not the result of changing capital, this will create a spurious effect, making the analysis of our main variable extremely difficult.

An important feature of the used data, is the holdings information, which is collected at the individual security level. Each security can be identified by its International Securities Identification Number (ISIN). Besides the ISIN, the data also contains information on the holding period, price, issuing sector, issuing country and other metrics. Moreover, for each ISIN the total outstanding market value is given as the market value a bank holds (number of securities held multiplied by the price). From this, we define the target variable in levels

[1]The number of RBGs differ per quarter.

[2]For a detailed description of the dataset one can consult ECB (2015) or `http://tiny.cc/hget5y`

$Holdings$.

$$Holdings_{itr} = \frac{\text{market value held by RBG } r \text{ of ISIN } i \text{ at time } t}{\text{total outstanding market value of ISIN } i \text{ at time } t} * 100$$

However, from an economic point of view we are interested in the change in holdings, $\Delta Holdings$.

In order to assess the capitalisation of a certain bank, we use the Tier 1 capital ratio, as is done by Acharya and Steffen (2015) and Van Lelyveld et al. (2019). The Tier 1 ratio is defined as the market value of the total equity divided by the market value of the total risk weighted assets. The data is obtained from the Statistical Data Warehouse and is incomplete. In order to still be able use these measures in analysis, the missing values are interpolated using a simple unweighted linear scheme.

For the ISINs there are several variables of importance, namely whether the ISIN is governmental, high-yield and/or is issued in the same country as a specific RBG is established in. These variables are all captured by indicator variables. $Domestic_{ir}$ equals one if security i is issued in the same country as RBG r is established in and zero otherwise. When ISIN i is issued by a government, $Government_i$ equals one and zero otherwise. To indicate whether a security has a high yield, we use two indicator variables, namely $HighYieldTotal_{it}$ and $HighYieldSector_{it}$. The first equals one if the yield to maturity of security i is above the median in quarter t and zero otherwise. The latter equals one if the yield to maturity of security i issued by sector s is higher than the median yield to maturity of all the securities issued by the same sector s in quarter t and zero otherwise.

We define two measures of high-yield as one takes the whole market in consideration, while the other only the sector. The latter is for example important for the regulatory arbitrage hypotheses, as we have to identify high-yield government securities, while government securities are probably all labelled as low-yield if we look at all ISINs. However, not all government securities have equal risk. Hence, we should also have a measurement of high-yield within the sector government. Furthermore, it could be the case that a specific ISIN becomes low-yield compared to all other ISINs, but stays high-yield in its own sector. However, if we find that including both measurements results in insignificant results, while for both holds that including that specific variable is significant, we include an interaction of both terms.

3.2 Summary statistics

Following the cleaning of the data set (Appendix A), a total of around 750 thousand observations and slightly less than 64 thousand unique long-term debt securities remain. On average, there are around 42 thousand observations and 20 thousand unique long-term debt securities per quarter. The holdings are concentrated in a limited set of securities. This is illustrated by the fact that the 1000 largest holdings of the long-term debt securities in each quarter account for 54% of the total holding values, while being only 5% of the total number of securities.

Overall, the holdings of long-term debt securities have been stable in the last years amounting to an average of 1147 billion (with standard deviation 84 billion). From Table 3.1, we can see that the banks on average hold more foreign securities than domestic ones. Note, however, that the average amount invested in a foreign bond is less than in a domestic bond. Additionally, Table 3.1 shows that the distribution between the issuers sectors remains stable, with the largest difference being in government and Monetary Financial Institutions (MFI) securities. Moreover, the percentage of government securities is the second lowest, but in each government bond is on average the highest amount invested compared to securities of other sectors. Furthermore, the standard deviation is by far the largest for the amount invested in government securities, which could be (partially) due to change in the number of securities being held. Furthermore, around 62% of the securities are low-yield compared to all other securities in that quarter. Banks invest a greater amount per low-yield bond than in a high-yield bond on average.

Table 3.1: Summary of kind of ISINs that are being held

	Percentage		Amount (in millions)	
	Average	Standard deviation	Average	Standard deviation
Country				
Foreign	63.84	0.73	21.61	3.21
Domestic	36.16	0.73	36.05	9.54
Issuer sector				
MFI	43.73	2.35	16.33	1.20
OFI	27.25	0.91	17.89	1.78
GOV	17.30	3.29	77.46	21.83
NFC	11.72	1.66	5.71	0.65
Yield[3]				
High-Yield	38.13	2.66	16.81	6.11
Low-Yield	61.87	2.66	35.95	6.35

Percentage: per subcategory of *country*, *sector* and *yield* the percentage invested in each subcategory w.r.t. the total.
Amount: per subcategory of *country*, *sector* and *yield* the absolute amount invested in each subcategory.
For both the average over time, ISINs and RBGs is given first, then the standard deviation.
Country: ISINs that are issued outside (inside) the country the RBG is established in is denoted as *foreign* (*domestic*).
Issuer sector: Monetary financial institutions (Banks) (*MFI*), Other financial institution (*OFI*), (General) Government (*GOV*) or Non-financial corporations (*NFC*)
Yield: ISINs with a yield to maturity that is higher (lower) than the median in a quarter are classified as *High-Yield* (*Low-Yield*).

Table 3.2 contains the summary statistics of the other main variables. The changes are quite small, which might be problematic. Overall, the change in holdings differ as expected from plus hundred to minus hundred. However, on average the change is small and considering all the data even the absolute value of the 1st and 99th percentile are below 20 percentage point change. For the 5th and 95th percentile, this would be below 3 percentage point change. The

[3]Note that from the unique ISINs in one quarter exactly 50% would be low-yield and the other 50% would be high-yield, as we compare it with the median. However, this table shows the average over time and over RBGs, which results in different percentages, as the same ISIN can be held by multiple RBGs.

relatively small observations with a high value for $\Delta Holdings$, could make it hard to predict correctly. However, the goal of the thesis is to quantify the effect of a capital change on the holdings and not to predict the holdings as well as possible.

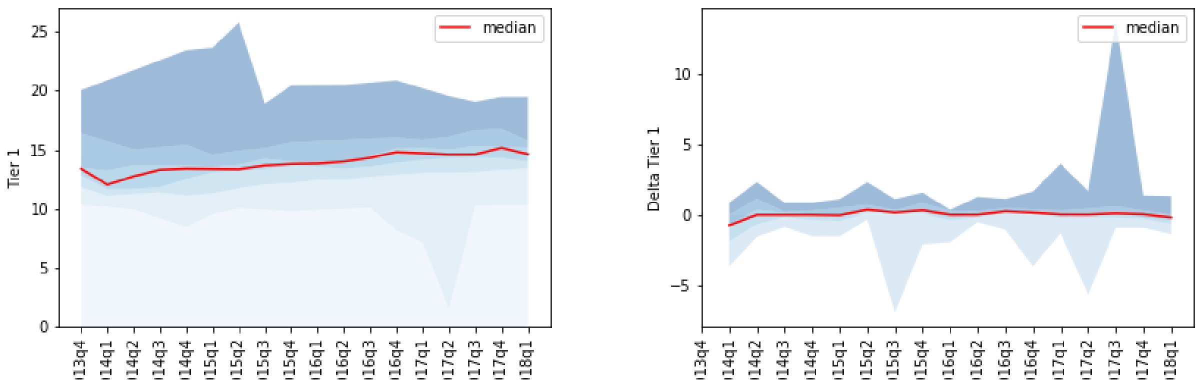
Table 3.2: Descriptives main variables

	2013Q4			2018Q1			Total			
	Mean	Med	SD	Mean	Med	SD	Mean	Med	SD	Obs
Tier1	13.5	12.5	1.8	14.8	14.8	1.7	13.8	13.4	1.8	126
Δ Tier1 %	.	.	.	-.3	-.4	.2	.1	.1	.6	119
Δ Holdings %	.	.	.	0	0	.2	0	.1	.5	119

Tier1: Tier 1 capital ratio

For 2 quarters and the total data set the mean, median and standard deviation is given for the Tier 1 capital ratio, the change in the Tier 1 capital ratio and the change in the holdings. For the total data set the minimum, maximum and number of observations ($\times 10^9$) is also given.

Figure 3.1 contains percentile timeplots for the capitalisation measure. Note that one line does not represent one bank, as these are percentile lines. Hence, the Tier 1 of different banks can form that specific percentile in different periods. We can see that overall the level of Tier 1 stays the same. Moreover, we see that the changes in Tier 1 are indeed quite small in general.



(a) Tier 1 (b) Change in Tier 1

Figure 3.1: Capital over time

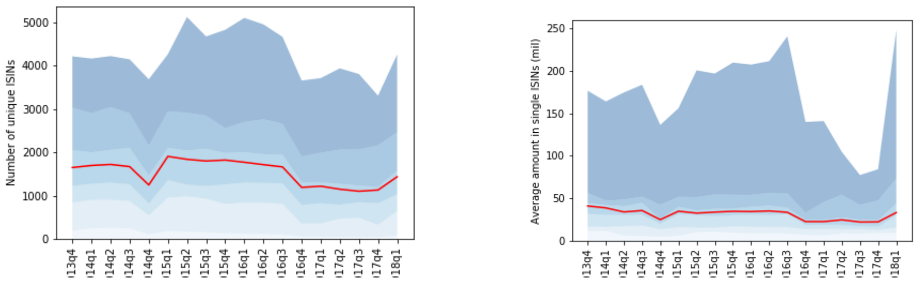
Table 3.3 summarizes the data for the Tier 1 ratio. Over the sample as a whole this indicator of bank capitalization ranges from 1.50 to 25.80 percent with an average of 13.82 percent. The measure is the lowest on average in 2014Q1 with a value of 12.88, and the highest on average in 2017Q4 with a value of 14.89. Note that a bank should have a Tier 1 capital ratio that is at least 8% (LLC, 2019). In the first and second quarter of 2017, there are values below 8. This is not an error in the data, as it belongs to a bank that was near collapse and was bailed out by its government.

Table 3.3: Extensive summary statistics | Tier 1 capital ratio

Date	mean	std	min	25%	50%	75%	max
2013 Q4	13.83	2.36	10.30	11.84	13.40	16.39	20.06
2014 Q1	12.88	2.14	10.18	11.30	12.04	13.85	20.90
2014 Q2	13.10	1.98	9.95	11.29	12.74	15.03	21.75
2014 Q3	13.09	1.96	9.20	11.44	13.30	14.35	22.59
2014 Q4	13.10	2.00	8.45	11.46	13.40	14.31	23.44
2015 Q1	13.02	1.95	9.55	11.31	13.37	14.16	23.65
2015 Q2	13.42	1.94	10.05	11.74	13.35	14.65	25.80
2015 Q3	13.62	1.87	9.92	12.08	13.67	15.03	18.92
2015 Q4	13.92	1.81	9.80	12.35	13.79	15.49	20.46
2016 Q1	13.73	1.75	9.90	12.55	13.84	14.90	20.48
2016 Q2	13.81	1.73	10.00	12.45	14.00	14.91	20.49
2016 Q3	14.02	1.70	10.10	12.72	14.32	15.13	20.69
2016 Q4	14.16	2.22	8.17	12.87	14.80	15.58	20.89
2017 Q1	14.33	1.99	7.16	13.05	14.70	15.83	20.22
2017 Q2	14.36	2.38	1.50	13.07	14.60	15.63	19.56
2017 Q3	14.81	1.88	10.28	13.41	14.61	16.30	19.08
2017 Q4	14.89	1.78	10.30	13.97	15.16	16.42	19.48
2018 Q1	14.58	1.78	10.30	13.59	14.63	15.76	19.48
mean	13.82	1.96	9.17	12.36	13.87	15.21	21.00

Per quarter the mean and standard deviation is given for the Tier 1 capital ratio. Moreover, the minimum value, the 25th percentile, the median (50th percentile), the 75th percentile and the maximum value are given. The last row gives the mean over time.

Furthermore, there also exists a difference between the RBGs. Figure 3.2 shows percentile graphs in which the different RBGs are ordered by the variable on the y-axis. The first graph shows the number of unique ISINs banking groups possess each quarter. The top 20% owns around 2.5 times as much as the bottom 60%. Additionally, the relative changes are larger for the RBGs that belong to the top 20%. From the second graph, a similar conclusion can be drawn for the average amount invested in a single long-term debt security. Here, the top 20 % invest around 4 times more in a single security than the bottom 80%, where the changes are again larger for the top.



(a) Number of unique ISINs (b) Average amount per ISIN

Figure 3.2: Holdings over time

Moreover, from the data we can see that better capitalised banking systems hold fewer government bonds, fewer domestic bonds and less risky bonds. However, if we also look at the changes of holdings and include holder area fixed effects – as in Table 3.4 – we can see that capitalisation is not related to the holdings in different kinds of securities.

Table 3.4: Effect of changes in capital on holdings per security type

	(1)	(2)	(3)
VARIABLES	$\Delta GovHolding$	$\Delta DomesticHolding$	$\Delta HighYieldHolding$
$\Delta Tier1$	0.6116	-0.4618	0.5283
	(2.0050)	(1.2626)	(1.2200)
Observations	119	119	119
R-squared	0.0005	0.0009	0.0038
Number of holder_area_enc	7	7	7
Holder area FE	Yes	Yes	Yes

Note: (Robust) standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. $\Delta Tier1$: is change in the Tier 1 capital ratio. Panel regression with holder area fixed effects predicting the change in the holdings of government securities, domestic securities and high-yield securities in column 1, 2 and 3 respectively.

The presented analysis is simple and can be influenced by other factors. Hence, in the next section, a security-by-security based analysis is described to test whether capitalisation influences the holdings of a RBG.

Chapter 4

Methodology

The methodology used in this thesis is based on the work of Bluwstein et al. (2019) and Joseph (2019). Firstly, in order to determine to what extent a bank adjusts its security holdings when its capital changes, a benchmark econometric procedure is performed. Subsequently, two machine learning techniques, which allow for a more flexible and non-linear specification of the model, are used to fit the data. These models are of particular interest, as they can approximate any well-behaved function arbitrary close if given enough training data (Bluwstein et al., 2019).

This section, firstly, discusses the division of the data and addresses a general problem with the data. Subsequently, the variables regarding the three hypotheses are discussed. Thereafter, the model methodology is elaborated on. Subsequently, we present the Shapley framework, which is used to construct OLS like coefficients, named SSC, for the machine learning methods in order to compare the estimated effects of the different methods. Lastly, we argue that the sign of the SSC values is constructed in an non-informative way and we propose a new method. This novel method could also be used to construct OLS like coefficients in a different way than in the Shapley framework.

First, we discuss the data division. All models are trained or estimated on the training set. In the econometric setting, decisions regarding including variables are based on tests using the estimated model, hence on the training set. However, in the machine learning setting the model construction is based on the training and validation set, as the hyper-parameters are tuned based on the performance of a trained model on a validation set.

Besides the training and validation set, we also have a test set. The latter is used to compare the performance of the different models with each other. We randomly assign the data per quarter to one of the three sets. With 70% chance a data point is assigned to the train set, while the probability to go to the validation or test sets both are 15%. The median that is used in the indicator variables for identifying high-yield securities, is determined per data set in order to avoid information leakage.

Second, we discuss some important variables and a problem with the dependent variable. In this framework, the variable of interest is the change in a bank's holdings. The holding of ISIN i by RGB r at time t is denoted as $Holdings_{irt}$. From the literature, one can conclude that the holdings can be influenced by the level of Tier 1, $Tier1_{rt}$, through the earlier mentioned doom loop theory. Additionally, due to prudential regulations one expects both the high yield measurements and the difference between government and non-government bonds, $Government_i$, are likely to play a role in the analysis. As mentioned in Chapter 3 we use two high yield dummies, where the one w.r.t. all other ISINs at time t is denoted as $HighYieldTotal_{it}$ and the one w.r.t. all other ISINs in the same sector is denoted as $HighYieldSector_{it}$. Furthermore, it is also known that home bias is present, despite banks might try to keep the efficient market portfolio. Hence, we should take this into account by including a dummy for domestic bonds, $Domestic_{ir}$, as a control variable.

A problem that all models have to overcome, is the problem of a relatively small number of observations with a high value for $|\Delta Holdings|$, as described in Chapter 3. One solution is to estimate the models on a subset of the data that contains only the observations for *total outstanding market value of ISIN i at time t > 5th percentile of all outstanding market values.* This likely excludes the high values of $|\Delta Holdings|$, as a RBG is unlikely to hold a high percentage of a ISIN with a large outstanding market value. Hence, besides regressing/training on the full train set, we also regress/train on a subset of it.

Third, we elaborate on the hypotheses. Given the formulated hypotheses, we expect the following coefficient signs. The flight-to-safety hypothesis suggests that the coefficient of $\Delta HighYield \times \Delta Tier1$ is positive, as in this case a decrease in capital will cause an increase in the holdings of the security for which the yield went down. Hence, when the capital decreases the holding in securities with a decreased risk increases. From the second hypothesis, which concerns moral suasion, we expect a negative coefficient of $Domestic \times \Delta Tier1$, as in this case a decrease in capital causes an increase in the holdings of domestic bonds. The last hypothesis, regulatory arbitrage, concerns the holdings of government securities. It implies that the coefficients of $Government \times \Delta Tier1$ and $Government \times \Delta HighYield \times \Delta Tier1$ are negative, as in this case a decrease in capital causes an increase in the holdings of government securities and high-yield government securities respectively.

The remainder of this chapter is composed as follows. Firstly, the methodology regarding the econometric regressions is discussed. Secondly, the methodology concerning the machine learning techniques, namely Neural Networks and XGBoost, is presented. Lastly, the Shapley framework, which is used to interpret the machine learning methods, is described.

4.1 Econometric regression

This section describes the structure of the econometric model. In order to investigate to what extent a bank adjusts its holdings when its capital changes, one can use a panel data framework, as the SHS-G data set is multidimensional data containing information per ISIN, RBG and quarter. The data consists of 18 time periods and has on average around 42 thousand cross-sectional units per time period, which are given by the number of unique RBG and ISIN combinations in a specific quarter.

One advantage of the panel data framework is the increase in the precision of the estimates, as one has more observations that can be used, as we have both a timeseries dimension and a cross-section dimension. Moreover, the framework also provides more control over time-invariant omitted variables, since panel data can be used to control for time invariant unobserved heterogeneity. This framework is therefore widely used for causality research. A limitation of panel data, however, is that time varying omitted variables may still be present. Nonetheless, the overall omitted variable bias is still smaller than with regular cross sectional data.

As previously described, the econometric model should include $Tier1_{rt}$, $Government_i$, $HighYieldTotal_{it}$, $HighYieldSector_{it}$ and $Domestic_{ir}$. The different interactions between these four variables are also meaningful and are denoted by *interactions*. Which interaction terms to include, is determined by minimisation of the the BIC and AIC indices and F-tests for joint significance.

Besides these variables and their interaction terms, there could also be a difference per RGB, ISIN and quarter, which are captured with α_r, α_i and α_t respectively. The time and RBG invariant factors that affect the holdings are contained in α_i. Hence, these factors are the ISIN specific effects, which model the impact of omitted time and RBG invariant variables on $Holdings_{irt}$. Likewise, α_r contains the time and ISIN invariant factors and α_t is cross-section invariant, which means it contains the ISIN and RBG invariant factors that change over time. This results in the model displayed in Equation (4.1).

$$\begin{aligned} Holdings_{irt} = {} & \beta_1 Tier1_{rt} + \beta_2 HighYieldTotal_{it} + \beta_3 HighYieldSector_{it} \\ & + \beta_4 Government_i + \beta_5 Domestic_{ir} + interactions'_{irt}\gamma + \alpha_i + \alpha_r + \alpha_t + \varepsilon_{irt} \end{aligned} \tag{4.1}$$

Including the effects of RBG and ISIN separately means that holdings in a specific ISIN i are influenced by the characteristics of that ISIN. Moreover, it also means that a specific RBG r could invest more in all ISINs it holds, compared to other RBGs, regardless of the characteristics of the ISINs. This might be restrictive, as it is quite unlikely that this would hold for all ISINs regardless of their risk. If instead of the separate effects α_i and α_r the RBG-ISIN combination α_{ri} is included, which means that the holding in ISIN i by a certain RGB are not only independent from other RGBs, but are also independent from other ISINs. This could be the case when the RBG knows a lot about each ISIN and adjusts its policy for each ISIN based on that information. However, this might be too costly to do for all ISINs, as transactions costs play a role. Hence, this option might be improbable.

There are three methods for estimating a model with fixed effects.[1] The first is to model the fixed effects directly using LSDV estimation. Another way is to use the within estimation or first differences estimation. As we are from an economic point of view interested in the change in holdings, $\Delta Holdings$, we use the first differences estimation. A benefit of this method is

[1] We assume that all time invariant effects are fixed effects, as they are likely to be correlated with the other regressors. This is formally tested with a Hausman test.

that we account for a possible unit root in the levels as taking the first difference leads to stationarity.[2]

Taking the first differences means losing the time invariant fixed effects, as it automatically controls for time-invariant unobserved factors. This results in the following equation where the interactions are now taken with respect to $\Delta Tier1$, $\Delta HighYieldTotal$ and $\Delta HighYieldSector$ instead of $Tier1$, $HighYieldTotal$ and $HighYieldSector$. The last step follows from including a new variable which is simply a linear combination of the other two.

$$\begin{aligned}\Delta Holdings_{irt} &= \beta_1 \Delta Tier1_{rt} + \beta_2 \Delta HighYieldTotal_{it} + \beta_3 \Delta HighYieldSector_{it} \\ &+ interactions'_{irt}\delta + \alpha_t - \alpha_{t-1} + \varepsilon_{irt} - \varepsilon_{ir,t-1} \\ &= \beta_1 \Delta Tier1_{rt} + \beta_2 \Delta HighYieldTotal_{it} + \beta_3 \Delta HighYieldSector_{it} \\ &+ interactions'_{irt}\delta + \gamma_t + \epsilon_{irt}\end{aligned}$$

Note that it is also possible to include ISIN or RBG specific fixed effects, denoted by κ_i and κ_r respectively. However, this would mean that the equation in levels has incidental trends: $\kappa_i t$ and $\kappa_r t$. This could be too restrictive, as it is a deterministic trend with heterogeneous loading, which means that each RBG has its own loading but that the loading is fixed over time. We could allow for ISIN-Time and RBG-Time effects that differ per ISIN/RBG and quarter by including $\alpha_{it}(= \alpha_i + \alpha_t + \eta_{it})$ and $\alpha_{rt}(= \alpha_r + \alpha_t + \eta_{rt})$ in Equation (4.1). Furthermore, note that the other fixed effects (including α_{ri}) then have to be excluded. However, if we include the latter, this would imply that we have to exclude $Tier_{rt}$, as failure to exclude this variable will lead to perfect multicollinearity. This is problematic, as we are interested in the effects of a change in capital. Hence, instead of α_{rt} we could also include $\alpha_{HolderArea,t}$, which would control for the effects that are time and holder area[3] specific. A similar problem arises with α_{it} and the high-yield indicators, hence α_{it} should be replaced by $\alpha_{IssuerSector,t}$.

When $\alpha_{IssuerSector,t}$ and $\alpha_{HolderArea,t}$ are included as fixed effects in (4.1), the effects that are time-varying do not drop out if we take first differences. It is reasonable to assume that specific RBGs react different in some quarters. The same holds for the reaction of a RBGs on specific ISINs in specific quarters, due to the different sentiments concerning the ISINs. Hence, there is some reason to believe that $\eta_{IssuerSector,t}$ and $\eta_{HolderArea,t}$ are significant. This implies that $\alpha_{IssuerSector,t}$ and $\alpha_{HolderArea,t}$ have to be included as fixed effect in (4.1). This results in the following equation, where we handle the effects that are time-varying, $\eta_{IssuerSector,t}$ and $\eta_{HolderArea,t}$, in the same manner as the α_t above. This means we include a linear combination of $\eta_{IssuerSector,t}$ and $\eta_{IssuerSector,t-1}$ denoted as $\gamma_{IssuerSector,t}$ and of $\eta_{HolderArea,t}$ and $\eta_{HolderArea,t-1}$ denoted as $\gamma_{HolderArea,t}$. Note, that $\gamma_{IssuerSector,t}$ and $\gamma_{HolderArea,t}$ are not fixed effects anymore

[2]If ε_{irt} is i.i.d. in the level equation, then the first difference equation will have a MA(1) structure. To test for this we use the Breusch-Godfrey test and the Wooldridge test. If the errors in the level equation are i.i.d., then we use GLS estimation technique in the first difference equation, which exploits the MA(1) covariance structure. This makes estimation not only consistent but also efficient.

[3]The holder area is the country in which the head office of the RBG is located.

in the sense that we can estimate them by taking first differences or performing the within transformation. However, they can be seen as control variables and can be included as dummies. This results in a pooled regression given by Equation (4.2).

$$\begin{aligned}\Delta Holdings_{irt} = {} & \beta_1 \Delta Tier1_{rt} + \beta_2 \Delta HighYieldTotal_{it} + \beta_3 \Delta HighYieldSector_{it} + interactions'_{irt}\delta \\ & + \gamma_{IssuerSector,t} + \gamma_{HolderArea,t} + \epsilon_{irt}\end{aligned} \tag{4.2}$$

Lastly, we consider the errors, ϵ_{irt}, which can be both heteroscedastic and serially correlated (clustered). However, as we only have 18 periods, a time-series analysis of the errors might not be meaningful. Hence, we use clustered standard errors, as they are valid whether or not there is heteroskedasticity and/or autocorrelation. Moreover, using non-robust standard errors can lead to an overestimation of the accuracy of the results. Bertrand, Duflo and Mullainathan (2004) illustrate the resulting downward bias in the computation of the standard error in the context of difference-in-differences estimation without the clustered errors. They find that the panel-robust clustered and panel bootstrap methods work relatively well.

The clusters can be taken in multiple levels of the data. It is important that in the chosen depth level, there should be at least 50 clusters, since otherwise the cluster-robust standard errors become inaccurate. This means that we can not cluster on RBG, as there are only 26 of them. If we cluster on the deepest level, given by the RBG-ISIN combinations, there could be arbitrary correlation within the individual time series of each RBG-ISIN combination, but the combinations themselves have to be independent of each other. This could be restrictive, as it might be the case that there is correlation within each issuer sector. Hence, using clusters on RBG-Issuer Sector is chosen. Having a framework for the econometric regression, we now move on to analyzing the machine learning methods, as those methods are able to capture non-linear relationships better and require less assumptions.

4.2 Machine Learning Models

The econometric regression, described above, relies on several assumptions. Moreover, it is linear in the coefficients, as the only non-linearities are artificially created by including interaction terms. On the other hand, machine learning methods rely on far less to no assumptions. Furthermore, they are known for their ability to capture nonlinear relationships. In this thesis two different machine learning methods are used, namely Neural Networks and XGBoost. Of each method, a high level explanation is given in the following. Implementation details are given in Appendix A.2.3 and Appendix A.2.4 respectively. We first discuss the variables that are given to NN and XGBoost, as they differ a bit from the variables used in the econometric regression.

We are still interested in the effect of a change in capital on the holdings in a specific security. For this we designed the first difference model as an econometric model, which caused some variables to drop out. Taking the first differences is not required when using NNs or XGBoost. Moreover, the interaction terms are redundant, as both methods should be able to capture the non-linearities in the underlying process itself. Hence, the input variables which are given to the NN are: $\Delta Tier1$, $\Delta HighYieldTotal$, $\Delta HighYieldSector$, $Government$, $Non-Government$, $Domestic$, $Foreign$, $IssuerSector$, $IssuerCountry$, $HoldingArea$, RBG dummies and $time$ dummies.

The RBGs, dates, issuer sector and country and the holding area are given as dummies so that the weights per subclass can differ. This is desirable, as this for example enables the change of the holding to differ per time period, which is more realistic than a constant time indifferent change. Moreover, indicator variables for domestic and foreign are given to the NN, as we cannot perform any mathematical operation with 0 for which we get a result different from 0. Hence, if we only included one indicator function, the NN could not learn anything for the securities for which the value is 0. This also holds for the indicator variables for government and non-government.

Moreover, the variables indicating the change in high-yield can have three values: down (-1), up (+1) and same (0). They represent respectively the security going from high-yield to low-yield, from low-yield to high-yield or staying the same.As with the other dummy variables, we need to pass these different values in different variables to the NN. This gives the model the possibility of differentiating between the different changes, can in any case lead to a better accuracy. If a difference between the changes is absent, the NN is able to simply combine the variables in a dummy like structure.

Note that the ISINs cannot be given to the network as dummies, as there is not enough memory to store the the dummies for thousands of ISINs.[4] To incorporate some information that is ISIN specific, we include the price of the ISINs, the ISIN encoding, the yield to maturity, the days left to maturity and the original days to maturity. The ISIN encoding is an unique number for each ISIN, as we cannot include the original ISIN names as those contain letters.

Lastly, the variables given to XGBoost are similar to the ones given to the NNs. However, we only include the indicator variable for domestic instead of for both foreign and domestic, as XGBoost can assign a value to the class that is zero. The same holds for government and non-government. Moreover, we do not make three different indicator variables for the three cases of a change in high-yield, as XGBoost can construct such a variable by itself by splitting the change in high-yield variable twice. Hence, the input variables which are given to XGBoost are: $\Delta Tier1$, $\Delta HighYieldTotal$, $\Delta HighYieldSector$, $Government$, $Domestic$, $IssuerSector$, $IssuerCountry$, $HoldingArea$, RBG dummies and $time$ dummies.

[4] In the econometric model, this is solved by using the within transformation instead of including the actual dummies as in LSDV. Initially, we do not demean the data before giving it to the machine learning models, as demeaning the data assumes a constant linear relationship over time per ISIN. This would be restrictive for non-linear models, as they would be able to capture non-linear relationships. However, for the NN we eventually demean the data as otherwise the estimated SSC values are dominated by ISIN specific variables.

4.2.1 Neural Networks

The first considered machine learning method is the (Artificial) Neural Network (NN), as this has been the most researched machine learning technique in recent years (Bluwstein et al., 2019). Moreover, the universal approximation theorem mathematically shows that any continuous function can be approximated to any degree of accuracy with a quite simple NN.[5] This section gives a brief explanation on NNs and specifies the different NNs we compare. The details of the implementations are given in Appendix A.2.3. Naturally, the literature on NN's is vast and for a more detailed explanation in papers relevant to this thesis one can consult either Bluwstein et al. (2019) or Berk (2008).

A Neural Network can be described as a network in which the network applies L functions on the data, one after the other. Hence, the final output $\widehat{\Delta Holdings_{irt}}$ is the output of the last function f^L, which is based on the computations of all previous functions $f^1, ..., f^{L-1}$:

$$\widehat{\Delta Holdings_{irt}} = f^L(f^{L-1}(...f^1(input, w^1)), w^L)$$

where w^j are the weights of the jth function (f^j) and the *input* for the first function are the values of the regressors and for the further functions, the input is the output of the previous function. The functions f are often simple, as they have to be computationally easy to compute and easy to differentiate. The functions could be linear with different weights for the inputs and non-linear such as the *tanh* function. The weights are learned by the network by iteratively minimizing the loss for which the previously mentioned easy differentiation is needed.

The terminology is as follows. The functions are called *activation functions* and the output of such a function, say the nth function, is called the *activation* and is denoted by a_n and could be a vector. Moreover, the number of output variables for each layer can also differ. If there are a total of L activation functions, then the NN is said to have L *layers* of which the last layer is the (final) *output layer* and the other $L - 1$ layers are called *hidden layers*.[6] If a NN uses a specific set of layers multiple times, these layers are called a *module*. Each layer consists of a number of *nodes* or *units* equal to the length of a, which represent the output variables of that layer. Lastly, a NN is called deep when L is large and wide when L is small and the number of nodes per layer is large.

Our last layer is linear with only one output, as our output has no restrictions in terms of permitted values. The number of hidden layers, the number of nodes per layer and the activation functions for each layer are hyperparameters that affect the outcome of the NN but are not optimized during training. Hence, we partition the data into a training, validation and test set and use the validation set to determine the right choices.

The experimental procedure is set up as follows. Firstly, a search grid is defined, which contains the values of the hyperparameters. Subsequently, a specific combination is selected and the resulting NN is trained on the train set. The performance of this NN is evaluated on

[5] The proof is for a two layer feed-forward network, given sufficiently many neurons in the hidden layer.

[6] Some authors do not count the last layer in the L layers. Hence, they would have L hidden layers with thereafter an output layer, which would make a total of L+1 functions that are applied.

the validation set. This is done for all combinations of the hyperparameters. The combination of hyperparameters with the best performance in the validation set is selected. Eventually, the performance of the selected NN is evaluated on the test set. This score is then compared to the scores of the other models.

Normally, the search grid contains three variables: L, which is the number of layers, and per layer $l \in [1, L]$: M_l and F_l which are respectively the number of units in layer l and the activation function in layer l. However, as we only have a relatively small number input variables to begin with, we keep this constant over all $L-3$ hidden layers.[7] We then include three fully connected linear layers in order to decrease the output gently to one value. Moreover, from previous research we know that deeper networks are preferred as deep NNs represent complex behaviour more efficiently than shallow architectures (Vinodhini and Chandrasekaran, 2016).

Furthermore, $tanh$ and different forms of the $ReLU$ (basis form: $ReLU = max(0, input)$) are the most common non-linear activation functions. They are preferred, as the function and its derivative are computationally easy to compute. Moreover, the gradients are quite strong, which is needed for updating the weights. Before each non-linear function, we place a linear function followed by a so called batch normalisation layer. The latter ensures that the output of the linear layer is normalized and centred around the origin so that the non-linear function can be fully utilized. Hence, we construct three different modules the following way. In the grid search procedure, we only search for the right combination of modules. From this and the three last layers the total number of layers L follows immediately.

1. linear function followed by a $tanh$
2. linear function followed by a $ReLU = max(0, input)$
3. linear function followed by a $LeakyReLU = max(input, 0.01 * input)$ (hence input if greater than zero and 0.01 times the input when it is lower than zero)

Note that the NN is expected to outperform the econometric benchmark regression, since the NN does not need handcrafted interaction terms as it can capture the nonlinear relation by itself by capturing more complex relationships than handcrafted interaction terms. However, the difference will be small, as a NN works better with continuous variables. We, on the other hand, have lots of binary indication functions, as the variables are simply binary or they are made binary (for example $HighYield$) for the analysis with the econometric method.[8] This suggests that a tree algorithm, like the one considered in the next section, should outperform both the NN and the econometric benchmark regression.

[7] We change the number to the closest number that is a power of 2 for computational easiness.

[8] In the main analyses we keep $HighYield$ as a variable instead of the yield-to-maturity, which would be a continuous variable, in order to be able to compare the output of the different models with each other. In the robustness section, however, we use the continuous variable.

4.2.2 XGBoost

This section contains the methodology concerning XGBoost, which stands for Extreme Gradient Boosting. It is a promising tree boosting algorithm, introduced by Chen and Guestrin (2016), that is used in a wide range of fields and performs well in more or less all cases (Chen and Guestrin, 2016).[9]

XGBoost uses several binary decision trees to make predictions. First, we describe a single tree. A decision tree is a model that successively splits the data into subsets by testing a single predictor in each node. As XGBoost uses binary trees, a single condition is tested (e.g. $Domestic = 1$ or $\Delta Tier1 > 0$). Hence, the observations for which the condition in the node is true go to one child node and the observations that do not satisfy the condition, go to the other child node.

At each end node, leaf, a specific subset of the data remains and for that set a prediction is made. Hence, in our case it predicts $\widehat{\Delta Holdings_{irt}}$ for different subsets of the data. XGBoost makes the subsets on its own by dividing the data, where it minimzes the total loss in each step it takes. Hence, the algorithm finds the best variable and the best place to split that variable on by calculating for each point the total loss and then selecting the combination that minimizes the loss. The loss is a measure of how far the predicted values are from the real values.[10] Often the squared error function is the chosen loss function. Splitting the tree continues until one of the stopping criteria is reached, for example when the maximum number of subsets are created. When it stops, a single tree is constructed.

After constructing the first tree, XGBoost constructs a new tree. However, instead of predicting the original variable, in our case $\Delta Holdings_{irt}$, this tree predicts the residuals that follow from the predictions of the previous tree.[11] The prediction of the newly constructed tree is then added to the prediction(s) of the previous tree(s).

Combining the trees in this way is called gradient boosting. This is a highly effective and frequently used method in machine learning (Chen and Guestrin, 2016). It combines many 'weak' or poor predictions into a 'strong' or accurate prediction (Berk, 2008).

XGBoost has more hyper-parameters than NNs which have to be chosen optimally, such as the maximum depth of a tree. In Appendix A.2.4 a table with the default values and a full description is given for each parameter. Moreover, in the same section the used tuning procedure can be found. It is inspired by the stepwise procedure of Xia et al. (2017) and uses 3-fold cross-validation to tune the data. Having analyzed the machine learning methods, we now study possible methods for statistical inference on non-parametric models.

[9]For a detailed explanation one can read Chen and Guestrin (2016).

[10]Besides this loss, XGBoost also adds a regularization term Ω to the function that has to be minimized. This is the main innovation of XGBoost and it reduces the chance of overfitting, as this term can cause the loss function to increase after a split (instead of decrease). If this holds for all split options, the algorithm will not make any further splits.

[11]Note that this is different than making the last tree deeper, since the algorithm looks at all the data again. Hence, it might find other variables that explain the difference, while it did not explain the difference in the specific subsets of the data in the leaves of the foregoing tree(s).

4.3 Shapley framework

A major benefit of the benchmark econometric estimation is the fact that it has interpretable coefficients for each included variable. This is in contrast with machine learning models, which only have limited interpretability. Hence, this would suggest an exclusion of these models in the case of applications in which model interpretability is important (Joseph, 2019). However, as Joseph (2019) proposes a framework (Shapley framework) for statistical inference on non-linear or non-parametric models, this needs not necessarily be the case. The main goal of this section is to explain the Shapley framework and consider its implementation in models used in this thesis. For more details, one can read Joseph (2019).

For simplicity we change the notation in this section in the following manner, as this prevents the subscripts from getting too long and benefits the legibility. One observation is denoted by the ISIN i, RBG r and time t as the irt subscript. Here, such an observation is denoted as i (different than the i of the ISIN). Moreover, we denote the dependent variable $\Delta Holdings$ as y. The variables will be denoted by the subscript j and there is a total of k variables.

The first step in the Shapley framework, is obtaining the Shapley values. Shapley values are a concept from the field of game theory. They define how much a particular player contributes in a cooperative game consisting of a group of players. This contribution depends on the players that are already in the game and on the order in which they enter. By considering all the different subgroups and orders for a single player, one arrives at the resulting marginal contribution, which is also called the Shapley value for that particular player.

In mathematical terms, the Shapley value for player j, denoted by ϕ_j, is given by:

$$\phi_j = \sum_{S \subseteq N \setminus j} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)]$$

where N be the set of all players and S a coalition of players with payoff $f(S)$.

A similar framework can be followed in the context of regressions, where the variables are seen as the players. Hence, the Shapley value for variable j is the contribution of that variable to the prediction, which can be obtained for each observation i and is denoted by ϕ_{ij}. Hence, $f(S)$ becomes $f_i(S)$, which is the predicted value for observation i and S is the set of included variables to obtain that prediction. The Shapley values are calculated with respect to the average mean predicted value of the training set, denoted by c. Hence, the predicted value of a model can be decomposed into the sum of the Shapley values and the mean predicted value: $\hat{y}_i = \sum_{j=1}^{k} \phi_{ij} + c$.[12]

Note, that variables that are not present in S, are left out. However, this is not possible with most machine learning methods. A solution to this would be to integrate these variables out by replacing them with all observed values in the validation set.[13]

[12]The Shapley values are calculated with the python package *SHAP*. The resulting Shapley values are approximations, hence the average is taken over 100 computations. A higher number is not achievable, due to the limited computer resources and time.

[13]Since the validation set is still quite large, we reduce the number of observations using k-means (as suggested

Although Shapley values are effective at measuring how much each variable drives the prediction, they provide no information on how well the variable predicts as they simply decompose the predicted value into the variables, but do not compare it to the true value. Hence, in order to get an estimate, Joseph (2019) suggests using Shapley regressions in order to judge the economic and statistical significance of the variables. This is the second step of the Shapley framework and entails regressing the dependent variable y on the Shapley value matrix Φ, which contains all ϕ_{ij} and a constant:

$$y = \Phi\beta^S + \varepsilon$$

The results from this regression determine whether the coefficients of the third step of the Shapley framework are robust and significant. The significance results from the testing the coefficients for each variable j, $\hat{\beta}_j$, against the null hypothesis $\mathcal{H}_0^j(\Omega) : \{\beta_j^S \leq 0|\Omega\}$. The coefficients are significant if the test rejects, as values close to zero imply that there is no relationship between y and the Shapley values which means that the estimates are not statistical significant. Moreover, the coefficients should not be negative as that implies that the effect of the variables would be the opposite compared to the effect estimated by the Shapley values. The dependence on a sub-domain of the data, Ω, follows from the fact that we use a region of the total dataset, D to construct the Shapley values with ($\Omega \subset D$). This implies that only local statements about the significance can be made (Joseph, 2019).

Furthermore, Joseph (2019) argues that for robustness of the results the values of β_j^S should be one, as that implies that the coefficients are unbiased and the machine learning model generalises well.[14] Hence, we have a robust estimate within the region Ω when $\mathcal{H}_0^j(\Omega) : \{\beta_j^S \leq 0|\Omega\}$ is rejected and $\mathcal{H}_1^j(\Omega) : \{\beta_j^S = 1|\Omega\}$ is not rejected. Instead of choosing a confidence level, we define an acceptable range for $\beta_j^S \in [0.9, 1.1]$, given that $\mathcal{H}_0^j$ can be rejected, which allows for a small amount of bias, as is done in Joseph (2019).

The last step in the Shapley framework is constructing the Shapley share coefficients (SSC) for each variable j. These coefficients can be interpreted as the coefficients of a linear model with the major difference that the results are only locally valid within the considered region Ω. The SSC for variable j, is denoted as Γ_k^S and is defined below where K is the number of observations in the subset Ω.

$$\Gamma_j^S = sign\big(\hat{\beta}_j^{OLS}\big)\frac{1}{K}\sum_{i\in\Omega}\left[\frac{|\phi_{ij}|}{\sum_{l=1}^{k}|\phi_{il}|}\right]$$

This statistic summarizes the contribution of variable j to the model over a region Ω. The sign is determined by the sign of the coefficient of a OLS, $\hat{\beta}_j^{OLS}$, regression of y on the original data (not the Shapley value matrix). The size of the coefficient is the fraction of the absolute

by the package). For the NN we chose k=100 to keep the computation time reasonable as the implementation for the NN uses an approximation. For XGBoost constructing the Shapley values takes less long, hence we use k=1000 there.

[14]For full derivation consult Joseph (2019).

Shapley value of variable j w.r.t. the summation of all Shapley values.[15] Since the Shapley values differ per observation, the average is taken over all observations in Ω.[16]

Furthermore, the standard errors can be constructed as the classical central limit theorem applies to the sampling distribution of Shapley values (Joseph, 2019). Moreover, they are interpretable as the SSC is normalized.[17]

Lastly, we could estimate the effect of interactions between variables within the Shapley framework. The first step changes in two ways. First, the Shapley interaction values $\phi^i_{r\times s}$ of variables r and s have to be calculated.[18] Second, the Shapley values of the original variables have to be adjusted in order to control for the interactions (Bluwstein et al., 2019). This means that when $\phi^i_{r\times s}$ is included, the Shapley value for r is adjusted as $\phi_r = \phi_r - \phi^i_{r\times s}$. The same holds for the Shapley value of s. Using the same notation as before, we get the following formula for the Shapley interaction value between r and s.

$$\phi^i_{r\times s} = \sum_{S\subseteq N\setminus\{r,s\}} \frac{|S|!(|N|-|S|-2)!}{2(|N|-1)!}[f(S\cup\{r,s\}) - f(S\cup\{r\}) - f(S\cup\{s\}) + f(S)]$$

Having discussed the Shapley framework for model inference, we now discuss a novel method for this purpose which has a number of preferable properties, which we discuss in the next section.

4.4 A novel framework to assess the sign

This section contains a novel approach in determining the sign of the SSC values from the Shapley framework. The proposed method is not limited to only determining the sign of the SSC values, but can be used in order to estimate the effects of certain variables, comparable to the coefficients found in econometric regressions. The method has a number of advantages over the Shapley framework.

Firstly, we know that the signs of the SSC values are determined using an econometric regression. There is, however, no check on whether the regression is correctly specified according to econometric theory. For example, there could be omitted variable bias. When the omitted variables are included, the sign of the coefficients of the variables we are interested in, could change. Hence, more attention should be paid to get the right signs from the econometric regression.

Secondly, another major critique on the method for determining the signs using the Shapley framework is the fact that the framework assumes that the sign from the econometric regression is correct and uses no information from the machine learning models to construct the sign. This

[15] The normalising factor accounts for localised properties of non-linear models.

[16] The Shapley values are calculated on the test set, Ω in our case, in order to prevent information leakage.

[17] The code provided by Bluwstein et al. (2019) is used to calculate the standard errors.

[18] For the XGBoost model, we use the package *SHAP* to calculate the Shapley interaction values, while for the NN model we implement the calculations ourselves as this is not implemented in the package.

implies that the sign for all models are the same. Hence, when we consider our hypotheses, the conclusion would be the same for all models. The size of the effect could differ per model.

Hence, another way to calculate the sign of the SSC values is required. The method we propose involves estimating the derivative w.r.t. each variable for different values of that variable. Hence, it is based on the output of the machine learning themselves, which means that it can differ per method of estimation. Moreover, the sign of the derivative should be the sign of the SSC, as the sign of the derivative shows in which direction a function is updated if the corresponding variable changes in value. Furthermore, the derivative itself actually already gives the size of the change, which means that the SSC value is not needed anymore. The remaining of this section first discusses the novel method for simple variables. Thereafter, variables that are interaction terms are considered.

4.4.1 Novel framework with regular variables

The proposed method consists of the following steps. Assume we are predicting y with a function $f(.)$ and want to obtain the coefficient for variable x_i. Hence, we want to transform a possible non-linear function $f(.)$ to a linear one by estimating the coefficient β_i for each variable x_i in $f(.)$ after which we can define:

$$f(x_1, ..., x_k) \approx \beta_1 x_1 + ... + \beta_k x_k$$

We could estimate the coefficient β_i by taking the derivative of $f(.)$ w.r.t. x_i. However, we cannot do this exactly, if $f(.)$ is a complex non-linear model or if we do not know the exact form of $f(.)$. Hence, we numerically differentiate $f(.)$ w.r.t. x_i using the finite difference, as is displayed below.

$$\hat{\delta}_i = \frac{f(X, x_i + \varepsilon) - f(X, x_i - \varepsilon)}{2 * \varepsilon} \tag{4.3}$$

where $\hat{\delta}_i$ is the estimated derivative of $f(.)$ w.r.t. x_i, X contains all the other variables that we do not change and ε is a arbitrarily small number.

Hence, the sign of $\hat{\delta}_i$ should be the sign of the SSC coefficient of variable x_i. However, with a complex non-linear function $f(.)$, the sign can differ per value of x_i, as the derivative changes for varying values of x_i. As it is impossible to evaluate all values of x_i, we define a grid of x_i values for which we will calculate $\hat{\delta}_i$ with for each grid point an arbitrarily small ε. Moreover, we keep the other variables X fixed, as we assume that the effect of X in $f(X, x_i + \varepsilon)$ is the same as the effect of X in $f(X, x_i - \varepsilon)$. Hence, by taking the difference, the effect of X drops out.[19]

There are different ways in which a grid can be constructed. However, for continuous variables we suggest taking the sample minimum and maximum of the variable and constructing

[19]Different fixed values of X can be used. In this thesis, we use the average value in the (used sub-sample of the) dataset. Binary or categorical variables can be rounded to an integer, as non integer values have no meaning. Keeping the values fixed also represent the ceteris paribus condition which applies when we interpreted OLS coefficients.

a grid of 100 uniformly distributed points. Moreover, if we know the theoretical minimum or maximum from economic theory or the construction of a variable then we extend the grid with 5 uniformly distributed values ranging from the theoretical minimum to the empirical minimum. The same is done for the theoretical maximum and empirical maximum. For categorical or binary variables only the possible points are included in the grid, as other point have no meaning and can never occur.

Evaluating K points in the grid leads to K estimates of $\hat{\delta}_i$. The final estimate of the derivative, hence the coefficient, is denoted by $\hat{\Delta}_i$ and is the average over $\hat{\delta}_{i1}, ..., \hat{\delta}_{iK}$. To formally test the sign of Δ we test the following, where Z is the standard normal distribution.

Test 1) $\mathcal{H}_0 : \Delta_0 \leq 0$ against $\mathcal{H}_a : \Delta_0 > 0$
rejects $\mathcal{H}_0$ at significance level α if $T = \frac{\hat{\Delta}}{SE(\hat{\Delta})} > z_\alpha \Leftrightarrow p-value = Pr[Z > T] < \alpha$

Test 2) $\mathcal{H}_0 : \Delta_0 \geq 0$ against $\mathcal{H}_a : \Delta_0 < 0$
rejects $\mathcal{H}_0$ at significance level α if $T = \frac{\hat{\Delta}}{SE(\hat{\Delta})} < -z_\alpha \Leftrightarrow p-value = Pr[Z < T] < \alpha$

Hence, when the first test rejects the null hypothesis while the second one does not, then the sign is positive. If it is the other way around, then the sign is negative. Note that these tests indicate how certain this framework is that the sign of the model, $f(.)$, is positive or negative. It does not imply anything about how certain the model is that the estimated sign is the same as the true sign.

Besides taking the average, a plot of $\hat{\delta}_i$ against the different values of x_i also indicates whether there are clearly different signs for different groups of values. This can be tested formally, for example with a test similar to the chowbreak test, which tests for structural breaks. This gives us a better understanding of the sign then the way the signs are now constructed.

Moreover, we could in this manner also find the magnitude of the coefficient, which is simply the size of $\hat{\Delta}_i$ and replaces the SSC value. In order to be able to estimate the significance of $\hat{\Delta}$, we need the standard error of the estimate, $SE(\hat{\Delta}_i) = \sqrt{Var(\hat{\Delta}_i)}$. Using the bootstrap algorithm, we can estimate the bootstrap standard error as follows. Firstly, draw N random observations with replacement from the dataset of size N. This new dataset is called a bootstrap sample. Second, estimate $\hat{\Delta}_i$ as described above for this bootstrap sample.[20] Then, we again sample a new bootstrap sample and estimate the $\hat{\Delta}_i$. This is done for B bootstrap samples, which results in an estimate for each sample: $\hat{\Delta}_{i1}, ..., \hat{\Delta}_{iB}$. The bootstrap standard error then becomes:

$$SE_{boot}(\hat{\Delta}_i^*) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\Delta}_{ib} - \hat{\Delta}_i^*)^2}$$

[20]For machine learning models we do not have to retrain the whole model on the new dataset as this would result in a model with the same predictions. This is the case, since a machine learning model is trained in multiple rounds with several stopping criteria. Hence, the model has seen the same data already multiple times and thus becomes less sensitive for this kind of changes of the data. For the econometric regression, one should estimate the model again on the new bootstrap dataset.

where $\hat{\Delta}_i^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\Delta}_{ib}$. With this SE, we can determine the significance of the estimated coefficient by evaluating $\mathcal{H}_0 : \Delta = 0$ against $\mathcal{H}_a : \Delta \neq 0$. This rejects $\mathcal{H}_0$ at significance level α if $|T = \frac{\hat{\Delta}^*}{SE_{boot}(\hat{\Delta}^*)}| > z_{\alpha/2} \Leftrightarrow p - value = 2 * Pr[Z > |T|] < \alpha$.

Note that, as in the case of Shapley framework, the estimated coefficient $\hat{\Delta}$ is also equivalent to the coefficient β in an OLS regression if $f(.)$ is the OLS model. This can be easily seen as for a linear model, $f(x_1, ..., x_k) = \beta_1 x_1 + ... + \beta_k x_k$, the following holds.

$$\hat{\delta}_i = \frac{f(X, x_i + \varepsilon) - f(X, x_i - \varepsilon)}{2 * \varepsilon} = \frac{2 * \varepsilon * \beta_i}{2 * \varepsilon} = \beta_i \quad \forall i$$

Hence, the average over different estimates of $\hat{\delta}_i$ will also be β_i for all variables x_i.

4.4.2 Novel framework with interaction terms

Until now, we have only considered regular variables, which are given directly to the model $f(.)$. In this section, we consider the estimation of the effect of interaction terms using this framework. Assuming once again that we are predicting y using a certain function $f(.)$, which can be any kind of model, we can construct a local finite difference method in order to estimate both the sign and magnitude of the effects of a certain variable, as is shown above. In the case of interaction terms, however, we need to consider changes in both variables when the interaction term itself is not given as input to $f(.)$.[21] Moreover, we have to adjust the estimation of the coefficients of the regular variables which are involved in the interaction. In the remaining of this section, we first consider the adjustments that have to be made to the estimates of regular variables. Thereafter, we define how the coefficients of interaction terms can be estimated.

First, we consider how the estimation of the coefficients of the regular variables changes. If we want to transform a possible non-linear function $f(.)$ to a linear one by including an interaction term, which is the multiplication of a variable x_i with a variable x_j, beside the regular variables $x_i, ..., x_k$, we get the following estimation.

$$f(x_1, ..., x_k) \approx \beta_1 x_1 + ... + \beta_k x_k + \beta_{ij} x_i x_j$$

In the previous section a procedure is described to obtain estimates of β_i, which are denoted as $\hat{\Delta}_i$. However, when x_i is involved in an included interaction term we have to correct the estimate as the approximated derivative now becomes $\hat{\Delta}_i \approx \frac{\partial f}{\partial x_i} = \beta_i + \beta_{ij} x_j$. Hence, from $\hat{\Delta}_i$ we have to substract the estimated of the coefficient of the interaction term, denoted as $\hat{\Delta}_{ij}$, times the value of x_j. As $\hat{\Delta}_{ij}$ is averaged, we also take the average of x_j.

Second, we elaborate on how the coefficients of interaction terms can be estimated. The main framework for determining $\hat{\delta}_{ij}$ which is the effect of an interaction term, given by the multiplication of a continuous variable x_i with a continuous variable x_j, can be summarised by

[21] In general, we do not give a machine learning model handcrafted features, such as interaction terms, as input, since the machine learning models will capture any relationship on its own. When an interaction term is given to the model as input, then its Δ can be constructed as if it were a normal variable.

the following formula.

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i \partial x_j} &\approx \frac{\frac{\partial}{\partial x_i} f(X, x_j + \varepsilon_j) - \frac{\partial}{\partial x_i} f(X, x_j - \varepsilon_j)}{2\varepsilon_j} \\
&\approx \frac{f(X, x_i + \varepsilon_i, x_j + \varepsilon_j) - f(X, x_i - \varepsilon_i, x_j + \varepsilon_j)}{4\varepsilon_i * \varepsilon_j} - \frac{f(X, x_i + \varepsilon_i, x_j - \varepsilon_j) - f(X, x_i - \varepsilon_i, x_j - \varepsilon_j)}{4\varepsilon_i * \varepsilon_j} \\
&= \hat{\delta}_{ij}
\end{aligned}$$

where $\frac{\partial f}{\partial x_i}$ is approximated by $\hat{\delta}_i$ defined in Equation (4.3) and X is again a matrix which contains the other variables besides x_i and x_j. As with $\hat{\delta}_i$ the estimates $\hat{\delta}_{ij}$ can differ per value of the interaction term. Hence, we again define a grid for which $\hat{\delta}_{ij}$ will be estimated.

The construction of the grid is closely related to the procedure done in estimating the univariate effect, but has now one additional dimension. Hence, we construct a $2D$ grid consisting of 100 points for variable x_a and 100 points for variable x_b, leading to a grid of a total $10,000$ points. By evaluating $\hat{\delta}_{ij}$ for a fixed x_i and all possible grid points of x_j and taking the average, we can construct $\hat{\delta}_{ij}$ for each level of x_i and x_j. The minimum and maximum value of the grid are constructed in an identical manner to the one presented in the previous subsection. The bootstrap set up and tests for significance, constructed in the previous section for the prior specified univariate method, remain identical for Δ_{ij}.

Note that we must make a distinction between continuous variables, which result from a multiplication of two continuous variables, described above, and censored continuous variables, which result from multiplication of a continuous variable with a dummy. In the case of having a dummy as one of the components of an interaction term, one must make an adjustment on the procedure specified above. As the interaction term has a considerable number of values for which it is equal to zero, since the dummy is equal to zero and hence also the product, the interaction term is censored. Hence, the marginal effect of this variable consists of two components. A jump component, in which the dummy goes from zero to one, and a continuous component, in which the dummy is one and the continuous variable changes values. In order to account for this censoring, we propose dividing the grid in two components as follows. First, we assume variable x_j is the dummy variable and x_i is the continuous variable. Subsequently, the first 5% of the grid values for x_i, hence closest to the minimum, will serve as an estimate for the jump, as for those observations we will use $f(X, x_i \pm \varepsilon_i, 1) - f(X, x_i \pm \varepsilon_i, 0)$ in the expression above. The remaining 95% will model the continuous part, for which we always use $x_j = 1$, hence x_j can then be seen as part of the other variables contained in X.

Chapter 5

Results

This chapter is made up of two sections. The first section presents the general results from the three estimation methods, namely the econometric regression, Neural Networks and XGBoost. For the machine learning methods, we report two estimated OLS-like coefficients. The first is the SSC value, which is constructed through the Shapley framework. Note that the sign is determined by our novel method instead of by the Shapley framework. The second estimate follows from our novel framework. In the second section, each hypothesis is elaborated on by considering the results from the three methods per hypothesis.

5.1 General results

This section presents the results of the three different estimation methods. The first method is the econometric regression. This is our benchmark as the economic theory is well developed and the model is easy to interpreted. However, the model assumes that the underlying process is linear. In order to better capture the possible non-linear underlying process, we also use a Neural Network and a XGBoost model. The Neural Network is used, as it is able to model any process with utmost precision given enough data. A drawback of this method is that it cannot handle dummy variables properly. On the other hand we have XGBoost, which is a machine learning model that can handle dummy variables. Moreover, another benefit of XGBoost is the fact that it is able to combine both high model performance and execution speed.

As discussed in Chapter 4 there might be a problem with estimating the models on the full dataset, as there is a relatively small number of observations with a high value for $|\Delta Holdings|$. Hence, it might be better to use a subsample of the data that contains only the 95% largest ISINs, where the size is measured by the total outstanding market value. Since, the estimated effects are similar, we only show the results for the econometric model.

5.1.1 Econometric regression

This section presents the results from the econometric regression. The details of the regression and specification of several used tests can be found in Appendix A.2.2. The resulting estimates are displayed in the first column of Table 5.1 for the results without accounting for time-varying effects and in the second column for the results with those effects. The coefficients and their significance changes slightly, but are comparable. The AIC and BIC values are lower for the model with the time-varying effects.

Hence, in the case of including time-varying effects, we now look into cluster-robust standard errors. There may be serial correlation within the cluster on the level of RBG-Issuer Sector combinations. The clustered standard errors correct for both heteroskedasticity and autocorrelation. Hence, we should try RBG-Issuer Sector cluster combinations. This results in column third column of Table 5.1. Note that the only difference between the second and third column are the standard errors, the coefficients are exactly the same. The normal standard errors are indeed lower than the cluster robust errors, hence the first leads to an overestimation of the accuracy of the results.

The last column of Table 5.1 shows the results of regressing on the sub-sample of the training set which contain only the larger ISINs, as we exclude the 5% smallest. This is done in order to overcome the problem of a relatively small number of observations with a high absolute value of $\Delta Holdings$.[1] The AIC and BIC values drop as expected, since the model does not have to explain very different changes in holdings. Excluding more values, results in even smaller AIC and BIC values. However, we prefer to keep more observations.

In all cases, the effect of a one percentage point drop in Tier 1 capital ratio results in an increase in the holdings for all ISINs of around 0.07 percentage point per ISIN.[2] This indicates that when the capital of the bank decreases, the bank increase its holdings for all ISINs. When a ISIN becomes high-yield w.r.t. all other ISINs in t while it was low-yield in $t-1$, the holdings in this ISIN are adjusted downwards with around 0.3 percentage point. This means that the RBGs do not like the increase in riskiness and hence lower their holdings. This would mean that there is always a flight-to-safety (hypothesis 1), regardless the capital of the bank. The same holds for $\Delta HighYieldSector$, although this effect is not always significant.

The other variables are individually not significant. Hence, we cannot determine directly whether the flight-to-safety is amplified when the capital of a bank decreases, nor if there is moral suasion or regulatory arbitrage. However, as the variables were jointly significant, the coefficient could still be considered to be meaningful. Section 5.2 contains further analysis regarding the hypotheses.

The presented regressions capture linear relationships in the coefficients. The interaction terms account for some non-linearity. However, there is still evidence that the model is not

[1]Elaborated on in Chapter 3 and Chapter 4.

[2]Note that the holdings are defined as the percentage of the total outstanding market value a RBG posses. Hence, the change in the holdings is a percentage point difference.

Table 5.1: Several first difference estimations of the change in holdings: $\Delta Holdings$

	(1)	(2)	(3)	(4)
$\Delta Tier1$	-0.0755***	-0.0653**	-0.0653	-0.0796*
	(0.0260)	(0.0294)	(0.0457)	(0.0427)
$\Delta HYTotal$	-0.3026***	-0.3143***	-0.3143**	-0.2822*
	(0.0764)	(0.0766)	(0.1520)	(0.1560)
$\Delta HYSector$	-0.1388**	-0.1361*	-0.1361	-0.1091
	(0.0698)	(0.0698)	(0.0879)	(0.0867)
$\Delta HY \times \Delta Tier1$	-0.0070	-0.0012	-0.0012	0.0155
	(0.0080)	(0.0080)	(0.0170)	(0.0152)
$GOV \times \Delta Tier1$	0.0907**	0.0394	0.0394	0.0406
	(0.0380)	(0.0412)	(0.0315)	(0.0301)
$DOM \times \Delta Tier1$	-0.0164	0.0195	0.0195	0.0279
	(0.0318)	(0.0329)	(0.0417)	(0.0373)
$GOV \times \Delta HY \times \Delta Tier1$	0.0243	0.0196	0.0196	0.0012
	(0.0156)	(0.0156)	(0.0137)	(0.0113)
$DOM \times \Delta HY \times \Delta Tier1$	0.0127*	0.0020	0.0020	-0.0176
	(0.0065)	(0.0065)	(0.0208)	(0.0181)
$GOV \times DOM \times \Delta HY \times \Delta Tier1$	0.0015	0.0100	0.0100	0.0290
	(0.0188)	(0.0187)	(0.0237)	(0.0215)
Observations	252,258	252,258	252,258	240,950
Adjusted R-squared	0.0004	0.0126	0.0126	0.0130
Adjusted Within R2	0.0004	0.0004	0.0004	0.0002
AIC	1640930	1637670	1637670	1526948
BIC	1641035	1637774	1637774	1527052
RMSE	6.26	6.22	6.22	5.75
Sub-sample high outstanding	No	No	No	Yes
Holder Area x Time	No	Yes	Yes	Yes
Issuer Sector x Time	No	Yes	Yes	Yes
Cluster robust errors	No	No	Yes	Yes

Standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Panel regression with fixed effects predicting the change in the holdings in a security. Changes in variables are denoted by Δ *variable name*. $HYTotal$ ($HYSector$) stands for the indicator function for $HighYieldTotal$ ($HighYieldSector$). HY is the interaction between both types of high-yield. GOV represents the indicator variable *Government*. Likewise, DOM stands for the indicator variable *Domestic*. The row sub-sample high outstanding indicates whether the subsample that contains all observations for which the outstanding amount of the ISIN is larger than the 5th percentile is used.

linear in the included parameters, as the Ramsey RESET test on the model including time-varying fixed effects and cluster robust errors, give a p-value of 0.0874. Hence, the model is not linear in the included parameters at a 10% significance level, as we reject the null hypothesis of being linear in the included variables. These possible non-linearities might be better captured with machine learning models.

5.1.2 Neural Networks

This section consists of the results obtained from the Neural Network method. In Appendix A.2.3 the implementation details are discussed. Moreover, the losses for different NNs are shown. From the models we tested, the NN consisting of one *tanh* module followed by three linear layers performs best. However, the differences are quite small. For this model, the RMSE on the training set is around 6.21, which is comparable with the RMSE from the econometric model (6.22). This is in line with the expectation that the NNs do not perform (much) better than the econometric benchmark, as there are lots of dummy variables and we can not include a dummy for each ISIN.

Analyzing the Shapley values show that the estimation is dominated by the variable $ISIN$ (see Appendix A.2.3). This is quite striking, as it is expected that a more informative variable such as the price of security or the yield-to-maturity is more informative than an encoded ISIN number. Moreover, the network does not pick up any effect for the variables of interest, as the ISIN effects dominates. A solution is to include only indicator functions for each ISIN instead of the other ISIN related variables. However, this is not possible due to memory limitations.

Another solution is to use a technique from econometrics as a data preprocessing step. Instead of including ISIN specific dummies, we demean for example $\Delta Holdings_{irt}$ w.r.t. the mean of $\Delta Holdings_i$ over time. Hence, each ISIN has its own mean, which capture the ISIN specific characteristics. This gives the model the opportunity to capture deviations from the mean, as demeaning centres the data. Besides the dependent variable, all explanatory variables are demeaned, except the included indicator variables as they are time invariant.

Demeaning the data in this way is used here. The RMSE on the training data is 5.21, which is much lower than without demeaning the data. The resulting SSC values of the variables we are interested in are shown in Table 5.2. Most variables are not significant, which is also the case in the econometric regression. Moreover, none of the coefficients are robust, which means that the model does not generalize well.

The SSC value for a change in the Tier 1 capital ratio is comparable with the coefficient from the econometric regression, -0.0855 and -0.0653 respectively. The other values are much lower then their econometric counterpart. Hence, the importance of the variables is lower. Furthermore, the signs of the SSC values, determined by our framework, do not always mach the sign of the econometric regression. This highlights the importance of having a framework in which the machine learning model itself determines the sign.

As the SSC coefficients are insignificant and non-robust, we also take a look at the results of our novel framework. The estimated coefficients are shown in the second column of Table 5.2. The absolute values of the coefficients are in general higher than the absolute SSC values. The signs are identical, as both signs are determined by our novel framework. Lastly, note that more coefficients are significant based on our novel framework compared to the SSC framework.

Table 5.2: Estimations of the change in holdings $\Delta Holdings$ with a NN after demeaning

Variable	Shapley framework SSC	Novel framework $\hat{\Delta}^*$
$\Delta Tier1$	-0.0855**	-0.2562
	(1.0851, 0.0309, No)	(1.6622, 0.8775)
$Government$	-0.0027	-0.0017***
	(0.0919, 0.2599, No)	(0.0000, 0.0000)
$Non-Government$	0.0069	0.0139
	(0.1796, 0.1107, No)	(0.0605, 0.8185)
$\Delta HYTotal\_down$	-0.0017	-0.0023***
	(0.0772, 0.3091, No)	(0.0000, 0.0000)
$\Delta HYTotal\_same$	-0.0053	-0.0089***
	(0.1720, 0.4721, No)	(0.0000, 0.0000)
$\Delta HYTotal\_up$	-0.0015	-0.0036
	(0.0721, 0.2387, No)	(0.0046, 0.4233)
$\Delta HYSector\_down$	0.0024	0.0040***
	(0.0978, 0.4614, No)	(0.0000, 0.0000)
$\Delta HYSector\_same$	-0.005	-0.0069***
	(0.1592, 0.2457, No)	(0.0000, 0.0000)
$\Delta HYSector\_up$	0.0064	0.0181**
	(0.2589, 0.4106, No)	(0.0089, 0.0429)
$Domestic$	-0.0308	-0.0157
	(0.3305, 0.3539, No)	(0.0296, 0.5965)
$Foreign$	0.0316	0.0113***
	(0.3324, 0.4754, No)	(0.0000, 0.0000)
$\Delta Tier1 \times HY$	0.000008	0.0168
	(0.0002, 0.2276, No)	(0.1197, 0.8887)
$\Delta Tier1 \times Domestic$	0.000017	0.0068
	(0.0002, 0.1385, No)	(0.0785,0.9307)
$\Delta Tier1 \times Government$	-0.000008**	-0.0380
	(0.0001, 0.0322, No)	(0.3228, 0.9062)

Importance of variables according to the NN. First column contains the Shapley share coefficients (SSC) from the Shapley framework. The sign is determined by our novel method and is NOT based on the econometric regressions as described by Shapley framework. Estimated standard errors ($\times 10^{-3}$) for the SSC, p-values and whether the SSC is robust are shown in parentheses. The second column contains the estimated coefficients from our novel framework. Estimated standard errors for $\hat{\Delta}^*$ and p-values are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Changes in variables are denoted as Δ *variable name*. $HYTotal$ ($HYSector$) stands for the indicator function for $HighYieldTotal$ ($HighYieldSector$). Moreover, *down*, *up* and *same* mean respectively that the security went from high-yield to low-yield, from low-yield to high-yield or it stayed the same. ($Non-$)$Government$ is one when a security is (not) issued by a government. Likewise, $Domestic$ ($Foreign$) is one when a variable is labelled domestic (foreign).

Instead of looking at the SSC value, we could also look at the underlying Shapley values for each original value of the variable. Figure 5.1 shows the Shapley values of a subset of variables. Overall, the change in Tier 1 capital ratio has the second largest average impact on the output of the NN. The variable with the largest impact is a country dummy.

As Figure 5.1 shows, one can see for most variables a clear distinction between the sign of the Shapley value and the feature value. Note, however, that the sign of the Shapley value does not say anything about the sign that the SSC should have, as the Shapley value only contains information how much a variable contributes to the model.

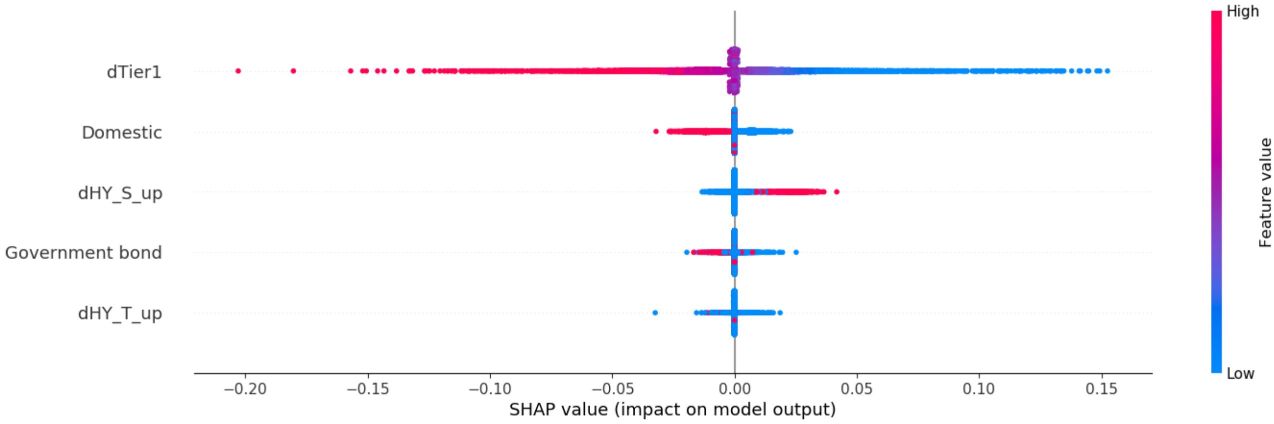


Figure 5.1: Shapley values

Concluding, the NN performs similar to the econometric regression without demeaned data. If the NN benefits from demeaning the data, the RMSE becomes much lower. However, demeaning the data assumes a linear relation, while if the ISINs could be given to the network as dummies, then the NN could account for non-linear relationships. Hence, demeaning might be restrictive for a non-linear model. A machine learning model that does not need the categorical variables as separate dummies, is XGBoost. The results for this model are considered next.

5.1.3 XGBoost

This section consists of the results obtained from XGBoost. The implementation details are discussed in Appendix A.2.4. Moreover, it shows the results of tuning the hyper-parameters. The tuned model consists of 50 trees (boosting rounds) with a maximum depth of only 5. Hence, it combines a lot of short trees to make one final prediction.

Compared to the NN and the econometric regression, XGBoost has a smaller in-sample RMSE when the models are trained/estimated on the same dataset. The RMSE on the train set is namely 5.74 for XGBoost. However, the NN on the demeaned data has an even lower RMSE.

Table 5.3 shows the SSC values for the variables we are interested in. From these variables, the change in Tier 1 capital ratio has the highest absolute SSC. The effect of -0.0251 is smaller than the effect of -0.0653 estimated with the econometric regression. Furthermore, the SSC value of $\Delta Tier1$ based on the NN is even lower than the econometric coefficient, namely -0.0855, and is significant. Hence, all methods capture a negative effect but they differ on the size and significance of it. Note, that the higher and significant result for the NN could be due to the demeaned train set. Hence, XGBoost might find another result when it is trained on a demeand data set.

The other SSC values are again smaller than the coefficients of the econometric regression. Moreover, only one SSC value is significant. Note that the estimate of the change in high-yield w.r.t. the sector is robust, while the other estimates are not.

The second column of Table 5.3 shows the estimated coefficients based on the novel method. As with the NN the absolute values of the coefficients are in general higher than the absolute SSC values. Moreover, the sign matches in both cases, as both signs are determined by our novel framework. Note, however, that an equal number of coefficients is significant based on the SSC and the novel coefficients. An explanation is that the novel framework has trouble with approximating the derivative of XGBoost as this is a (complicated) step-function instead of a continuous function.

Table 5.3: Estimations of the change in holdings $\Delta Holdings$ with XGBoost and interactions

Variable	Shapley framework SSC	Novel framework $\hat{\Delta}^*$
$\Delta Tier1$	-0.0251	-0.0138
	(0.0003, 0.1169, No)	(0.0812, 0.8645)
$Government$	0.0000	-0.0207***
	(0.0000, 0.4111, No)	(0.0028, 0.0000)
$\Delta HYTotal$	0.0034[a]	0.0000
	(0.0001, 0.1415, No)	(0.0004, 0.9045)
$\Delta HYSector$	0.0039[a]	0.0000
	(0.0002, 0.2584, Yes)	(0.0003, 0.9948)
$Domestic$	0.0147	0.0130**
	(0.0002, 0.4428, No)	(0.0065, 0.0462)
$\Delta Tier1 \times \Delta HYTotal$	0.000065	0.0021
	(0.000003, 0.2768, No)	(0.0170, 0.9003)
$\Delta Tier1 \times \Delta HYSector$	0.0001**	0.0004
	(0.000003, 0.0165, No)	(0.0164, 0.9821)
$\Delta Tier1 \times Domestic$	-0.0003	-0.0033
	(0.000008, 0.3534, No)	(0.0173, 0.8493)
$\Delta Tier1 \times Government$	0.0000	-0.0007
	(0.000, 0.3452, No)	(0.0152, 0.9643)

Importance of variables according to XGBoost. First column contains the Shapley share coefficients (SSC) from the Shapley framework. The sign is determined by our novel method and is NOT based on the econometric regressions as described by Shapley framework (SSC[a] means that the sign could not be determined as none of the hypothesis rejected). Estimated standard errors for the SSC, p-values and robustness of the SSC are shown in parentheses. The second column contains the estimated coefficients of our novel framework. Estimated standard errors for $\hat{\Delta}^*$ and p-values are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Changes in variables are denoted by Δ *variable name*. $HYTotal$ ($HYSector$) denotes the indicator function for $HighYieldTotal$ ($HighYieldSector$). $Government$ denotes the indicator function for when a security is issued by a government. Likewise, $Domestic$ denotes the indicator function for when a variable is labelled domestic.

Instead of looking at the SSC value, we could also look at the underlying Shapley values for each original value of the variable. Figure 5.2 shows the Shapley values of a subset of variables, including the variable with the larges absolute Shapley values, namely $resid\_mat$, which is the maturity to expiration. Note that this is not the $ISIN$ variable as it was with the NN.

As Figure 5.2 shows, for most variables there is a clear distinction between the original value of that variable and the Shapley value of that variable for that observation. For example, observations for which $Domestic$ is 0, hence foreign securities, have a Shapley value for the variable $Domestic$ of (around) zero, while the observations with an original value of 1 have a positive Shapley value for the variable $Domestic$.

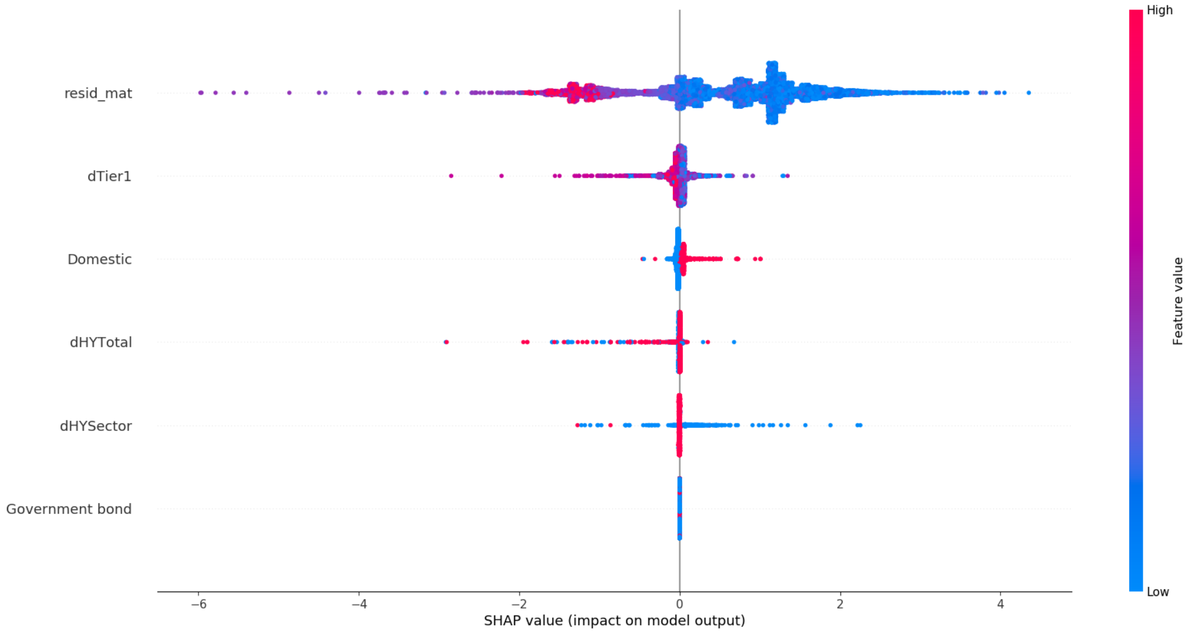


Figure 5.2: Shapley values

This section described the general results of the last model. We did not discuss the interaction terms, as those capture the different hypotheses and are discussed for all methods per hypotheses in the next section.

5.2 Hypotheses

This section reviews the results in the context of the formulated hypotheses of Section 2.1.3. Per hypothesis the three estimation methods are considered. One could argue that the method with the lowest out-of-sample RMSE has best captured the underlying process. Hence, the coefficients of that model are closest to the truth. From Table 5.4 we can conclude that the NN on demeaned data has the best out-of-sample performance. Note, however, that the models are not trained with exactly the same variables. Moreover, for the NN the data is first demeaned.[3]

Besides the difference in methods, there is also a difference in dataset, as we estimate each model on all observations in the training set and on a subset of the training set which contains only the larger ISINs as we exclude the 5% smallest. This is meaningful, as one could argue that the decision to change the holdings in ISIN i slightly depends on other issues than the

[3]Note that for the NN the out-of-sample RMSE is lower than the RMSE on the training set. In general the training error will underestimate both the validation error and test error. However, in our case the latter errors are overestimated. A possible explanation for this could be that the test and validation set contain easier observation, which do not significantly differ from the "average" model output. Moreover, one can relate this to the demeaning step, as the data is demeaned per subset of the data such that there is no information leakage and ISINs that are not present in the train data can still be used in the analysis.

Table 5.4: RMSE on the test set

	Full test set	Subsample of test set
Econometric regression	6.45	5.97
NN demeaned	4.62	4.39
XGBoost	6.09	5.58

decision to step entirely in or out the ISIN, causing large changes. However, despite the models perform better on the subset, the effect of the variables are in line with the effects estimated on the entire dataset. Hence, we only compare the models based on the entire dataset.

The remainder of this section is organized as follows. Firstly, the flight-to-safety hypothesis is discussed. Subsequently, the moral suasion hypothesis is analysed. Thereafter, the last hypothesis of regulatory arbitrage is elaborated on.

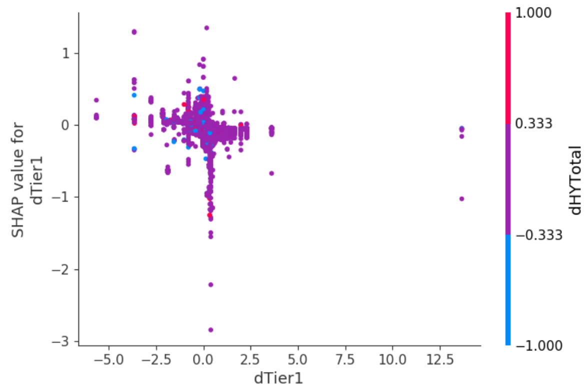
5.2.1 Flight-to-safety

This section contains the analysis regarding the flight-to-safety hypothesis. The flight-to-safety hypothesis implies that the coefficient of $\Delta HighYield \times \Delta Tier1$ is positive, as in this case a decrease in capital will cause an increase in the holdings of the security for which the yield went down. Hence, when the capital decreases the holdings in securities with a decreased risk increases.

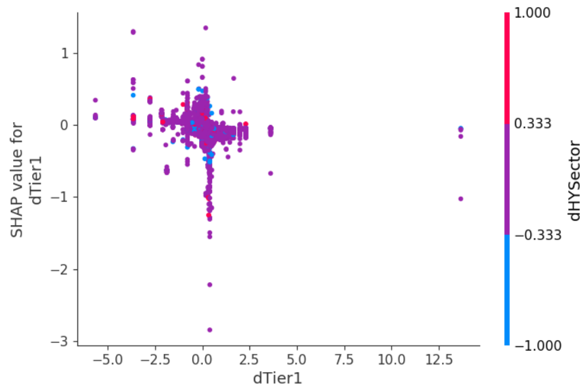
With the econometric regression, the coefficient for a single variable is the same for all values of that variable. However, with the machine learning models the actual contribution of a variable may differ for the different values of a variable. By following the Shapley framework, we construct a single coefficient-like value, namely the SSC. As the underlying Shapley values for a variable differ per observation, we first take a look at a plot for the Shapley values.

Figure 5.3 and Figure 5.4 show the Shapley values of $\Delta Tier1$ w.r.t. the different high-yield variables for the XGBoost and NN model respectively. In Figure 5.3 red dots denote observations for which the security went from low- to high-yield. Purple stands for securities that did not change from yield bin and blue represents the securities that went from high- to low yield. As Figure 5.4 colors on $\Delta HighYieldTotal\_down$ and $\Delta HighYieldSector\_down$ red means from high- to low-yield and blue captures the securities that stayed in the same bin or went from low- to high-yield.

The figures show that there is no clear support for the flight-to-safety hypothesis based on the Shapley values themselves, as we would then see that the the colors are grouped. For example, if in Figure 5.3 the points for which the Shapley value is negative were all red, then we could say that we expect the sign of an interaction between the change in Tier 1 and the indicator becoming high-yield would be negative, which would support the flight-to-safety hypothesis.

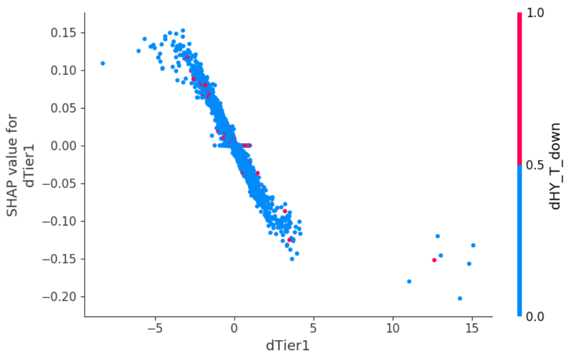


(a) Interaction $\Delta Tier1$ and $\Delta HighYieldTotal$

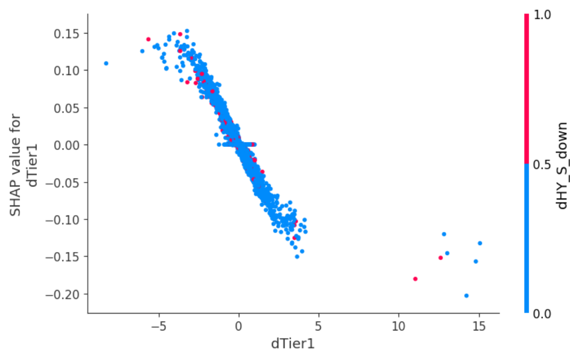


(b) Interaction $\Delta Tier1$ and $\Delta HighYieldSector$

Figure 5.3: XGBoost | Shapley values for a change in Tier 1 capital ratio (denoted as dTier1)



(a) Interaction $\Delta Tier1$ and $\Delta HighYieldTotal\_down$



(b) Interaction $\Delta Tier1$ and $\Delta HighYieldSector\_down$

Figure 5.4: NN | Shapley values for a change in Tier 1 capital ratio (denoted as dTier1)

Instead of analyzing possible interactions graphically, we could also calculate the Shapley interaction terms or the coefficient of the interaction. Table 5.5 shows relevant coefficients and SSC values for the three methods. Note that the methods do not estimate exactly the same variables. Furthermore, the econometric regression seems to overestimate the effect of each variable as the absolute values are much higher than SSC values of the machine learning methods. However, as the by our novel framework estimated coefficients $\hat{\Delta}^*$ are (much) higher than the SSC values, it is also possible that the SSC values underestimate the effect.

The econometric benchmark regression does not find prove in favour of the flight-to-safety hypothesis as the sign of $\Delta HighYield \times \Delta Tier1$ is negative. However, it finds results in favour of the hypothesis when it looks at domestic securities, although not significant. Furthermore, the machine learning methods find a positive SSC and $\hat{\Delta}^*$ value, which is in line with the hypothesis. XGBoost finds a larger and positive effect for the yield measurement w.r.t. the other securities in the same sector than the measure w.r.t. all other securities. For the SSC values this extra effect is even significant. Hence, our results suggest that some evidence exist in favour of the flight-to-safety hypothesis. Moreover, the effect might be amplified for domestic securities and securities that have a low-yield w.r.t. the other securities in the same sector.

Furthermore, note that the signs estimated by the machine learning methods for $\Delta HighYield \times \Delta Tier1$ are not equal to the sign of the econometric regression coefficients. Hence, based on this coefficient the econometric model would conclude that the hypothesis is false, while the machine learning models find prove in favour of the hypothesis. This illustrates the importance of letting the machine learning methods determine their own sign.

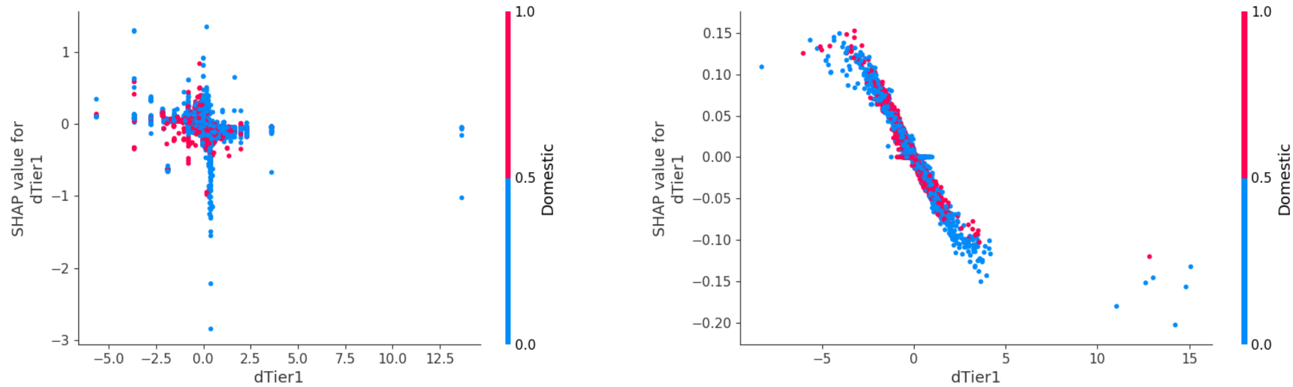
Table 5.5: Coefficients relevant to flight-to-safety hypothesis

	Econometric	NN		XGBoost			
	regression	SSC	$\hat{\Delta}^*$	SSC		$\hat{\Delta}^*$	
Yield measurement:	*combination*	*combination*	*w.r.t. total*	*w.r.t. total*	*w.r.t. sector*	*w.r.t. total*	*w.r.t. sector*
$\Delta HighYield \times \Delta Tier1$	-0.0012	0.000008	0.0168	0.000065	0.000115**	0.0021	0.0004
$Government \times \Delta HighYield \times \Delta Tier1$	0.0196	-	-	-	-	-	-
$Domestic \times \Delta HighYield \times \Delta Tier1$	0.0020	-	-	-	-	-	-

Relevant coefficients for the flight-to-safety hypothesis estimated by the different methods. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Changes in variables are denoted as Δ *variable name*. $(Non-)Government$ is one when a security is (not) issued by a government. Likewise, $Domestic$ $(Foreign)$ is one when a variable is labelled domestic (foreign).

5.2.2 Moral suasion

From the second hypothesis, concerning moral suasion, a negative coefficient of $Domestic \times \Delta Tier1$ is implied, as in this case a decrease in capital causes an increase in the holdings of domestic securities. In order to analyze the results for this hypothesis, we follow the same steps as in analyzing the previous hypothesis. Hence, we first consider the Shapley values themselves. Subsequently, we compare the different SSC and coefficients for the variables of interest.



(a) Interaction $\Delta Tier1$ and $Domestic$ in XGBoost

(b) Interaction $\Delta Tier1$ and $Domestic$ in NN

Figure 5.5: Shapley values for a change in Tier 1 capital ratio (denoted as dTier1)

The figures of the Shapley values show that there might be meaningful interactions. Figure 5.5 shows the Shapley values for a change in the Tier 1 capital ratio. Both figures color all the domestic securities red and the foreign securities blue. Moreover, they both show that negative changes in the Tier 1 capital ratio have in general a positive Shapley value. This is clearer for the values of the NN model than the ones of XGBoost. Furthermore, from both figures we see that the lowest Shapley values belong to foreign bonds. This means that a decrease in Tier 1 capital causes a larger increase in foreign securities, as the Shapley values are more negative for foreign securities. This in turn implies that there is no evidence for the moral suasion hypothesis as the holdings in foreign instead of domestic securities increase.

Table 5.6: Coefficients relevant to moral suasion hypothesis

	Econometric regression	NN SSC	NN $\hat{\Delta}^*$	XGBoost SSC	XGBoost $\hat{\Delta}^*$
$Domestic \times \Delta Tier1$	0.0195	0.000017	0.0068	-0.0003	-0.0033
$Domestic \times \Delta HighYield \times \Delta Tier1$	0.0020	-	-	-	-
$Domestic \times Government \times \Delta HighYield \times \Delta Tier1$	0.0100	-	-	-	-

Relevant coefficients for the flight-to-safety hypothesis estimated by the different methods. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Changes in variables are denoted as Δ *variable name*. $(Non-)Government$ is one when a security is (not) issued by a government. Likewise, $Domestic$ $(Foreign)$ is one when a variable is labelled domestic (foreign).

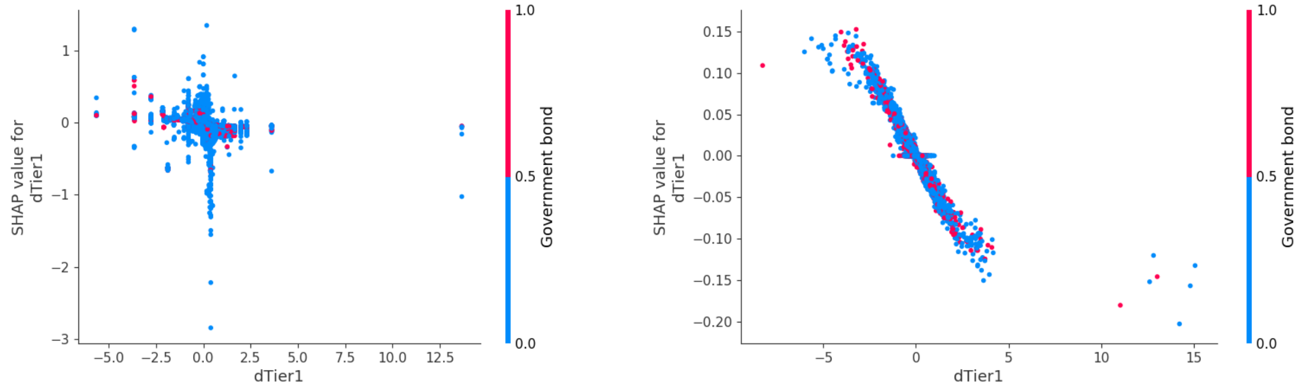
Table 5.6 contains the coefficients of the relevant variables. The econometric regression and the NN assign a positive effect to $Domestic \times \Delta Tier1$, which is in conflict with the hypothesis. On the other hand XGBoost finds a negative effect, which is inline with the hypothesis. Moreover, for $Domestic \times \Delta HighYield \times \Delta Tier1$ the econometric model has a positive coefficient. This means that when the capital decreases, the holdings in domestic securities that went from high- to low-yield increases, as for those securities $\Delta HighYield$ is negative. The positive coefficient of $Domestic \times Government \times \Delta HighYield \times \Delta Tier1$ implies that this effect is amplified when the securities are issued by the RBGs government. This could imply that there is indeed moral suasion.

However, as all coefficient are insignificant, one is unable to reject or confirm the moral suasion hypothesis. Hence, we conclude that the way a RGB adjusts its holdings for both domestic and foreign securities does not differ significantly.

5.2.3 Regulatory arbitrage

The last hypothesis, regulatory arbitrage, concerns the holdings of government securities. It implies that the coefficients of $Government \times \Delta Tier1$ and $Government \times \Delta HighYield \times \Delta Tier1$ are negative, as in this case a decrease in capital causes an increase in the holdings of government securities and high-yield government securities respectively. In order to analyze the results for this hypothesis, we again first consider the Shapley values themselves. Thereafter, the different SSC and coefficients for the variables of interest are compared.

Figure 5.6 shows the Shapley values for a change in the Tier 1 capital ratio. In both plots, the dots are red when the security is issued by a government and blue when it is not. Figure 5.6a show that the Shapley values for government securities are in general smaller than the Shapley values for non-government securities. There is no clear difference for the NN.



(a) Interaction $\Delta Tier1$ and $Government$ in XGBoost

(b) Interaction $\Delta Tier1$ and $Government$ in NN

Figure 5.6: Shapley values for a change in Tier 1 capital ratio (denoted as dTier1)

Table 5.7 shows the coefficients that are relevant to the regulatory arbitrage hypothesis. The results contradict each other. The econometric regression finds an effect opposite to what is expected from the hypothesis, as the coefficient is positive. The NN and XGBoost find an effect in line with the hypothesis. However, none of the results are significant. The results of the other interaction terms $Government \times \Delta HighYield \times \Delta Tier1$ and $Government \times Domestic \times \Delta HighYield \times \Delta Tier1$ are also not significant.

Hence, as with the moral suasion hypothesis, we are unable to reject or confirm the regulatory arbitrage hypothesis. Hence, we conclude that the way a RGB adjusts its holdings for both government and non-government securities does not differ significantly.

Table 5.7: Coefficients relevant to regulatory arbitrage hypothesis

	Econometric regression	NN SSC	NN $\hat{\Delta}^*$	XGBoost SSC	XGBoost $\hat{\Delta}^*$
$Government \times \Delta Tier1$	0.0394	-0.000008	-0.0380	0.0000	-0.0007
$Government \times \Delta HighYield \times \Delta Tier1$	0.0196	-	-	-	-
$Government \times Domestic \times \Delta HighYield \times \Delta Tier1$	0.0100	-	-	-	-

Relevant coefficients for the flight-to-safety hypothesis estimated by the different methods. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Changes in variables are denoted as Δ *variable name*. $(Non-)Government$ is one when a security is (not) issued by a government. Likewise, $Domestic$ $(Foreign)$ is one when a variable is labelled domestic (foreign).

Chapter 6

Endogeneity issues

The so far presented analyzes does not take endogeneity issues into consideration, even though this is an important topic in econometric research. As mentioned in Chapter 4 this is done in order to be able to compare the results of the machine learning methods with the econometric regression. However, since we know that results might change when we do account for endogeneity, this chapter elaborates on this matter.

The econometric field has lots of research on the endogeneity problem, as endogenous regressors makes OLS inconsistent. In our first differences regression, Equation (4.2), $\Delta Tier1$ might be endogenous, as we can argue that $Tier1$ in the level equation, Equation (4.1), is predetermined. $Tier1$ could be seen as predetermined, as a unpredictable error at t can mean that the RBG believes that the profits of this ISIN are going to be really high or really low. The first option will of course effect the Tier1, as the profit will be made in the future. The losses, however, are taken immediately, which could mean that Tier1 is even endogenous. This means that IV is needed in the first difference estimation, as predetermined variables in the level equation become endogenous variables in the first difference equation. Hence, all the interactions with $\Delta Tier1$ are also endogenous by construction.

Given this endogeneity issue, one needs instrumental variables as this enables us to obtain instrumental variable (IV) estimators. A benefit of panel data is that one can use internal instruments, i.e. lagged values of the regressors. However, this is only possible when there is no autocorrelation in the levels (Equation (4.1)). As stated in Appendix A.2.2, we have autocorrelation in the levels, hence we can only use external instruments.

Cohen and Scatigna (2016) investigate the increase in bank capital ratios. They describe several channels in which a bank can realize an increase in its risk-adjusted capital ratio. One way to increase the capital ratio is to increase profits. A bank could make attempts to increase its profits by increasing the interest rates it charges for loans and decreasing the rates payed for deposits. This would effect the demand for loans and the supply of deposits. Another way is to slow down lending growth or run down the loan portfolio, which allows the capital to increase. Hence, the loan and deposits could serve as an instrument for $\Delta Tier1$, as when the Tier 1 capital ratio of a bank changes, the loans and deposits of that bank might change as well. Instead of using the loans and deposits separately, we use the loan-to-deposit ratio as this is standardised for all banks and combines both effects. Moreover, we use the first lag as the change in loans and deposits are not simultaneously with the change in Tier 1 capital ratio. It can take a while before the effects of a change in loans and deposits let the capital change. Moreover, taking the lagged variables into account means that we have exogenous instruments as they are realized in the past. Note, that we cannot test this formally as we have exactly the same amount of instruments as we have endogenous variables.[1]

Table 6.1 shows the results from IV. The first column contains the normal regression without instruments for $\Delta Tier1$ and its interactions. The second and third column contain the results with IV without and with clustered standard errors respectively. The underidentification test rejects the null hypothesis of underidentification, i.e. instruments that are not relevant (p-value $= 0.0047 < 0.05$) Hence, the instruments are relevant. In general, the coefficients estimated with IV are (much) larger then the coefficients estimated without instruments. However, they remain insignificant. The following paragraphs elaborate on the variables that are relevant to our hypotheses, as we want to determine whether our results regarding those hypotheses are robust to endogenous misspecification.

Regarding the flight-to-safety hypothesis, we see that $\Delta HighYield \times \Delta Tier1$ is now positive instead of negative. Hence, this suports the hypothesis. However, it is still insignificant as are $Government \times \Delta HighYield \times \Delta Tier1$ and $Government \times \Delta HighYield \times \Delta Tier1$. Hence, the conclusion regarding the flight-to-safety hypothesis does not change.

The same holds for the second hypothesis regarding moral suasion, as the coefficients of $Domestic \times \Delta Tier1$, $Domestic \times \Delta HighYield \times \Delta Tier1$ and $Government \times Domestic \times \Delta HighYield \times \Delta Tier1$ are still insignificant. Note, however, that the first two have opposite signs.

Moreover, the conclusion regarding the third hypothesis of regulatory arbitrage does not change either, as $Government \times \Delta Tier1$ is still positive and insignificant when the appropriate standard errors are used. The other two relevant variables for this hypothesis, $Government \times \Delta HighYield \times \Delta Tier1$ and $Government \times \Delta HighYield \times \Delta Tier1$, remain also insignificant.

Since we now know that the instruments are relevant and valid, we test whether IV was really needed using a Hausman test, which is perform in two steps with clustered standard errors. The Hausman test indeed rejects the null hypothesis of exogeneity, hence the use of IV regression is needed. However, as we have seen from the results in Table 6.1 the conclusions do not change.

There is, to the best of our knowledge, no theory in the machine learning field regarding the problem of endogenous variables. Hence, there is not a known procedure to determine whether the conclusions based on machine learning techniques will change when there is an endogeneity

[1]Note that Van Lelyveld et al. (2019) also conduct IV analyzes. They use government profits as an instrument for the Tier 1 capital, which is logical, as they have the aggregated Tier 1 capital ratio over all banks in a country. However, we cannot use this as we have the Tier 1 capital ratio for specific RBGs which have multiple branches around the world.

issue nor a procedure to test whether endogeneity is an issue at all. Hence, we do not elaborate on this issue further as this is out of the scope of this thesis. Moreover, as our results are robust for correcting for endogeneity in the econometric model, we have reason to believe the same will hold for the machine learning models.

Table 6.1: IV estimations of the change in holdings: $\Delta Holdings$

	(1)	(2)	(3)
$\Delta Tier1$	-0.0653	-10.1869**	-10.1869
	(0.0457)	(5.1807)	(50.6847)
$\Delta HighYieldTotal$	-0.3143**	-0.7134	-0.7134
	(0.1520)	(0.5801)	(1.7401)
$\Delta HighYieldSector$	-0.1361	-0.7314	-0.7314
	(0.0879)	(0.4815)	(1.3717)
$\Delta HighYield \times \Delta Tier1$	-0.0012	0.1223	0.1223
	(0.0170)	(0.0810)	(0.1186)
$Government \times \Delta Tier1$	0.0394	6.9393**	6.9393
	(0.0315)	(3.2936)	(30.2181)
$Domestic \times \Delta Tier1$	0.0195	-0.0772	-0.0772
	(0.0417)	(1.0204)	(6.6250)
$Government \times \Delta HighYield \times \Delta Tier1$	0.0196	-0.1080	-0.1080
	(0.0137)	(0.0896)	(0.1108)
$Domestic \times \Delta HighYield \times \Delta Tier1$	0.0020	-0.0094	-0.0094
	(0.0208)	(0.0512)	(0.2889)
$Government \times Domestic \times \Delta HighYield \times \Delta Tier1$	0.0100	0.0637	0.0637
	(0.0237)	(0.1084)	(0.1624)
Observations	252,258	130,978	130,978
RMSE	6.22	7.98	7.98
Holder Area x Time	Yes	Yes	Yes
Issuer Sector x Time	Yes	Yes	Yes
Cluster robust errors	Yes	No	Yes

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. IV regression with fixed effects predicting the change in the holdings in a security. Changes in variables are denoted as Δ *variable name*. $HYTotal$ ($HYSector$) stands for the indicator function for $HighYieldTotal$ ($HighYieldSector$). HY is the interaction between both types of high-yield. GOV represents the indicator variable $Government$. Likewise, DOM stands for the indicator variable $Domestic$.

Chapter 7

Conclusion

This thesis has analyzed to what extent an RBG adjusts its holdings when its own capital changes. This was done by constructing three models, consisting of an econometric model, a neural network and a XGBoost model. By using Shapley framework, as proposed by Joseph (2019), we were able to construct coefficient for the machine learning methods, which can be compared to the econometric regression coefficients. However, we have argued that a shortcoming of this framework is that the sign of the coefficients is determined by an OLS regression instead of the machine learning method that is being analyzed. Hence, we have proposed a novel method based on numerically differentiating the output of the machine learning model. This method can also be used as framework to estimate coefficients for machine learning models.

Using the three models, we have analyzed three hypothesis which contribute to current literature into the matter. We have found some evidence in favour of the first hypothesis of the phenomena of flight-to-safety, which means that when the capital of a bank decreases, it rebalances towards safer securities, i.e. securities with a lower yield-to-maturity. We do not find significant evidence in favour nor against the moral suasion and regulatory arbitrage hypothesis, as the results of the different models are insignificant and in contradiction with each other. Hence, we can conclude that there is no statistical evidence for moral suasion, which implies that the way a RGB adjusts its holdings for either domestic or foreign securities, do not significantly differ from each other. Furthermore, we conclude that the way a RGB adjusts its holdings for both government and non-government securities, do also not differ significantly from each other, as there is no statistical evidence for the regulatory arbitrage hypothesis.

The remaining of this chapter contains a discussion on the presented analysis.Firstly, possible flaws in the data are discussed. Secondly, remarks are made on the estimation methods. Lastly, the concerns regarding the Shapley framework and the novel framework are mentioned.

The first problem with the data is that the bank specific data, e.g. Tier 1 capital ratio, is incomplete. As mentioned in chapter 3, we use a simple linear interpolation scheme to fill in the missing values. This is done, as otherwise too many observations are absent. However, this method of interpolation assumes a linear relation between the Tier 1 capital ratio of a bank over time. If this assumption does not hold, then the variable $\Delta Tier1$ and its interaction terms are a poor summary of reality, implying that the estimated effects are invalid. A more realistic way of filling in the missing values, could get the Tier 1 capital ratio closer to its true value. One could for example estimate a model to predict the Tier 1 capital ratio. Furthermore, besides using the Tier 1 capital ratio as a capital measure, we could also use a different measure, such as the leverage or the common Tier 1 capital ratio. Moreover, one could also make a distinction between the types of capital changes. These can be grouped into changes due to a change in regulation and changes due to market developments, as banks could reject differently in the two cases.

Subsequently, we discuss the obstacles in using machine learning model. A major concern regarding the machine learning models is that they can assign effect to variables that are unlikely to have effect. Analyzing the different models, we can clearly conclude that XGBoost performs best. The performance of the econometric benchmark regression is similar to that of the NN. However, using the SSC values we can conclude that the NN assigns the highest contribution to the $ISIN$ variable, which is the encoding for the ISINs. This is unexpected, as other ISIN specific variables, such as the price and maturity, are more meaningful. The problem is that the ISINs could not be given as dummies to the machine learning network, due to a memory limit. Hence, more research is required in the behaviour of machine learning methods in the case of omitted variables. This raises the question to what extent the obtained SSC values represent the actual coefficients. Moreover, it illustrates that we cannot blindly use machine learning techniques in economic or regulatory decision making.

Lastly, we discuss the concerns regrading the Shapley framework and the novel framework. Firstly, constructing the Shapley values requires a lot of time for the NN. In order to keep the time manageable, we could only use 100 background observations from which the variables that are not included in the coalition are integrated out with. These 100 values are constructed by applying the k-means algorithm on the validation set. However, this might not be the best way to obtain background observations. Other possible techniques, like Guassian Mixture Models, may be able to provide better results. Moreover, more research must be done in how to account for variables that are not included in the coalition.

Moreover, in the Shapley framework the sign of the SSC values are determined using an econometric regression. However, there is no check on whether the regression is correct according to econometric theory. For example, there could be omitted variable bias. When the omitted variables are included, the sign of the coefficients of the variables we are interested in, could change. Hence, more attention should be paid to get the right signs from econometric regression. Another possible solution is to determine the sign based on a approximation of the derivative in several points, as proposed in the novel framework. A benefit of this framework is that the sign can differ per method, which is preferred when we test hypotheses such as the ones in this thesis. A point of improvement in the novel framework is that values could be integrated out instead of taking the average. However, this would take much longer to compute.

Appendix A

Appendix

A.1 Data | Cleaning

This section describes the steps taken to clean the SHS-G data. The steps are the same as in Van Lelyveld et al. (2019). The data covers the two main types of security: debt securities and equity securities. The two security types each consist of two sub-types, namely short- and long-term debt securities (codes F31 and F32) and listed shares and investment fund shares (codes F511 and F52E). We focus on long-term debt (F32) and apply the following cleaning steps to the SHS data disseminated by the ECB.

1. Only keep ISIN (drop Cusip, etc.)
2. Drop if observation is not a stock position
3. Drop if stock position does not have a market value
4. Drop all holders that are not from the euro area
5. Exclude International organizations
6. Exclude short positions
7. Exclude third party holdings except households
8. Exclude FDI
9. Combine observations which are recorded twice in terms of functional category
10. Combine observations which are recorded twice in terms of source
11. Combine observations which are recorded twice in terms of reference area
12. Keep only euro denominated securities
13. Only keep market values, drop defaulted and estimated prices
14. Only keep bonds in case of F31 and F32
15. Issuer side: drop irrelevant sectors and aggregate the remaining ones to a partition of interest
16. Project-specic cleaning process

A.2 Models | Implementation details

A.2.1 Level estimation

In this section, the level equation is estimated in order to determine whether either fixed or random effects are present. The results for several estimations are presented in Table A.1. When we regress $Holdings$ only on $Tier1$, the estimated effect is positive, which means that a RBG invests more in ISINs when its Tier1 equity is higher. However, one omits meaningful variables in this case, such as whether the holdings are in a domestic, high yield and/or governmental bond. If those are included, the effect of extra Tier1 capital on the holdings becomes negative. Moreover, it seems that the holdings in governmental and domestic bonds are lower, while the holdings increase when a bond is labeled as high yield. This suggest that RBGs in general take on more risk.

However, it is likely that there are still omitted variables. In our model, those variables will be captured by fixed effects. Adding RBG specific effects, both RBG specific and time effects, ISIN specific effects, ISIN specific and RBG specific effects, ISIN specific and RBG specific and time effects do not change much besides the high yield coefficient becoming negative. However, in the last case the effect for Tier1 becomes insignificant at a 5% significance level.

Instead of including RBG specific and ISIN specific effects separately, one could also include effects for possible RBG and ISIN combinations. The effect captured by these effects could be divided into RBG specific, ISIN specific and RBG-ISIN combination specific effects. Note that there is a large increase in the number of included effects, namely nearly 2.5 times as much. The regression based on this framework, results in insignificant negative effects for Tier1. However, when time effects are also included, the effect of Tier1 becomes positive and significant.

As there are many reasons for including the effects either separately or in combination, it is unclear which model represents the underlying distribution best. However, we use the latter to test whether the effects are fixed or random, as it includes the first case with an extra combination effect.

To test the null hypothesis of random effects, the model with RBG-ISIN specific effects is used, as it already incorporates the RBG and ISIN specific effects. The null hypothesis is rejected with a Hausman test (p-value $< 10^{-4}$). Hence, a fixed effects panel data model should be used.

Table A.1: Estimation of the holdings in ISIN i by RBG r at time t

	(1)	(2)	(3)	(4)
Tier1	0.1192***	-0.0637***	-0.0472***	0.0201
	(0.0134)	(0.0133)	(0.0141)	(0.0145)
HYTotal		-1.1762***	-0.2375***	-0.2523***
		(0.1166)	(0.0563)	(0.0550)
HYSector		2.1454***	-0.1940***	-0.2249***
		(0.1166)	(0.0564)	(0.0550)
GOV		-3.9866***	-0.1585	-0.6342*
		(0.0720)	(0.3560)	(0.3523)
DOM		-6.7395***	-1.6169***	-1.6926***
		(0.0568)	(0.0371)	(0.4852)
Observations	525,526	525,526	511,773	484,721
R-squared	0.0001	0.0339	0.8634	0.9076
AIC	4637019	4618998	3455078	3103331
BIC	4637042	4619065	3455334	3103587
RMSE	19.94	19.60	7.41	6.67
RBG	No	No	Yes	No
Time	No	No	Yes	Yes
ISIN	No	No	Yes	No
RBG-ISIN	No	No	No	Yes

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Panel regression with fixed effects predicting the the holdings in a security. $Tier1$ is the Tier 1 capital ratio of a RBG. $HYTotal$ ($HYSector$) denotes the indicator function for $HighYieldTotal$ ($HighYieldSector$). GOV represents the indicator variable *Government*. Likewise, DOM denotes the indicator variable *Domestic*.

A.2.2 First differences estimation

In this section, the first difference model is constructed. The final results are presented in Chapter 5. This section starts by determining whether we can use OLS in the first difference model. Thereafter, we test if and which interaction terms should be included in the estimation. Subsequently, a look is taken at which time-effects should be included. In Chapter 5, we test for clustered standard errors.

As discussed in the methodology, we firstly test for fixed effects and i.i.d. errors in the level equation. Having random effects instead of fixed effects is formally rejected by a Hausman test (p-value $< 10^{-4}$, see Appendix A.2.1) for the equation in levels, Equation (4.1). Hence, a fixed effects panel data model should be used. Moreover, the LM test (Breusch-Godfrey test) and the Wooldridge test reject the null hypothesis of no serial correlation in the level equation. Hence, the errors can not be independently distributed and we can use pooled OLS instead of GLS in the first difference estimation.

In order to estimate Equation (4.2), we need to know which interaction terms should be included. The results without any interactions is given in the first column of Table A.2. The Ramsey RESET test rejects the null hypothesis of no omitted variables (p-value $= 0.0002 < 0.05$). Hence, the model is not linear in the included variables.

From theory, we know that factors of influence are given by the effects of (a change in) Tier1 in combination with an ISIN being domestic, high yield or governmental. However, there might be more combinations that have an effect. After testing which effects should be included, we take a look into the time-varying effects of the other control variables.

Firstly, we formally test which interactions have to be included in the model. The decision to include a series of interaction terms is based on the AIC or BIC measure and the joint significance F-test. The AIC and BIC measures combine both the fit and complexity of a model. Fit is measured negatively by -2*ln(likelihood); the larger the value, the worse the fit (negative of log-likelihood). Complexity is measured positively, either by 2*number of parameters (AIC) or ln(number of observations)*number of parameters (BIC).

We test the additional variables by first adding all first order interactions with the change in Tier1. Hence, we first add $\Delta HighYieldSector \Delta Tier1$, $\Delta HighYieldTotal \Delta Tier1$, $Government \Delta Tier1$ and $Domestic \Delta Tier1$ at the same time. Then we add interactions with two variables with respect to the change in Tier1. Lastly, we add the interaction consisting of all four variables interacted with the change in Tier1, hence $Government * Domestic * \Delta HighYield * \Delta Tier1$, where $\Delta HighYield = \Delta HighYieldTotal * \Delta HighYieldSector$ as defined in Chapter 3.

A summary of the results are shown in the second and third column of Table A.2. In each step the variables are added, based on the joint significance. The AIC value is in each step slightly lower, but the BIC value gets slightly higher. An explanation for this could be the fact that the BIC enforces a larger penalty for larger models. The interactions with the two metrics of the changes in high-yield are not significant. However, when both are combined, by simply multiplying them, some terms become jointly significant. Lastly, we look at different combinations of included interactions but all are jointly significant at a 10% level.

Subsequently, we look at which time-effects are missing, as they control for time-varying but cross-section invariant variables. We could simply include a date effect labelled γ_t. However, this might be restrictive as this means that the date-specific effect is common to all RBG-ISIN combinations. Another way for capturing the time-effects is by controlling for RBG Holder Area-date effect and/or a Issuer sector-date effect, denoted as $\gamma_{HolderArea,t}$ and $\gamma_{IssuerSector,t}$ respectively. The first effect captures the effects common for all RBGs in a specific country in a specific quarter, which is the same for all ISINs that the RBGs hold. The second includes the effect for a specific issuer sector in a specific quarter, which is the same for all RBGs.

Including both $\gamma_{HolderArea,t}$ and $\gamma_{IssuerSector,t}$ leads to the results in the last column of Table A.2. The changes are small, but the AIC and BIC are a bit lower.

Concluding, the fourth column of Table A.2 presents the final form of the estimated first difference equation. It shows the different interaction terms we include. Moreover, time-varying holder area and time-varying issuer sector effects are taken into account. The main results of this section are presented in Chapter 5. Moreover, we address there also a final concern about the used standard errors.

Table A.2: Estimation of the change in holdings in ISIN i by RBG r at time t

	(1)	(2)	(3)	(4)
$\Delta Tier1$	-0.0640***	-0.0825***	-0.0755***	-0.0653**
	(0.0153)	(0.0259)	(0.0260)	(0.0294)
$\Delta HYtotal$	-0.2457***	-0.2971***	-0.3026***	-0.3143***
	(0.0539)	(0.0757)	(0.0764)	(0.0766)
$\Delta HYsector$	-0.1616***	-0.1955***	-0.1388**	-0.1361*
	(0.0536)	(0.0661)	(0.0698)	(0.0698)
$\Delta HY \Delta Tier1$		0.0062	-0.0070	-0.0012
		(0.0067)	(0.0080)	(0.0080)
GOV$\Delta Tier1$		0.1029***	0.0907**	0.0394
		(0.0378)	(0.0380)	(0.0412)
DOM$\Delta Tier1$		-0.0096	-0.0164	0.0195
		(0.0317)	(0.0318)	(0.0329)
GOV$\Delta HY \Delta Tier1$			0.0243	0.0196
			(0.0156)	(0.0156)
DOM$\Delta HY \Delta Tier1$			0.0127*	0.0020
			(0.0065)	(0.0065)
GOV_DOM$\Delta HY \Delta Tier1$			0.0015	0.0100
			(0.0188)	(0.0187)
Observations	252,258	252,258	252,258	252,258
Adjusted R-squared	0.0004	0.0004	0.0004	0.0126
Adjusted Within R2	.	0.0004	0.0004	0.0004
AIC	1640939	1640937	1640930	1637670
BIC	1640981	1641010	1641035	1637774
RMSE	6.26	6.26	6.26	6.22
Subsample high outstanding	No	No	No	No
Holder Area x Time	No	No	No	Yes
Issuer Sector x Time	No	No	No	Yes

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Panel regression with fixed effects predicting the change in the holdings in a security. Changes in variables are denoted by Δ *variable name*. $Tier1$ is the Tier 1 capital ratio of a RBG. *HYTotal* (*HYSector*) stands for the indicator function for *HighYieldTotal* (*HighYieldSector*). *HY* is the interaction term between both types of high-yield. *GOV* represents the indicator variable *Government*. Likewise, *DOM* stands for the indicator variable *Domestic*.

A.2.3 Neural Networks

This section consists of the implementation details for the Neural Networks, which are implemented in python. To implement the NNs we use the package *keras*, which is based on the package *TensorFlow*. Thereafter, some results of the analysis of a NN with the sub-sample of the data which contain only the larger ISINS as we exclude the 5% smallest are given.

Construct NN using all the data

As discussed in the methodology, we start by making a model with the 3 linear layers in the end and one module in front of that, which result in three different NNs.

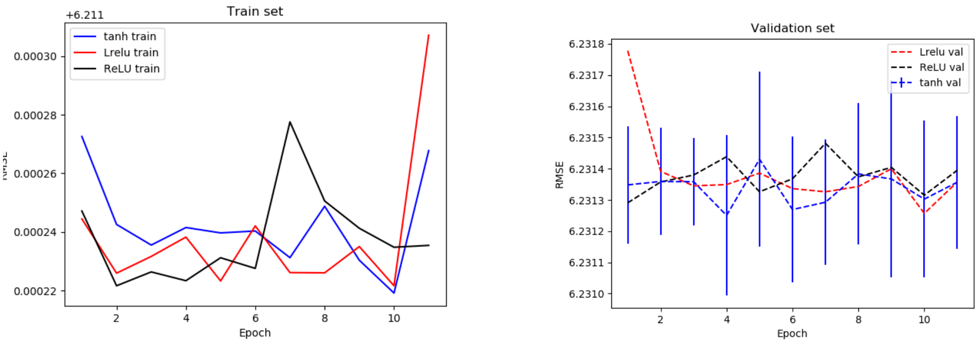
Each NN is trained a 100 times on the training data set. Hence, we estimate each NN 100 times. Per estimation, we impose a maximum of 30 epochs. One epoch entails that the NN has seen all training instances once. In each epoch, we train the NN with mini-batches of size 512, which means that we feed the NN 512 observations at a time and update the parameters after it has calculated the loss of those 512 observations[1]. At the end of each epoch the loss is calculated on the validation set.

We allow for early stopping, which means that if the mean squared error loss of the validation set changes less than 0.01 for five times in a row, then we do not go to the next epoch and the training stops. We then construct a new NN, which we train again for at most 30 epochs.

For all 3*100 NNs the validation and test RMSE are stored and in Figure A.1 the average RMSE per implementation of the NN is given. For the *tanh* module we also plot the standard errors for the validation RMSE. We see that the train RMSE is a bit lower than the validation RMSE, which is to be expected, as the NN should be better in explaining observations on which it is trained. Moreover, we can see from the validation RMSE that the NN with the *tanh* module produces the lowest RMSE in most epochs. Moreover, we can also see that the other errors fall within the 95% CI. Hence, the t-test would reject that there is a difference between the RMSEs of the models. As the NN with the *tanh* module has the lowest RMSE for nearly all epochs, we continue our analysis with this NN.

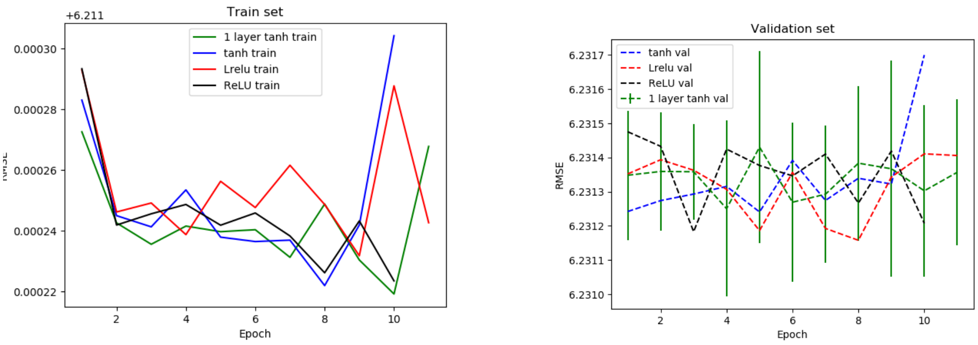
The procedure is repeated for a NN consisting of a *tanh* module, one extra module and the three end layers. The results are shown in Figure A.2. We have also plotted the validation RMSE of the NN with only a *tanh* module. As we can see the RMSE does not improve significantly, as most of the RMSEs fall within the 95% CI. Hence, we conclude that within the model specification we are looking into, the NN with one *tanh* module and three linear layers performs the best on the validation set.

[1] We could also update after each observation or after seeing all the data (hence once per epoch), but this is not preferred as the first case results in volatile updates and the last case has no stochasticity at all and assumes that the train data represents all the data correctly, which is quite unlikely. Hence, updating in mini-batches leads to the right amount of stochasticity.



(a) Train RMSE (b) Validation RMSE

Figure A.1: RMSE for the NNs with one module + 3 linear layers



(a) Train RMSE (b) Validation RMSE

Figure A.2: RMSE for the NNs with *tanh* module + extra module + 3 linear layers

Results from trained NN on normal data (not demeaned)

This section contains the results for the NN consisting of one *tanh* module followed by three linear layers. The model is trained and tested on the normal, i.e. not demeaned, dataset. The analysis that follows shows that the effects, SSC values, estimated by this model are dominated by ISIN specific variables. In order to get meaningful SSC values for the variables of interest, we could demean the data first. The results after demeaning are shown in Chapter 5.

The results from the previous section show that the RMSE on the train set is around 6.211 for the NN consisting of one *tanh* module followed by three linear layers. This is comparable with the RMSE for the econometric model, which is 6.22 evaluated on the training set.

Although the performance of the NN is comparable to the performance of the econometric benchmark regression, the driving variables behind the prediction are different. This can be seen from Table A.3, where the first column, SSC, is comparable with the coefficients from the econometric benchmark regression. The table only shows the economically relevant variables, hence the variables in which we are interested. However, the regression is performed with all the variables that the NN gets. The SSC values indicate that none of the variables we are interested in have any effect. Note that some of these SSCs are significant, but none of them are robust.

Table A.3: Estimations of the change in holdings $\Delta Holdings$ with a NN

Variable	Sign	SSC $\times 10^6$	SE $\times 10^7$	p-value	Robust
$\Delta Tier1$	+	2.6	8.0	0.4227	No
Government	+	2.9	8.0	0.0005	No
Non-Government	+	1.7	5.0	0.1917	No
$\Delta HYTotal\_down$	-	0.0	0.0	0.0000	No
$\Delta HYTotal\_same$	+	0.70	4.0	0.2062	No
$\Delta HYTotal\_up$	-	0.0	0.0	0.0000	No
$\Delta HYSector\_down$	+	0.0	0.0	0.0689	No
$\Delta HYSector\_same$	-	0.50	4.0	0.0053	No
$\Delta HYSector\_up$	-	0.0	0.0	0.0000	No
Domestic	-	2.0	6.0	0.1478	No
Foreign	-	2.6	9.0	0.2070	No

Importance of variables according to the Shapley framework applied to a NN. Changes in variables are denoted by Δ *variable name*. $HYTotal$ ($HYSector$) stands for the indicator function for $HighYieldTotal$ ($HighYieldSector$). Moreover, *down*, *up* and *same* mean that the security went from high-yield to low-yield, from low-yield to high-yield or it stayed the same respectively. ($Non-$)*Government* is indicator function for when a security is (not) issued by a government. Likewise, *Domestic* ($Foreign$) is an indicator function for when a variable is labelled domestic (foreign). The sign is determined on the basis of the novel framework. The SSC is constructed through the Shapley framework. The standard errors are given in the SE column. The p-value and robustness follow from a regression of the dependent variable on the Shapley values (step 2 of the Shapley framework).

As the SSC are not significant, we analyze the underlying Shapley values. To determine which variables do have a large effect on the prediction, we plot the Shapley values for all variables. Figure A.3 shows the results of a subset of these variables, including the variable with the larges absolute Shapley values, namely $ISIN$. This is a randomly selected encoding for all the ISINs.[2] This is quite striking as there is no consistent logical order in the numeration of the ISINs. Hence, it is strange that the NN can retrieve this much information from this variable and nothing from a more meaningful variable such as the change in Tier 1 capital ratio. Moreover, we have also included more meaningful variables to replace the indicator variables for each ISIN, such as the price of a ISIN and the yield to maturity. However, the NN still gets the highest Shapley values from the ISIN encoding. In order to stimulate the NN to find the effect of the variables we are interested in, we could first demean the data. The reasons behind this approach and the results are presented in Chapter 5.

[2]The ISIN codes contain numbers and letters, hence to include them in a regression all unique ISINs get a unique number.

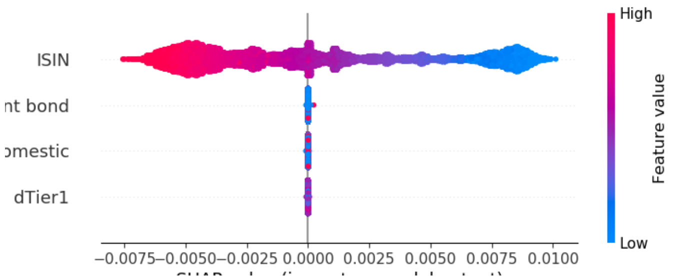


Figure A.3: Shapley values

A.2.4 XGBoost

This section consists of the implementation details for XGBoost, which is again implemented in python. We us the package *xgboost*. First the different hyper-parameters are given. Thereafter, we discuss tuning procedure and show immediately the results.

Parameters in XGBoost

Below the table that describes the parameters of XGBoost.

Table A.4: Hyper-parameters of XGBoost

Parameter	Default	Range	Description
η	0.3	$[0,1]$	Learning rate, this is the rate by which each prediction (i.e. weight) in a leaf is shrinked. Makes the boosting process more conservative.
γ	0	$[0,\infty)$	This is the gamma from the regularisation term Ω, thus it is the minimum gain a split needs to have. The higher the gamma, the more conservative XGBoost is.
maximum depth	6	$[0,\infty)$	This is the maximum depth of a tree. However, a value of 0 means there is no limit. The higher this value is, the more complex the model becomes. However, there is also more change of overfitting.
minimum child weight	1	$[0,\infty)$	This is the minimum sum of instance weight in a child. In this thesis linear regression is used, thus here this is the number of observations (instances) that have to be contained by a node. Thus if a leaf node has a sum that is less than this value, the algorithm will not continue splitting there.
subsample	1	$(0,1]$	This value is the ratio of the training observations that will be used in constructing a tree. XGBoost selects randomly the right number of observations. This prevents overfitting.
colsample by tree	1	(0, 1]	This value is the ratio of the variables that are used in constructing a tree. Again the variables are randomly selected.
λ	1		This is the λ from the regularisation term Ω. It is a correction on the predictions (weights) in a leaf. The higher this value, the more conservative the model becomes.
α	0		This is the alpha from the regularisation term Ω. It is a correction on the number of leafs. The higher this value, the more conservative the model becomes.
number of boosting rounds			The number of boosting rounds that are preformed. This is thus the number of trees that will be constructed.

Tuning procedure

This subsection describes the used tuning procedure of XGBoost and is inspired by the stepwise procedure of Xia et al. (2017). However, they use a logistic function and in this thesis, the linear function (*squared error*) is used, because this is the default and it is applicable since OLS can be used.

In the first tuning step, all parameters are left in the default settings, except the number of boosting rounds. The number of boosting rounds is determined by 3-fold cross-validation with the help of early stopping rounds (build in in *xgboost*). Figure A.4 shows the validation and training RMSE, where the validation set is a part of the original training set since we use cross-validation. We set the *early_stopping_rounds* criterion to 10, which means that the validation RMSE needs to improve at least once in every 10 rounds to continue training. As the training stops at 25, we choose $25 - 10 = 15$ boosting rounds.

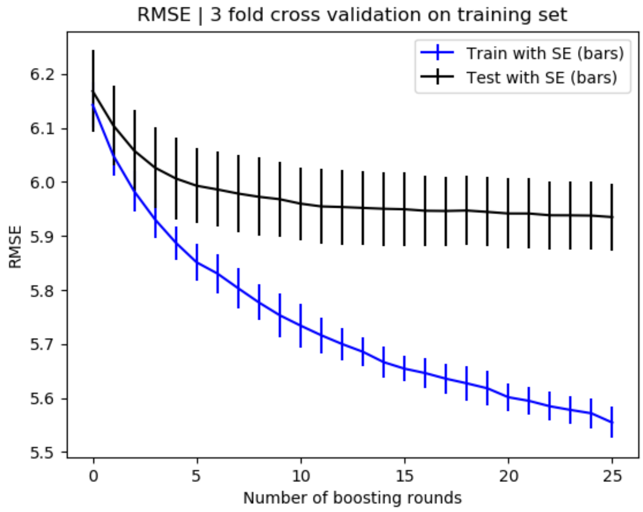


Figure A.4: RMSE during first tuning step

The second tuning step consists of tuning the other hyper-parameters, except the learning rate, λ and α. The learning rate is left in the default setting, since the learning rate influences the optimal number of boosting rounds. Moreover, λ and α from the regularisation term Ω are not adjusted, since γ (form the regularisation) is already altered. The other parameters are tuned by randomised search. This means that several randomly chosen sets of parameter values are tried, which might not result in the optimal set of parameter values. Moreover, because of the randomised search, the values that are chosen, differ each time, resulting in a different optimal set of parameter values. 3-fold cross-validation (from the *sklearn* package) is used to get the optimal values. We use randomised search, as other search methods consume to much time and computational power. This optimization scheme ultimately leads to a setting that is at or close to an optimum. The best parameters are: subsample = 0.9, minimum child weight = 8, maximum depth = 5, $\gamma = 0.9$ and colsample by tree = 0.6 (see Table A.4 for definitions).

In the last tuning step, the number of boosting rounds is again determined by 3-fold cross-validation with the help of early stopping rounds (build in in XGBoost). Figure A.5 shows the validation and train RMSE. We select 50 boosting rounds, as we use early stopping of 10 and the number of boosting rounds go until 60.

This is the procedure that is used in this thesis to tune the XGBoost models. The models are always tuned on the training datasets.

In the tuning steps, K-fold cross-validation is used in order to prevent overfitting. For K-fold cross-validation the training data is split in K distinct, equally sized subsets (folds). The model is each time t is trained on all K sets except number t. After training, it is tested on set t. This is done for all t in K. Then, it takes the average out-of-sample error as estimate for

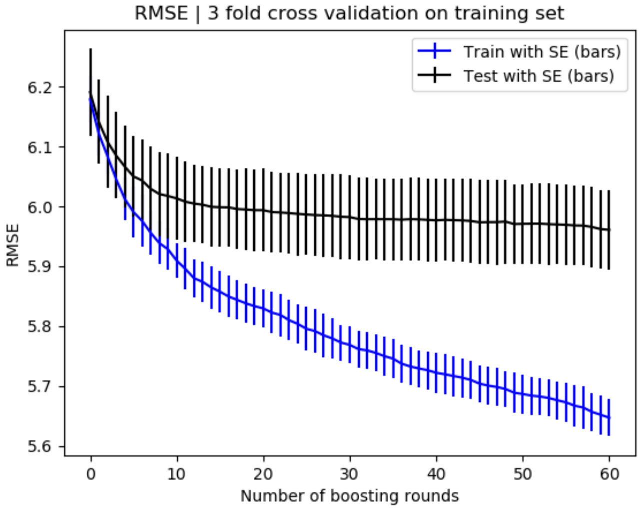


Figure A.5: RMSE during the last tuning step

the true out-of-sample rate. When tuning parameter values, it will choose the parameter values that minimize the estimate of the out-of-sample error.

Bibliography

Abbassi, P., Iyer, R., Peydró, J.-L., and Tous, F. R. (2016). Securities trading by banks and credit supply: Micro-evidence from the crisis. *Journal of Financial Economics*, 121(3):569–594.

Acharya, V., Drechsler, I., and Schnabl, P. (2014). A pyrrhic victory? Bank bailouts and sovereign credit risk. *Journal of Finance*, 69(6):2689–2739.

Acharya, V. V. and Steffen, S. (2015). The ”greatest” carry trade ever? Understanding eurozone bank risks. *Journal of Financial Economics*, 115(2):215–236.

Berk, R. A. (2008). *Statistical Learning from a Regression Perspective*, volume 14. Springer.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., and Simsek, O. (2019). Financial crisis prediction with machine learning and the (surprising) power of the yield curve. *Mimeo*.

Buch, C. M. and Prieto, E. (2014). Do better capitalized banks lend less? Long-run panel evidence from germany. *International Finance*, 17(1):1–23.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM.

Cohen, B. H. and Scatigna, M. (2016). Banks and capital requirements: channels of adjustment. *Journal of Banking & Finance*, 69:S56–S69.

De Jonghe, O. (2010). Back to the basics in banking? A micro-analysis of banking system stability. *Journal of Financial Intermediation*, 19(3):387–417.

Drechsler, I., Drechsel, T., Marques-Ibanez, D., and Schnabl, P. (2016). Who borrows from the lender of last resort? *Journal of Finance*, 71(5):1933–1974.

ECB (2015). Who holds what? New information on securities holdings. *Economic Bulletin*, 2:72–84.

Farhi, E. and Tirole, J. (2017). Deadly embrace: Sovereign and financial balance sheets doom loops. *Review of Economic Studies*, 85(3):1781–1823.

Giannetti, M. and Laeven, L. (2012). The flight home effect: Evidence from the syndicated loan market during financial crises. *Journal of Financial Economics*, 104(1):23–43.

Joseph, A. (2019). Shapley regressions: A framework for statistical inference on machine learning models. *arXiv preprint arXiv:1903.04209*.

King, M. R. (2010). Mapping capital and liquidity requirements to bank lending spreads.

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

LLC, M. (2019). Kapitaalvereisten voor de banksector. `https://www.consilium.europa.eu/nl/policies/banking-union/single-rulebook/capital-requirements/`.

Martinez-Miera, D. and Suarez, J. (2014). Banks' endogenous systemic risk taking. *Mimeo, CEMFI.*

Martynova, N. (2015). Effect of bank capital requirements on economic growth: a survey.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410.

Timmer, Y. (2018). Cyclical investment behavior across financial institutions. *Journal of Financial Economics*, 129(2):268–286.

Van Lelyveld, I., Mink, M., and Ramcharan, R. (2019). Banks, governments and distress. *Mimeo.*

Vinodhini, G. and Chandrasekaran, R. (2016). A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University-Computer and Information Sciences*, 28(1):2–12.

Xia, Y., Liu, C., Li, Y., and Liu, N. (2017). A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78:225–241.