HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES

# Predicting Sovereign Bond Spreads: Empirical Analyses using Machine Learning Techniques

Fakultät für Wirtschafts- und Sozialwissenschaften
der Hochschule Osnabrück
Studiengang "Angewandte Volkswirtschaftslehre"

## Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Arts

vorgelegt von

**Dennis Prüß**

| | |
|---|---|
| **Matrikelnummer:** | 1002800 |
| **Ausgabedatum:** | 10.01.2025 |
| **Abgabedatum:** | 28.02.2025 |
| **Erstprüfer:** | Prof. Dr. Markovic-Bredthauer |
| **Zweitprüfer:** | Prof. Dr. Skala |

# Abstract

In my bachelor thesis, I explored how machine learning (ML) techniques can predict 10-year sovereign bond spreads in the euro area. I used macroeconomic, financial market, and policy data from 2007 to 2023. My goal was to compare Random Forest and Support Vector Regression (SVR) against a simple linear model. Random Forest performed best, showing that credit risk (measured by CDS spreads) and unemployment rates strongly influence bond spreads. The SVR model also outperformed the linear model and highlighted other important factors like inflation and credit ratings. Both ML models outperformed the linear model, suggesting that bond spreads in the euro area follow non-linear patterns. Greece's special role in the European debt crisis and their high bond spreads during this time skewed the results. When removing Greece from the dataset, all models worked better and had lower error metrics. The SVR model benefited the most, shifting its focus to a broader range of predictors including country-specific factors and global financial indicators. Random Forest continued its emphasis on credit risk variables. My research demonstrates how ML techniques can help our understanding of financial markets by identifying complex relationships. This provides an alternative perspective beyond traditional econometric approaches.

**JEL Classification:** G12, C53, E44.

**Keywords:** Sovereign bond spreads, Machine learning, Random Forest, Support Vector Regression, Euro area.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **BIS** | Bank for International Settlements |
| **CBOE** | Chicago Board Options Exchange |
| **CDS** | Credit Default Swap |
| **COVID-19** | Coronavirus Disease 2019 |
| **ECB** | European Central Bank |
| **EMU** | European Monetary Union |
| **Euribor** | Euro InterBank Offered Rate |
| **FRED** | Federal Reserve Economic Data |
| **IQR** | Interquartile Range |
| **MAE** | Mean Absolute Error |
| **ML** | Machine Learning |
| **MSE** | Mean Squared Error |
| **OECD** | Organisation for Economic Co-operation and Development |
| **OMT** | Outright Monetary Transactions |
| **OOB** | Out-of-Bag Error |
| **PD** | Probability of Default |
| **QE** | Quantitative Easing |
| **REER** | Real Effective Exchange Rate |
| **RF** | Random Forest |
| **RMSE** | Root Mean Square Error |
| **RSS** | Residual Sum of Squares |
| **S&P** | Standard & Poor's |
| **SVM** | Support Vector Machine |
| **SVR** | Support Vector Regression |
| **VIX** | CBOE Volatility Index |
| **XGBoost** | Extreme Gradient Boosting |

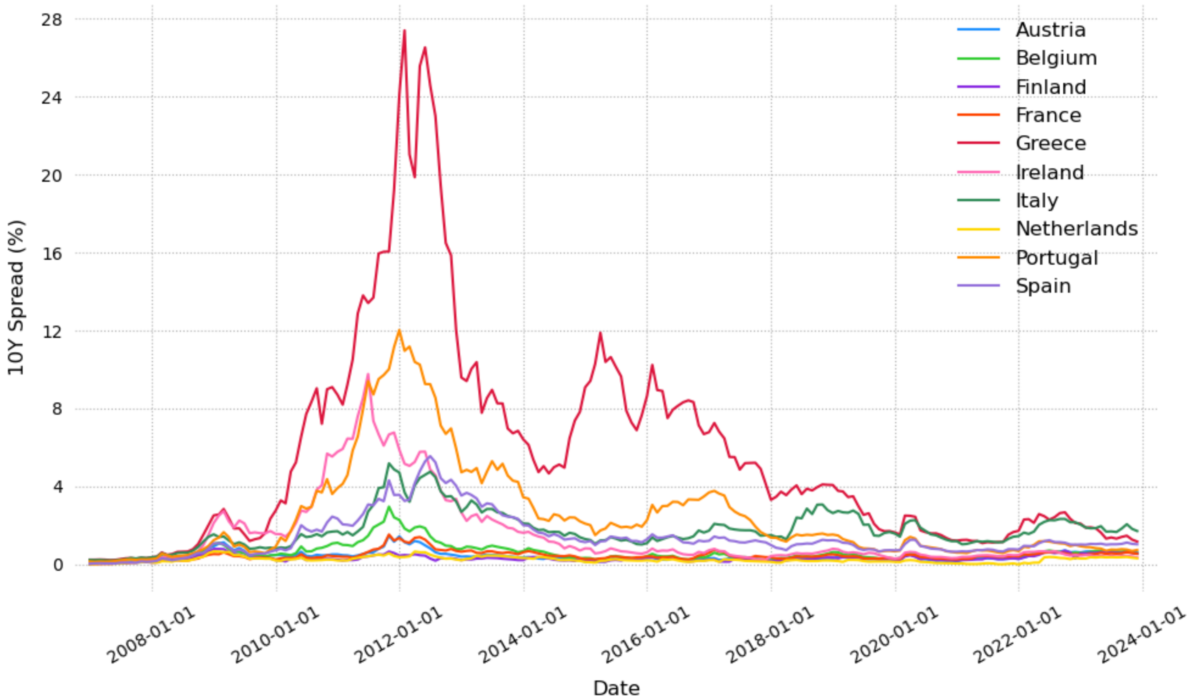
# 1 Introduction

Sovereign bond spreads refer to the difference in yields between a country's bond yield and those of a benchmark. They typically reflect the risk premium investors demand for holding the bonds compared to the benchmark bond. Within the euro area, these spreads are often measured against the German bonds, because they are considered almost risk-free. These spreads have led to great research on the factors driving them.[1] Government bond spreads matter to both policymakers and economists. They show that when public debt increases, borrowing costs also rise because yields increase as well. This is especially important in the European Union (EU) since member states cannot simply use inflation to lower their debt levels.[2] Since the Economic and Monetary Union (EMU) was created and the euro introduced in 1999, bond spreads in all member countries have stayed very low and have mostly been on a downward trend.[3] Removing exchange rate and inflation risks in the euro area led to lower yields. Borrowing costs fell along with liquidity and credit risks. After a period of convergence, markets stabilized and focused more on corporate bonds and other debt types. But after Lehman Brothers collapsed in 2008, focus shifted back to eurozone government bonds.[4] The global financial crisis deeply impacted the eurozone bond market. Economic conditions got worse and financial stress raised fears of rising deficits and debt. By 2009, bond yields surged, especially in countries with weak banking systems. The dependence on sovereign bonds became a serious concern for many countries.[5] The yield spread between Greek and German bonds reached almost 300 basis points in early 2009. One basis point equals 0.01%, so a 300 basis point spread means a 3% difference. By 2010, the spread had risen to over 1000 basis points (10%). Investors began to doubt if some EU governments could repay their debt. As a result, they demanded much higher risk premiums in the form of higher yields.[6] The debt crisis spread to other financially weak EU countries. Greece, Ireland, and Portugal lost access to market financing and had to rely on support from the European crisis resolution and the International Monetary Fund. Larger eurozone countries like Spain and Italy also faced

[1]cf. Matei and Cheptea, 2012, p. 2.
[2]cf. Bernoth, Von Hagen, and Schuknecht, 2012, pp. 975–976.
[3]cf. Pagano, 2004, p. 18.
[4]cf. Kilponen, Laakkonen, and Vilmunen, 2012, p. 3.
[5]cf. Matei and Cheptea, 2012, p. 2.
[6]cf. Costantini, Fragetta, and Melina, 2014, p. 337.

rising bond yields, which made it harder for them to refinance their debt.[7] In a speech on July 26, 2012, Mario Draghi, then President of the European Central Bank (ECB), pledged to do "whatever it takes" to protect the euro. This speech and the announcement of the Outright Monetary Transactions (OMT) programme significantly reduced sovereign bond yields in less stable countries of the EU.[8] After the European debt crisis, quantitative easing (QE) by the ECB played a key role in stabilizing sovereign bond markets and improving fundamentals in the euro area. It helped boost gross domestic product (GDP) growth, changed current account deficits into surpluses, and kept debt-to-GDP ratios stable.[9] Over a decade after the European debt crisis and the interventions of the ECB with unconventional actions like QE, eurozone countries still face different sovereign bond yields. After the COVID-19 pandemic, euro area countries had to refinance large amounts of debt. This was difficult because the economy was weak and market confidence was generally low. Countries had to take on more debt to deal with social problems.[10] Figure 1.1 shows the 10-Year bond spreads over the period 2007 to 2023 for the ten EU countries in this analysis. Greece is a notable exception when compared to the other EU nations, due to the sovereign debt crisis.

Figure 1.1: 10-Year Sovereign Bond Spreads (against German)




Source: Own illustration based on data from OECD

Machine Learning (ML) is very good at predicting asset prices, which is key to forecasting

[7] cf. Kilponen, Laakkonen, and Vilmunen, 2012, p. 4.
[8] cf. Arakelian et al., 2019, p. 1.
[9] cf. De Grauwe, Ji, and Macchiarelli, 2017, Chapter 4.
[10] cf. Belly et al., 2022, p. 1.

risk premiums. Traditional econometric methods struggle with many or too correlated variables. ML methods can handle complex data well by selecting important variables and reducing dimensions. They can also capture non-linear relationships, which gives ML methods an advantage over traditional linear models.[11] This paper analyzes and predicts sovereign bond spreads for ten European countries. Germany is used as the benchmark. This paper uses data from 2007 to 2023 and covers multiple key events like the Great Financial Crisis, the European debt crisis, and the COVID-19 pandemic. It includes macroeconomic data, financial market data, and economic policy data, to identify the main factors that influence sovereign bond spreads in Europe. ML methods, specifically Random Forest (RF) and Support Vector Regression (SVR), are used to predict and analyze key factors in spreads. This bachelor thesis contributes to the research of finance, economics, and data science. It shows how ML can provide new insights into sovereign bond markets and help better explain bond spreads.

[11] cf. Gu, Kelly, and Xiu, 2018, pp. 3–4.

# 2 Literature Review

Sovereign bond spreads have been widely studied in the economic and financial literature. The focus was mainly on their key determinants and the role of fiscal, macroeconomic, and market risk factors. Over time, the relationship between spreads and their factors has evolved. Major financial crises and monetary policy interventions had their influence. This chapter shows how research in this field has changed over time. I start with traditional econometric methods and later including ML techniques. Bernoth, Von Hagen, and Schuknecht (2012) studied bond yield spreads in 15 European countries from 1993 to 2009. They found that government debt issues, liquidity risks, and market sentiment were the main factors explaining bond spreads. After the EMU was introduced, investors paid less attention to total debt levels and focused more on budget deficits and debt payments.[1] Building on this, Matei and Cheptea (2012) found similar results in their study of sovereign bond spreads in 25 EU countries from 2003 to 2010. They showed that countries with better government budgets pay lower borrowing costs. They also highlight that higher GDP growth leads to lower bond spreads. Therefore, a stronger economy can decrease the bond yields.[2] Despite these similar findings, research pointed to a significant shift in investor behavior. Afonso, Arghyrou, and Kontonikas (2012) studied bond spreads from 1999 to 2010 and found that before the global financial crisis, economic and fiscal factors had little impact. From the middle of 2007 onward, these factors became more important. Investors started paying more attention to debt levels, liquidity issues, and risks related to bond maturity when evaluating sovereign yields.[3] The 2008 great financial crisis and later the European debt crisis changed how investors evaluated sovereign bond yields and their spreads. Favero and Missale (2012) studied bond spreads from 2006 to 2011 and found that government debt and deficits affect spreads in a complex way. Instead of being judged on their own, a country's spreads were compared to those of other nations. This meant weaker economies faced more pressure if they were close to struggling countries. The study also showed that market sentiment and the spread of financial stress made sovereign spreads higher than what economic factors alone would suggest.[4] Georgoutsos and Migiakis (2013) showed that the factors affect-

[1] cf. Bernoth, Von Hagen, and Schuknecht, 2012, pp. 982–983.
[2] cf. Matei and Cheptea, 2012, p. 17.
[3] cf. Antonio Afonso, Arghyrou, and Kontonikas, 2012, p. 23.
[4] cf. Favero and Missale, 2012, pp. 14–15.

ing sovereign bond spreads vary across Eurozone countries. Analyzing data from 1999 to 2011, they find that no single explanation always applied. This is especially the case before and after the 2008 crisis. They highlight that market sentiment and economic confidence strongly influence how investors viewed sovereign spreads. But the impact differs between countries.[5] As the crisis deepened, policy responses began playing a more central role. Kilponen, Laakkonen, and Vilmunen (2012) focused on how ECB policies affected bond markets from 2007 to 2012. They found that higher credit risk, measured by credit default swap (CDS) spreads, and liquidity risks, measured by bid-ask spreads, increased bond yields. However, in countries like Ireland and Portugal, ECB actions helped improve liquidity. Their study suggests that ECB interventions reduced financial stress in some crisis-heavy nations. Still, sovereign risk was mainly driven by credit risk rather than overall market sentiment.[6] Taking a broader view of policy impacts, Afonso, Jalles, and Kazemi (2020), studied data from 1999 to 2016 to see how policy announcements affected bond spreads. They found that bad news, like an increase in the debt-to-GDP ratio, increased spreads, while good news decreased them. ECB actions, like interest rate changes, bond refinancing announcements, and the first bond purchase program, helped stabilized bond markets. Their study showed that monetary policy has been highly influential in determining sovereign spreads, especially during crises.[7] Monetary policy has become more important in sovereign bond markets after the great financial crisis. De Grauwe, Ji, and Macchiarelli (2017) studied how ECB actions affected bond spreads from 2000 to 2017. They found that factors like debt-to-GDP ratios, economic growth, and trade balances mattered, but policy measures such as the OMT program in 2012 and QE in 2015 helped lower spreads and stabilized financial markets.[8] This shift in market dynamics was further explored in smaller country-specific analyses. Guirola and Pérez (2023) studied how sovereign bond spreads relate to debt levels in Spain, France, and Italy from 1980 to 2019. They found that before 2012, higher debt-to-GDP ratios led to higher spreads. However, after 2012 investors became more tolerant of high debt levels. This is likely because of monetary policy support. The study also showed that each country has a different level of debt tolerance. The impact of debt on spreads therefore varies between nations.[9] Eijffinger and Pieterse-Bloem (2023) analyzed Eurozone sovereign spreads from 1999 to 2021. Their results suggest that macroeconomic fundamentals alone cannot consistently explain spreads. Market risk indicators such as CDS spreads, bid-ask spreads, and ECB policy measures provide a more stable explanatory model. Over time, ECB monetary policy has increasingly shaped sovereign bond pricing, particularly through its

[5] cf. Georgoutsos and Migiakis, 2013, p. 4663.
[6] cf. Kilponen, Laakkonen, and Vilmunen, 2012, p. 15.
[7] cf. António Afonso, Jalles, and Kazemi, 2020, p. 11.
[8] cf. De Grauwe, Ji, and Macchiarelli, 2017, Chapter 4.
[9] cf. Guirola and Pérez, 2023, pp. 3974–3976.

QE programs.[10] Traditional econometric models have been widely used to study sovereign bond spreads, but recent research looked at how ML can improve predictions. Arakelian et al. (2019) studied sovereign risk in Europe from 2008 to 2017 using CDS spreads as the dependent variable. They grouped countries into three risk levels based on debt levels, unemployment, and financial contagion. During the crisis (2008-2013), market contagion had a big impact, but after 2013, debt-to-GDP and unemployment became more important factors. They found that QE helped lower CDS spreads by reducing contagion effects. They also showed that ML models like RF predict sovereign risk better than traditional econometric models.[11] Unlike this study, which focuses on CDS spreads as the dependent variable, my research looks at 10-year bond spreads against Germany. Belly et al. (2022) compared ML methods with traditional models in 10 EU countries from 2004 to 2019. They found that ML models, especially Extreme Gradient Boosting (XGBoost) and RF, made better forecasts of sovereign spreads. They also showed that new data sources, like Google Trends, Quanto CDS spreads (which measure redenomination risk), and ECB speech sentiment, offer valuable insights. Their study highlighted that investor expectations, central bank speeches, and fears of currency escape impact bond spreads, along with traditional economic factors.[12] Overall, the existing literature on sovereign bond spreads in europe showed that the factors have changed over time. Government finances and macroeconomic factors have always played a role. However, their impact had changed over time. The great financial crisis changed the most important determinants of spreads. Market sentiment, liquidity risk, and sovereign credit risk were more important than macroeconomic factors during and after the great financial crisis. In recent years, ECB policies have become a main factor in determining bond spreads across europe. Unconventional policies like asset purchase programs, have stabilized the markets. ML methods have become more interesting because they can capture non-linear patterns and complex relationships better than traditional methods.

[10]cf. Eijffinger and Pieterse-Bloem, 2023, p. 14.
[11]cf. Arakelian et al., 2019, p. 21.
[12]cf. Belly et al., 2022, p. 34.

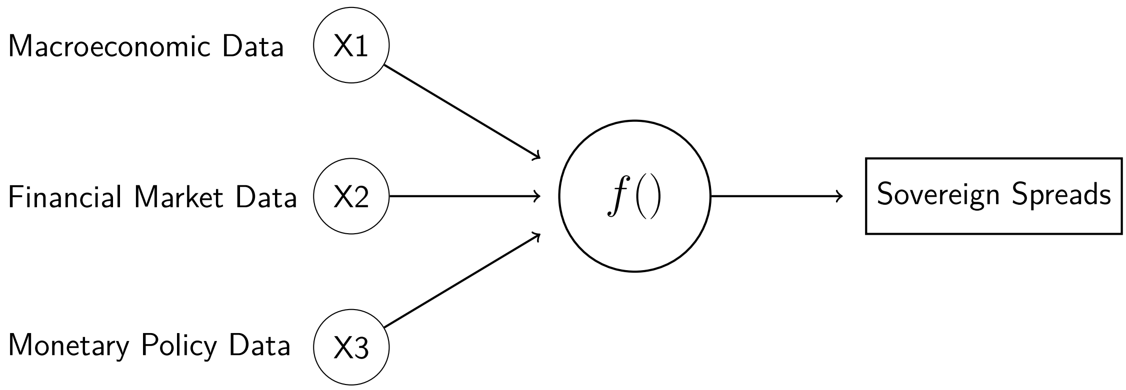
# 3 Methodology

In this chapter, I describe the data, the ML methods, and the statistical framework used in the models. The dataset includes macroeconomic indicators, financial market data, and monetary policy variables. This data helps to estimate a function $f()$ that links the inputs to the target variable. Figure 3.1 shows this process. The general model is written as

$$Y = f(X) + \epsilon \tag{3.1}$$

Here, $Y$ represents the target variable, the sovereign bond spreads. The function $f(X)$ represents the unknown relationship between the predictor variables $X_1 ... X_p$ and $Y$. The term $\epsilon$ captures the irreducible error due to unmeasurable variations.[1] The main goal of this analysis is to use RF and SVR as the ML techniques to model the relationship between the input variables and the target variable. The full analysis is performed in Python, mainly using the scikit-learn library for the ML models and statistical methods.[2]

Figure 3.1: Diagram representing the data flow for sovereign spread model




Source: Own illustration

This analysis answers three main questions. First, I check how well RF and SVR predict sovereign bond spreads in ten European countries from 2007 to 2023. Second, I find the key factors that influence bond spreads. Third, I test how removing Greece, which had a major sovereign debt crisis, changes the model's accuracy and the importance of different factors. I propose three hypotheses. **H1:** RF and SVR (RBF) will predict bond spreads

[1] cf. James et al., 2023, p. 15.
[2] cf. Pedregosa et al., 2011.

well because they capture complex patterns. I compare them to a simple linear model to see if ML methods improve accuracy. **H2:** Important macroeconomic indicators, such as debt-to-GDP, GDP growth, and inflation, will strongly affect bond spreads. These should rank high in the RF and SVR models. **H3:** Removing Greece improves predictions and changes which factors matter most, since Greece had unique debt issues. First, time series forecasting tests how well RF and SVR predict bond spreads over time, compared to a linear model. Second, variable importance measures identify the key factors affecting bond spreads. Third, I compare the results with and without Greece in the dataset to see how it influences predictions. The next sections explain the dataset, the ML models, and the statistical framework.

## 3.1 Data

The dataset used for this analysis includes 151 independent variables. The data was collected from a few different sources, including the S&P Capital IQ Platform provided by the University, the ECB Data Portal, the Organization for Economic Cooperation and Development (OECD), the World Bank, the Federal Reserve Economic Data Portal (FRED), the BIS Data Portal from the Bank for International Settlements (BIS), and the World Governments Bonds Website. A detailed overview of all variables, their descriptions, frequencies, and links to the data sources is provided in Table A.1 in the appendix. The dataset includes the period from January 2007 to December 2023 on a monthly frequency. In total, the dataset has 2040 observations in a panel form and covers about 204 monthly observations for 10 countries. The period used for this analysis is based on a trade-off between data availability and accessible data sources. This time frame provides a long enough series of data to support a solid empirical analysis without losing variables for some countries.

### 3.1.1 Dependent variable: sovereign bond spreads

The dependent variable is the 10-year sovereign bond spreads of 10 euro area countries including Austria, Belgium, Finland, France, Greece, Ireland, Italy, the Netherlands, Portugal, and Spain. The spread is calculated against the 10-year German government bond yield. The German bond yield for the 10-year serves in this paper as the benchmark and as a risk-free rate. Table 3.1 contains the descriptive statistics of the dependent variable.

Table 3.1: Descriptive statistics of the euro area sovereign bond spreads in percentage points from January 2007 - December 2023

| Country | Mean | Median | St. Dev. | Min | Max | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| Austria | 0.42 | 0.35 | 0.25 | 0.04 | 1.49 | 1.58 | 3.26 |
| Belgium | 0.61 | 0.47 | 0.45 | 0.04 | 2.97 | 2.33 | 6.75 |
| Finland | 0.32 | 0.30 | 0.15 | 0.03 | 0.80 | 0.87 | 0.58 |
| France | 0.46 | 0.42 | 0.24 | 0.04 | 1.54 | 1.73 | 4.90 |
| Greece | 5.58 | 3.76 | 5.51 | 0.23 | 27.39 | 1.81 | 3.58 |
| Ireland | 1.51 | 0.65 | 1.88 | 0.02 | 9.77 | 2.10 | 3.87 |
| Italy | 1.78 | 1.60 | 0.97 | 0.21 | 5.19 | 1.12 | 1.67 |
| Netherlands | 0.25 | 0.22 | 0.15 | 0.01 | 0.68 | 0.74 | 0.45 |
| Portugal | 2.47 | 1.49 | 2.58 | 0.14 | 12.03 | 1.91 | 3.21 |
| Spain | 1.41 | 1.12 | 1.08 | 0.05 | 5.56 | 1.69 | 2.76 |

Source: Own calculations based on data from OECD

### 3.1.2 Independent variables

The independent variables can be classified into three categories, which are macroeconomic data, financial market data, and monetary policy data. I will discuss each category and the variables they include. A further detailed description of the data, frequency, and the data sources can be found in table A.1 in the appendix.

**Macroeconomic data** has been an important factor in most findings related to the analysis of spreads, as highlighted in the literature review. For all countries, I included the following variables: the inflation rate, the unemployment rate, retail sales, real GDP growth rate, current account balance as a percentage of GDP, debt-to-GDP ratio, debt growth rate, the consumer confidence index, and the political stability and absence of violence/terrorism index. I also calculated for each variable their difference to the benchmark, Germany. Additionally, all macroeconomic variables were included with lags of 1, 3, and 12 months.

**Financial market data** includes the monthly and yearly percentage changes in stock market returns for each country's main index. These numbers show how well each country's equity market is performing and its current trends. The Chicago Board Options Exchange (CBOE) Volatility Index (VIX) was also included. The VIX measures the expected 30-day volatility of the United States (U.S.) S&P 500 stock market Index by using call and put options. It is widely used in studies on euro area government bond spreads and serves as an indicator of global risk.[3] Additional market sentiment is captured by two indicators: the U.S. yield spread and the Baa corporate bond spread. The U.S. yield spread is the difference between the 10-year and 1-year Treasury yields. The

[3] cf. Beber, Brandt, and Kavajecz, 2009, p. 945.

Baa corporate bond spread is the difference between Moody's Seasoned Baa Corporate Bonds and the 10-year U.S. Treasury yield.[4] The real effective exchange rate (REER) is also included and measures a country's inflation-adjusted currency value relative to its trading partners. REER reflects shifts in competitiveness and can serve as an early warning indicator. A real appreciation can harm competitiveness, leading to current account deficits and debt issues.[5] For credit risk, the Credit Default Swap (CDS) spreads and the sovereign credit ratings are among the most common variables, which I included. CDS data represent the annual insurance premium (in basis points) that buyers pay to protect against a default. Essentially, a CDS transfers default risk from the buyer to the seller. If a default occurs, the seller compensates the buyer for the loss. Higher CDS spreads signal increasing perceived credit risk.[6] De Grauwe et al. (2017) argue that CDS spreads should not be treated as independent variables in sovereign bond pricing models. They point out that CDS spreads and bond spreads tend to rise together during crises and are driven by increased market fear and uncertainty. This means CDS spreads are not truly separate from the factors driving bond spreads but are instead influenced by the same market conditions. As a result, adding CDS spreads to a model may make the numbers look better statistically but doesn't necessarily improve the actual understanding of bond spreads.[7] Fontana & Scheicher (2016) investigated whether the same factors drive both CDS spreads and bond spreads, using a panel regression analysis. They show that CDS spreads contain important credit market information and can help explain sovereign bond spreads. Their results suggest that while both CDS spreads and bond spreads respond to credit market conditions, the effect is stronger for CDS spreads. They find that the iTraxx corporate index, a key measure of credit risk in Europe, is consistently significant in explaining CDS movements. This suggests that CDS spreads reflect broader credit risk dynamics beyond what is captured in bond spreads. These findings support using CDS spreads as an independent variable in bond pricing models.[8] Several studies have included CDS spreads in their analysis of sovereign bonds as an independent variable. Beber et al. (2009) analyze the determinants of sovereign bond yield spreads in the euro area and focus on the roles of credit quality and liquidity. To measure credit risk, they use CDS data, stating that they "obtain an exogenous estimate of credit quality for each of the countries in the sample."[9] Similarly, Eijffinger and Pieterse-Bloem (2023) include CDS spreads in their analysis, arguing that CDS spreads are a key proxy for credit risk.[10] To address potential endogeneity concerns, I use only 3, 6, and 12-month lagged CDS spreads as independent variables. This ensures that credit risk affects bond spreads rather than

[4] cf. Eijffinger and Pieterse-Bloem, 2023, pp. 5–6.
[5] cf. De Grauwe, Ji, and Macchiarelli, 2017, Chapter 3.3.
[6] cf. Fontana and Scheicher, 2016, pp. 8–9.
[7] cf. De Grauwe, Ji, and Macchiarelli, 2017, Chapter 3.3.
[8] cf. Fontana and Scheicher, 2016, p. 17.
[9] Beber, Brandt, and Kavajecz, 2009, p. 926.
[10] cf. Eijffinger and Pieterse-Bloem, 2023, p. 12.

the other way around. Using lagged values helps reduce the risk of reverse causality and bias in the regression model.[11] The CDS spreads are measured relative to the German CDS as a benchmark. To improve clarity, the scale is converted from basis points to percentage points. Another key measure of credit risk is sovereign credit ratings, which are given by credit rating agencies such as Moody's, Standard & Poor's (S&P), and Fitch. To measure this into the regression model, I took the average ratings of the three rating agencies and converted the rating into a numeric scale from 1 (AAA, highest) to 22 (D, default). Additionally, I calculated the annualized probabilities of default (PD) for each country. The PD values originate from a Credit Benchmark White Paper (2016). In Appendix 2 the PD values are calculated using an average of credit risk estimates from at least three banks. These values are then mapped to Credit Benchmark Consensus breakpoints and are similar to the credit ratings.[12] The exact mapping is in Table A.3 in the appendix. As with the CDS data, I have also converted these values to percentage points for consistency.

**Monetary policy data** includes both conventional and unconventional policies. To measure the interest rate, I use the 3-month Euro Interbank Offered Rate (Euribor). The Euribor is the average interest rate at which leading European banks lend unsecured funds to one another in the interbank market.[13] The Euribor is both in percentage and its first difference included. To measure the interest rate differential between the U.S. and the Eurozone, the difference between the Federal Reserve's interest rate and the ECB's main interest rate was calculated and included as a variable. For unconventional policy data, I use the month-to-month change in Eurosystem total assets, both in absolute terms and as a percentage change relative to the previous month. The Eurosystem total assets change should reflect the impact of QE. To capture the impact of QE more precisely, the total net ECB asset purchases per month were included as an independent variable. This combines purchases from the public sector purchase programme and the pandemic emergency purchase programme. The securities markets programme was not included and the OMT programme was left out since it was never used.

**Dummy variables and interaction terms:** As Eijffinger and Pieterse-Bloem (2023) highlight, regime shifts play an important role in the movement of sovereign bond spreads in the euro area. They identify three key periods: (1) 1999–2010, when markets stabilized after the euro was introduced; (2) 2010–2013, when the sovereign debt crisis increased risks and led to ECB rescue programs; and (3) 2014–2021, when the ECB used negative interest rates and quantitative easing to prevent financial instability.[14] Therefore, I constructed dummy variables for these regimes to help my ML models better capture and

[11] cf. Fontana and Scheicher, 2016, p. 23.
[12] cf. Credit Benchmark, 2016, p. 15.
[13] cf. European Central Bank, 2013, p. 71.
[14] cf. Eijffinger and Pieterse-Bloem, 2023, p. 10.

understand their effects. In addition, I created interaction terms for each dummy variable to improve the analysis. For the pre-crisis period, I multiplied the dummy variable with the difference between debt and GDP compared to Germany to highlight the role of macroeconomic and fiscal dominance. For the period of the sovereign debt crisis, I interacted the dummy with debt growth to reflect the increasing concern about sovereign debt. Finally, for the post-crisis period, I used the QE dummy with the inflation rate to capture the relationship between monetary policy interventions and inflation dynamics.

## 3.2 Machine Learning Models

### 3.2.1 Random Forest

To understand the RF algorithm I use in my analysis, it's important to understand the basics of them: decision trees. Decision trees are a simple and powerful tools for making predictions. They work by dividing the data into rectangular regions ($R$), with each region getting its own simple model. In a regression tree, the process starts by splitting the entire space into two regions. In each region the predicted value is simply the mean of the target variable, $Y$. The best split is chosen by finding the variable and split-point that give the best fit. Then, one or both of these regions are split further. This process continues until a stopping rule is reached.[15] Terminal nodes, also known as leaves, are where the final predictions are made. They can be seen at the bottom of the tree. A decision tree is usually drawn with the root at the top and the leaves at the bottom. The tree splits the data at points called internal nodes. The lines connecting the nodes are called branches. Branches show the path from the top of the tree to the leaves.[16] "Various algorithmic rules are used to decide which variables to split and which splitting value to take at each step of the tree's construction."[17] In theory, regions may have any shape. For simplicity and clearer interpretation, rectangular regions are chosen. The objective is to identify regions that minimize the residual sum of squares (RSS). The RSS is defined as

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \tag{3.2}$$

This equation measures the total error in a regression tree. It adds up the squared differences between actual values ($y_i$) and the average prediction ($\hat{y}_{R_j}$) in each region ($R_j$). The goal is to make this error as small as possible, so the tree makes better predictions. Because it is not practical to consider every possible way to split the data, a top-down method called recursive binary splitting is used. The best split is chosen at

[15] cf. T. Hastie, Tibshirani, and Friedman, 2009, p. 305.
[16] cf. James et al., 2023, p. 333.
[17] Efron and T. J. Hastie, 2016, p. 126.

each step by selecting the predictor and cut-point that lead to the largest reduction in RSS.[18] "For each splitting variable, the determination of the split point $(s)$ can be done very quickly and hence by scanning through all of the inputs, determination of the best pair $(j, s)$ is feasible."[19] Although this method can fit the training data very well, it often overfits and performs poorly on new data. To reduce overfitting, a common approach is to first grow a very large tree and then prune it back to a smaller subtree. This method is known as cost complexity pruning.[20] A tuning parameter $\alpha$ controls how much pruning is done. Greater values for $\alpha$ produce a smaller tree, while a lower $\alpha$ results in a larger tree. Cost complexity pruning removes one less useful internal node at a time.[21] Rather than considering every possible subtree, it considers a number of trees indexed by the tuning parameter $\alpha$. The equation

$$\sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha|T| \tag{3.3}$$

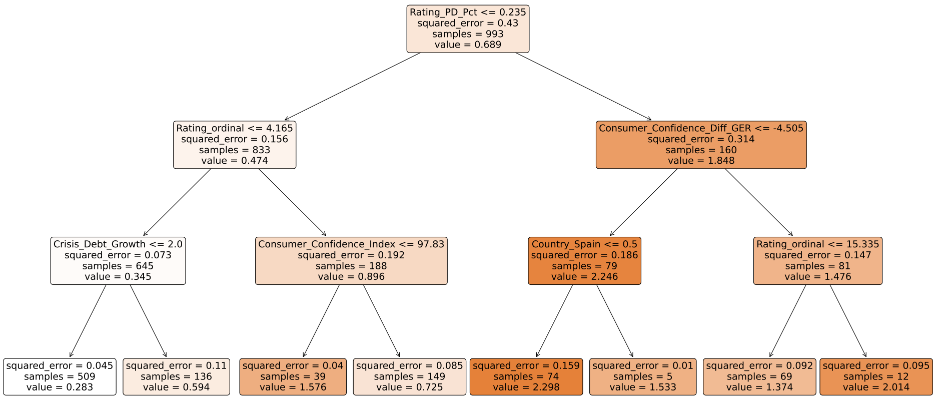
is expanded by the tuning parameter $\alpha$. Here $|T|$ is the number of final groups (terminal nodes) in the tree and $R_m$ represents the section of data that belongs to each group. The first part of the equation measures the total prediction error and the second part adds a penalty for having too many groups. When $\alpha = 0$, there is no penalty. The tree grows then unlimited to reduce prediction error. However, this can make the tree too complex and fit too closely to the training data (overfitting). When $\alpha$ increases, the penalty grows and forcing the tree to be smaller and simpler by reducing the number of groups. The best value of $\alpha$ can be chosen by testing different options on a validation set or using cross-validation. After choosing $\alpha$, the corresponding subtree is built using the full dataset.[22] Regression trees are simple and easy to understand, which makes them good for explaining results. However, they can be unstable and sometimes give extreme jumps in predictions, so they are not always reliable for direct estimation. Instead, they are often used as parts of larger prediction models.[23] Another disadvantage of decision trees is their sensitivity to data changes. A small change can lead to very different splits. An error at the top of the tree affects all the splits that follow. Even with better splitting rules, some instability remains. This is the cost of using a simple tree model.[24] Ensemble methods combine many simple models to create one strong model. These simple models are called weak learners because they do not perform well on their own. However, when they are combined, they can give very good predictions. Examples include bagging, RF, boosting, and Bayesian

[18] cf. James et al., 2023, p. 334.
[19] T. Hastie, Tibshirani, and Friedman, 2009, p. 307.
[20] cf. James et al., 2023, pp. 335–336.
[21] cf. T. Hastie, Tibshirani, and Friedman, 2009, p. 308.
[22] cf. James et al., 2023, p. 336.
[23] cf. Efron and T. J. Hastie, 2016, p. 127.
[24] cf. T. Hastie, Tibshirani, and Friedman, 2009, p. 312.

additive regression trees. In these methods, a decision tree is used as the basic building block.[25] Bagging and RF both use bootstrap as a resampling method. The bootstrap is used to test how well a model works. It creates many new datasets by randomly selecting data points from the original dataset. Each new dataset is the same size as the original. The model is then fitted to each new dataset. The differences in the model's performance across these datasets estimate its prediction error. This helps show how the model might perform on new data.[26] Bagging, also known as bootstrap aggregation, builds on bootstrapping. First, many new training sets are created by randomly sampling the original data with replacement. Then, a separate model is trained on each of these bootstrapped samples. Finally, the predictions from all the models are averaged. This process reduces the variance in the model and improves accuracy on new data.[27] Boosting, on the other hand, works by building a number of models, where each new model focuses on correcting the mistakes of the previous one. Unlike methods like bagging, boosting gives more weight to the models that perform better and result in an overall stronger model.[28] RF improves bagging by adding randomness when choosing splits. In a RF, many trees are built on bootstrapped samples. At each split, only a random subset of predictors is considered. Specifically, if $p$ is the total number of features, a random set of $m$ predictors is chosen at each split. Typical values of $m$ are $\sqrt{p}$ or $p/3$.[29] Unlike bagged trees, which can produce highly correlated predictions, RF avoids this issue by selecting random subsets of predictors at each split. Furthermore, adding more trees does not lead to overfitting.[30] Figure 3.2 shows a single decision tree from the RF model I used in my model. It visualizes the first tree using the `plot_tree` function from the scikit-learn library. This tree is one of many in the ensemble and helps understand how the model makes predictions. Here, the tree has a depth of 3 and starts at the root node with 993 samples. At each split, a feature and threshold are selected to divide the data into two groups. This reduces the variance and improves the prediction accuracy. For example, the first split occurs at Rating PD Pct $\leq 0.235$, separating observations into different branches. Rating PD Pct is here the variable for the credit rating in PD values in percent. Each node shows key details such as the number of samples, the squared error, and the predicted value (mean target value). The process continues until the tree reaches leaf nodes, where the final predictions are made. This tree is only one little part of the full RF, which averages predictions from many trees to reduce overfitting and improve accuracy.

[25] cf. James et al., 2023, p. 343.
[26] cf. T. Hastie, Tibshirani, and Friedman, 2009, p. 249.
[27] cf. James et al., 2023, p. 343.
[28] cf. T. Hastie, Tibshirani, and Friedman, 2009, pp. 337–338.
[29] cf. Efron and T. J. Hastie, 2016, pp. 326–327.
[30] cf. James et al., 2023, p. 347.

Figure 3.2: Decision Tree from the Random Forest Model




Source: Own illustration using the plot_tree function in scikit-learn

RF use out-of-bag (OOB) samples to estimate model performance. When building each tree, some data points are left out of the bootstrap sample. These unused data points are called OOB samples and are used to test the model. The OOB error is similar to the error from cross-validation, but it is calculated during training. This means RF can be trained in one step without needing separate cross-validation.[31] When the number of trees ($B$) is large enough, the OOB error gives almost the same result as cross-validation. This makes OOB estimation very useful, especially for large datasets, where cross-validation would take too much time.[32] In this analysis, I will still focus on the test error because I want to compare the RF model with the SVR models. RF make strong predictions but do not simply show how the model makes decisions. Each tree in the forest uses different predictor variables, some more often than others. To understand which variables are most important, the model tracks how much each split improves the prediction. These improvements are then added up across all trees to create a variable importance measure. A variable importance plot can show the most influential features in the model. It measures the relative importance.[33] I use the variable importance plot to identify the key factors that explain the spreads. The `RandomForestRegressor` function from the scikit-learn library was used to build the RF model in my analysis.

[31] cf. T. Hastie, Tibshirani, and Friedman, 2009, pp. 592–593.
[32] cf. James et al., 2023, p. 345.
[33] cf. Efron and T. J. Hastie, 2016, p. 332.

### 3.2.2 Support Vector Regression

SVR is an extension of Support Vector Machines (SVM) that is used for predicting continuous values. SVM is typically used for classification problems. With some modifications, SVM can be applied to regression tasks.[34] Vladimir Vapnik and some colleagues developed the SVM. The idea was first introduced in a paper by Boser et al. in 1992, which presented the optimal margin classifier (also called the optimal separating hyperplane). The concept quickly gained attention from a large community of researchers and later expanded into the broader category of kernel methods.[35] I will first explain the SVM for classification problems in detail and before moving on to SVR and regression problems. This approach ensures that the foundation of SVR is also explained.

The concept of an optimal margin classifier is based on a hyperplane. In two dimensions, a hyperplane is just a line. In three dimensions, it is a flat surface (a plane). In a space with $p$-dimensions, a hyperplane is a flat subspace with $p-1$ dimensions. This can be hard to visualize. The mathematical definition in a $p$-dimensional space is defined as

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p = 0 \tag{3.4}$$

where $\beta_0$ is the intercept and $\beta_1, \beta_2, ..., \beta_p$ are the other coefficients in a $p$-dimensional space. $X_1, X_2, ..., X_p$ are the coordinates (or features) of a point in that space. The equation is set equal to zero because it defines the hyperplane as the flat subspace where the sum of the weighted coordinates (plus the intercept) equals zero. Any point that fulfills the equation lies exactly on the hyperplane. Considering a point $X$ that does not lie exactly on the hyperplane. If

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p > 0, \tag{3.5}$$

then the point lies on one side of the hyperplane. If

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p < 0, \tag{3.6}$$

then the point $X$ lies on the other side of the hyperplane. In this way, the hyperplane divides the $p$-dimensional space into two regions. Evaluating the left-hand side of equation (3.4) for each point reveals which side of the hyperplane the point lies on.[36] In a classification problem with two dimensional space, the optimal separating hyperplane is the line that creates the biggest gap, know as margin, between the classes. By doing that on the training data, the hope is that it also works well on the test data.[37] The

[34] cf. James et al., 2023, p. 383.
[35] cf. Efron and T. J. Hastie, 2016, p. 390.
[36] cf. James et al., 2023, p. 386.
[37] cf. Efron and T. J. Hastie, 2016, p. 376.

classification of the test observation $x^*$ is based on the sign of

$$f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + ... + \beta_p x_p^*. \tag{3.7}$$

If $f(x^*)$ is positive, the observation is assigned to class 1. If it is negative, it is assigned to class $-1$. In addition to that, the distance of $f(x^*)$ can be measured. A value far from zero indicates that $f(x^*)$ lies far away from the hyperplane. This leads to a more confident classification. In contrast, a value near zero means that $f(x^*)$ is close to the hyperplane and the classification tends to less certain classification.[38] SVMs find a hyperplane that maximizes the gap between classes. The closest points to the hyperplane determine the margin. Finding this hyperplane is solved through an optimization problem.[39] Support vectors are the training points closest to the hyperplane and define the margin. These observations are called support vectors because each can be represented as a vector with $p$-dimensions. A small change in these points will shift the hyperplane. Moving other points does not affect the hyperplane unless they become support vectors.[40] The optimization problem is expressed as follows:

$$\max_{\beta_0,\beta_1,\ldots,\beta_p,M} M \tag{3.8}$$

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 = 1, \tag{3.9}$$

$$y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right) \geq M \quad \forall i = 1, \ldots, n. \tag{3.10}$$

In (3.8), $M$ is the margin to be maximized. It represents the smallest distance between any data point and the hyperplane. By selecting $\beta_0, \beta_1, \ldots \beta_p$ to maximize $M$, the hyperplane is placed as far as possible from the closest data points. This creates the best separation between the two classes. In (3.9), the constraint does not define the hyperplane itself but normalizes the coefficients. It ensures that the sum of the squared values of $\beta_j$ equals 1. Without this, multiplying all $\beta$ values by a constant would result in the same hyperplane, making $M$ meaningless. This constraint ensures that $M$ represents the true distance from each point to the hyperplane. This normalization is later relevant when scaling the input variables for the SVR models, which I will later discuss in 3.3. The expression (3.10) ensures that each data point is classified correctly. The term inside the parentheses represents the rule (seen in 3.4) that determines the classification of a point. Multiplying by $y_i$ makes sure the result is positive when the classification is correct. The inequality $\geq M$ means that each point must be at least a distance $M$ from the hyperplane. If $M$ is positive, the two classes are clearly separated. The term $\forall i = 1, \ldots, n$ means that this

[38] cf. James et al., 2023, pp. 369–370.
[39] cf. Awad and Khanna, 2015, p. 44.
[40] cf. James et al., 2023, p. 371.

condition applies to every data point. This guarantees that all observations are classified correctly and have at least a margin of $M$.[41] When the data cannot be fully separated, slack variables are added to allow some misclassification. Instead of finding a perfect separation, SVM becomes a soft-margin classifier. This means it correctly classifies most points but allows some points to be on the wrong side of the margin.[42] This helps when the data is not perfectly separable and also improves generalization to new test data. The optimization problem

$$\max_{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,M} M \tag{3.11}$$

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 = 1, \tag{3.12}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M(1 - \epsilon_i), \tag{3.13}$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C, \tag{3.14}$$

introduces $C$, a nonnegative tuning parameter. As in (3.8), $M$ is still the margin with the goal to maximize it. However, the constraint now includes slack variables $\epsilon_i$, which allow some data points to be on the wrong side of the margin or even the hyperplane. If $\epsilon_i = 0$, the $i$th point is correctly classified and on the right side of the margin. If $\epsilon_i > 0$, then the point is misclassified and on the wrong side of the hyperplane. If $0 < \epsilon_i \leq 1$, the observation is on the wrong side of the margin. In (3.14), $C$ sets a limit on the total sum of slack variables. This controls how many points can violate the margin and by how much. $C$ acts like a budget for margin violations. A small $C$ allows fewer violations and makes the margin stricter, while a larger $C$ allows more violations and makes the margin more flexible. $C$ controls therefore the balance between bias and variance.[43] When the data cannot be separated with a straight line, soft-margin SVM may not work well. It can cause more misclassifications and poor generalization. A kernel helps by transforming the data into a higher-dimensional space. In this new space, the data becomes easier to separate. Instead of searching for a complicated boundary, SVM finds a simple linear separator. This approach is effective because it improves learning without adding much extra computation.[44] SVMs build on top of the support vector classifier by expanding the feature space using these kernels. Instead of using the original data points, SVMs use inner products between observations. This makes the calculations easier. The linear

[41]cf. James et al., 2023, p. 372.
[42]cf. Awad and Khanna, 2015, p. 46.
[43]cf. James et al., 2023, pp. 375–376.
[44]cf. Awad and Khanna, 2015, p. 48.

support vector classifier is written as:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle \tag{3.15}$$

This equation shows that classification is based on a weighted sum of inner products between a new point $x$ and the training points $x_i$. Only the support vectors influence the final classification. If a training point is no support vector, its weight $\alpha_i$ is zero, so it has no effect on the final equation. To handle more complex cases where a linear boundary is not enough, the kernel trick is introduced. Instead of computing inner products directly, SVM replaces them with a function that measures similarity between points. This function $K$ is called kernel. It generalizes the inner product:

$$K(x_i, x_i') \tag{3.16}$$

This function measures how similar two data points are without calculating their positions in a higher-dimensional space. For a linear kernel, the similarity is measured using the standard Pearson correlation.[45] $K$ can have different functions. Besides the linear kernel discussed here, other common types include the polynomial and radial kernels.[46] The SVM method used for classification can also be applied to regression. This is known as SVR. SVR is efficient because its complexity does not depend on the number of input variables. It also provides accurate predictions and generalizes well to new data. SVR works by introducing an $\epsilon$-insensitive region, called the $\epsilon$-tube, around the function. This tube changes the optimization problem. The goal is to find the best tube that fits the data while keeping the model simple and accurate.[47] This tube sets a limit for errors where no penalty is given, similar to the tuning parameter $C$ in (3.14) for SVMs. Points outside this tube add to the loss function and change the optimization process. To handle this, SVR uses the following optimization problem:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{3.17}$$

The first part in the equation, $\frac{1}{2}\|w\|^2$, represents the model complexity. Here, $\|w\|$ shows how big the vector is. This vector points away from the surface the model is making. It helps decide the shape and position of the line. $\frac{1}{2}$ is only included to simplify the math when solving the optimization problem. The second part, $C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$, controls how much error is allowed. $C$ is a tuning parameter that decides the balance between keeping the model simple and allowing some mistakes. The sum $\sum_{i=1}^{N}(\xi_i + \xi_i^*)$ adds up the

[45] cf. James et al., 2023, pp. 379–380.
[46] cf. T. Hastie, Tibshirani, and Friedman, 2009, p. 424.
[47] cf. Awad and Khanna, 2015, p. 67.

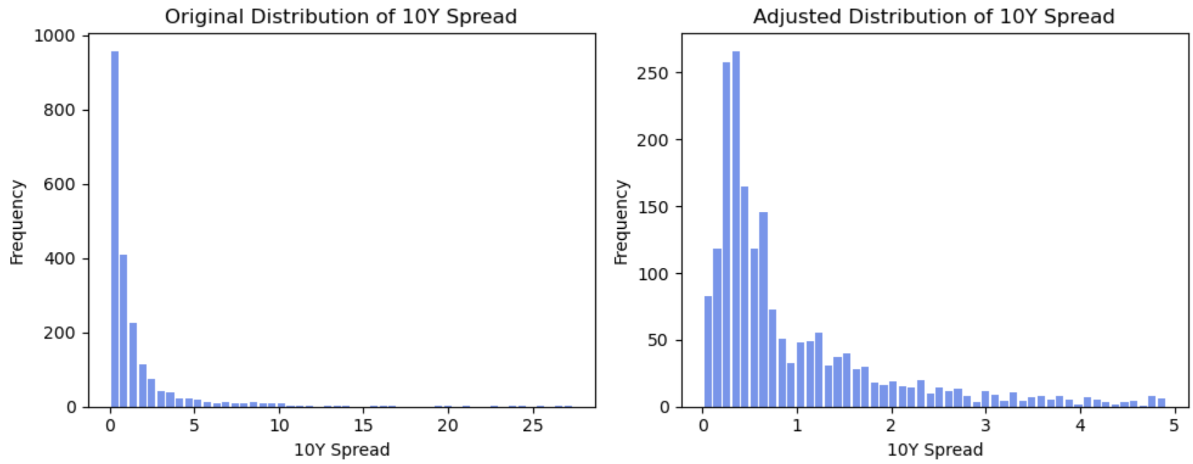
errors beyond the $\epsilon$-tube, the slack variables. A small $C$ allows more flexibility, meaning the model will tolerate more errors. A large $C$ makes the model stricter, reducing errors but possibly making it too complex.[48] For non-linear functions, the data is mapped to a higher-dimensional space called the kernel space. This improves accuracy. Since kernels were already introduced in the SVM section, they are now used to handle complex patterns in SVR. Like in classification, this helps SVR work better with non-linear relationships. Replacing $x$ with a kernel function $k(x_i, x_j)$ changes how the problem is written. This creates new equations for the weight vector and the final prediction function.[49] In my analysis, I use SVR with a linear kernel as the base model. Its performance is compared with that of SVR with a radial kernel (RBF) and a RF model discussed earlier. I use the `SVR` function from the scikit-learn library.

## 3.3 Statistical Framework

The sovereign bond spreads are highly right-skewed. Therefore, I removed outliers that could negatively impact the model's predictive performance. To reduce outliers in the 10-year bond spreads, I used the Interquartile Range (IQR) method. First, I calculated the first quartile (Q1) and third quartile (Q3) of the 10-year spread variable. The IQR was then set as the difference between Q3 and Q1. Outliers were identified using the standard rule: any values below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ were considered outliers.[50] Finally, I removed all data points outside this range to create a more balanced distribution for the analysis. As a result, the dataset was reduced from 2040 rows to 1812 rows including Greece and from 1836 to 1666 in the dataset without Greece. Figure 3.3 shows two histograms, that show the distribution of the target variable, the 10-year spreads. The left histogram in Figure 3.3 shows the original 10-year bond spreads, which is highly right-skewed. This means there are many observations clustered at lower values and a few observations with much higher values. These extreme outliers come primarily from the sovereign bond crisis. The right histogram shows the spreads after removing these outliers through the IQR method. This makes the distribution more balanced and helps the model work better.

[48] cf. Awad and Khanna, 2015, pp. 68–71.
[49] cf. Awad and Khanna, 2015, p. 72.
[50] cf. Han, Pei, and Kamber, 2012, p. 49.

Figure 3.3: Impact of Outlier Removal on the Distribution of 10-Year Spreads




Source: Own illustration based on data from OECD

In a first model development, I trained the RF and SVR models without removing outliers. The RF model performed well with an $R^2$ of 0.73, while the SVR model had a negative $R^2$, indicating it struggled with the data. This suggests that the RF model is more robust to outliers, whereas the SVR model is sensitive to extreme values. After removing outliers, the SVR model's performance improved significantly. The right histogram of Figure 3.3 shows that the distribution is still right-skewed after removing outliers. To handle this further, I applied the natural logarithm transformation to the spread variable. This transformation helps the variable further to normalize the data and make the distribution less skewed. Table 3.1 shows that all values are greater than zero. Logarithmic transformations don't work with zero or negative values. So this ensures it can be applied correctly. In the next step, I focused on feature selection to further improve the model. The dataset originally has 151 independent variables. This is a lot for a dataset with a small sample size. To reduce the number of variables, I used feature selection to remove unimportant variables and keep only the relevant ones. In the first step, I applied the Boruta feature selection algorithm. Boruta uses the RF algorithm internally and therefore fits in my analysis. RF is fast, works without much tuning, and provides a numerical measure of feature importance. It creates many decision trees from random samples of the data. These trees combine their predictions to reach the final result. A feature is considered important if shuffling its values leads to a drop in accuracy. This drop in accuracy is calculated for every tree that uses the feature. Then, the average and standard deviation of these drops are computed.[51] For each feature, a shuffled copy (called a shadow feature) is created. Then, a RF is run on all features. The importance of every feature is measured. A shadow feature's score is random. Its score is used as a reference. The shuffling is repeated several times to get reliable results. The algorithm stops when

[51] cf. Kursa and Rudnicki, 2010, p. 2.

all features have been evaluated.[52] In this analysis, I use the `BorutaPy` algorithm from the boruta library. In the second step, I further reduced the features. This time only for the SVR base model with a linear kernel. RF and SVR with an RBF kernel handle multicollinearity well, so this step is not needed for them. However, for the linear SVR model, reducing multicollinearity is an important step. I computed a correlation matrix using the reduced features selected by the Boruta algorithm. Then, I removed features that showed a correlation above 0.75 with any other feature. I only kept the variable with the higher correlation to the target variable. This step helped reduce multicollinearity in the linear model. For accurate cross-validation, I used the PanelSplit package.[53] While the scikit-learn library in Python offers methods like `TimeSeriesSplit`, these do not work well with panel data. The `PanelSplit` function is specially designed for panel data structure. It splits the data and maintains the chronological order for each country. This approach ensures there is no data leakage, meaning that only past information is used during training. I split the dataset into a training period from 2007 to 2018 and a testing period from 2019 to 2023. Cross-validation is the resampling method used in my analysis. Resampling methods help create many samples from the training data. With modern computing power, it is now possible to create many subsets of data. In cross-validation, the training set is split into multiple sets. The test set remains unchanged. This process helps build a model that performs better on unseen data. Often, there is a gap between training error and test error. Cross-validation reduces this gap by improving the model's ability to generalize better on unseen data.[54] In order to get the best performance of each ML model, I used a hyper-parameter tuning step in combination with cross-validation. This calibrates the parameters of each model. In ML, hyper-parameters are important for model performance. Automated hyper-parameter optimization sets parameters automatically. This boosts their performance. It also reduces manual work and customizes models for specific tasks. It makes studies more reproducible and fair by tuning every model equally.[55] I used `GridSearchCV` from the scikit-learn library to find the best hyper-parameters. Grid search is a basic method for automated hyper-parameter optimization. In grid search, a set of values is chosen for each hyper-parameter. Then, every possible combination is tested. A disadvantage is that as more hyper-parameters or values are added, the number of combinations grows quickly. This can make the process very expensive computationally.[56] I chose GridSearchCV because my small dataset makes it manageable to test all hyper-parameter combinations. In my models, I tuned the following hyper-parameters. For RF, I tuned the number of trees (n_estimators), the maximum tree depth (max_depth), and the number of features used at each split

[52] cf. Kursa and Rudnicki, 2010, p. 3.
[53] cf. Frey and Seimon, 2024, panelsplit (v0.4.2).
[54] cf. James et al., 2023, pp. 201–202.
[55] cf. Feurer and Hutter, 2019, pp. 3–4.
[56] cf. Feurer and Hutter, 2019, p. 7.

(max_features). For both SVR models, I used the regularization strength (C) and the margin of tolerance (epsilon). The optimal hyper-parameters determined through GridSearchCV are presented in Table A.2 in the Appendix. The model performance was measured using the negative mean squared error. In many ML algorithms, the units of the data can highly impact the model performance. If the variable is expressed in smaller units, its numerical range increases. This can give the variable a larger effect on the model. To avoid this issue, the data is often normalized or standardized. Scaling the data to a common range such as $[-1, 1]$ or $[0, 1]$ helps ensure all features contribute fairly to the model. This normalization is important for algorithms like neural networks, nearest-neighbor classification, and clustering. The scale of the input features can significantly influence the outcome.[57] SVR is also sensitive to the scale of features. As earlier discussed in the SVR optimization problem in (3.17), the model balances complexity and errors. These terms depend on the input features, so their scale affects the model. If features have different ranges, some will have more influence than others. This would make the model unbalanced. Scaling is needed to ensure all features contribute equally. In my analysis, I used the `MinMaxScaler` function from the scikit-learn library. This function transforms all features into the range $[0, 1]$. The minimum value becomes 0 and the maximum becomes 1. To compare the models, I used different evaluation metrics based on the test data, not the training data. This is important because good training results do not always mean good test results. The main goal is to check how well the model works on new data. The first metric is the mean squared error (MSE), which is a common metric for regression problems. The MSE is defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2, \tag{3.18}$$

where $\hat{f}(x_i)$ represents the predicted value for the $i$th observation and $y_i$ is the actual observed value. The error for each observation is squared and summed, and then averaged over all observations. A lower MSE indicates that the predictions are closer to the true test data.[58] Besides MSE, I also calculated the root mean squared error (RMSE). RMSE is the square root of MSE. It converts the error back to the same unit as the target variable. This makes the results easier to interpret. I also computed the mean absolute error (MAE), which measures the average absolute difference between actual and predicted values. Furthermore, I included this metric because MAE could be more reliable than RMSE because RMSE is affected by the error variance and sample size. Willmott and Matsuura (2005) highlight this in climate research, but the same reasoning applies to economic modeling.[59] As the last evaluation metric, I chose the $R^2$ statistic. It measures

[57]cf. Han, Pei, and Kamber, 2012, p. 113.
[58]cf. James et al., 2023, pp. 28–29.
[59]cf. Willmott and Matsuura, 2005, p. 82.

how much of the target variable can be explained by the model. $R^2$ always takes a value between 0 and 1, which makes it easy to interpret. A value close to 1 means the model explains most of the variability in the data, while a value near 0 means it explains very little. This can happen if the model does not fit well or if the data has a lot of randomness.[60] Besides the evaluation metrics, I included the actual vs. predicted plots for each model to better illustrate their performance. In these plots, each point represents a prediction made by the model. The x-axis shows the actual bond spread value, while the y-axis shows the predicted values. If a model were perfect, all points would lie exactly on the 45-degree reference line. The closer the points are to this line, the better the model's predictions. If a point is above the line, the model has overestimated the bond spread and therefore predicted a higher value than the actual one. If a point is below the line, the model has underestimated the bond spread and predicted a lower value than the actual one. I also included residual plots to show the errors in predictions. Here, the x-axis represents the predicted bond spreads, while the y-axis represents the residuals. Residuals are the difference between actual and predicted values. A good model should have residuals evenly spread around zero. If there is a pattern, it means the model is missing something in the data. One possible explanation for this could be heteroscedasticity. This means that the variance of the residuals changes across different values of the predicted variable.[61]

[60]cf. James et al., 2023, p. 79.
[61]cf. James et al., 2023, p. 103.

# 4 Empirical Results

In this chapter, I present my empirical findings on predicting 10-year sovereign bond spreads. I use RF and SVR with a radial kernel (RBF) to predict bond spreads for the following ten euro area countries: Austria, Belgium, Finland, France, Greece, Ireland, Italy, the Netherlands, Portugal, and Spain. I compare these ML models to a linear SVR model, which serves as a benchmark. This way I can evaluate if the ML models can capture non-linear patterns in the data and perform better against the linear benchmark. Therefore, I test **H1**, which states that RF and SVR (RBF) predict sovereign bond spreads more accurately than the linear SVR baseline model. In the second part, I analyze which variables have the most influence in the models through variable importance measures. **H2** suggests that macroeconomic indicators such as debt-to-GDP, GDP growth, and inflation are the main determinants in bond spreads. Last, I test if excluding Greece from the dataset affects the model's in a positive way and changes the main determinants. Due to the high bond spreads Greece had during the sovereign debt crisis, it can be seen as an outlier. Therefore, **H3** explores whether removing Greece as an outlier improves predictions in all three models.

## 4.1 Predictive Performance and Model Evaluation

The following section presents the overall evaluation metrics for the test period from 2019 to 2023, the metrics across the years to see if there are any trends or notable years. This section uses MSE, RMSE, MAE, and $R^2$ Statistic. After this, I analyze further the predicted vs actual and residual plots.

Table 4.1: Evaluation Metrics for Sovereign Bond Spreads

| Model | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|
| Random Forest | 0.079 | 0.281 | 0.219 | 0.780 |
| SVR Radial | 0.098 | 0.312 | 0.227 | 0.729 |
| SVR Linear | 0.218 | 0.467 | 0.313 | 0.395 |

Source: Own calculations

The results show that the RF model performs better than both SVR models across all metrics. RF has the lowest MSE (0.079), RMSE (0.281), and MAE (0.219) and can explain 78% of the variance. The RMSE of 0.281 means that, on average, the prediction of the RF model is 0.281 percentage points away from the actual spreads. The SVR (RBF) model has slightly higher metrics for MSE, RMSE, and MAE, but can also perform well. It can explain 72% of the variance in the data. The SVR linear model has significantly higher metrics for MSE, RMSE, and MAE. This makes the predictions of the model further away from the real values. With a $R^2$ of just 0.395, it can only explain 39.5% of the variance in the data. This indicates that the relationship between the features and the target variable is non-linear. Overall, the RF and SVR (RBF) models can predict accurately sovereign bond spreads, whereas the linear model struggles.

Table 4.2: MSE for Sovereign Bond Spreads by Year

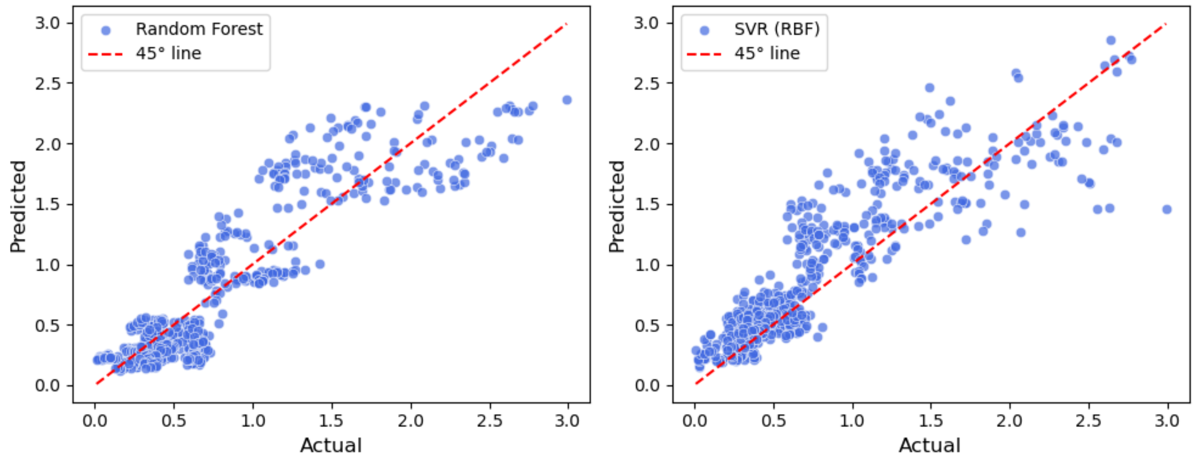
| Model | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| Random Forest | 0.053 | 0.090 | 0.095 | 0.081 | 0.075 |
| SVR Radial | 0.093 | 0.139 | 0.166 | 0.052 | 0.037 |
| SVR Linear | 0.226 | 0.282 | 0.077 | 0.072 | 0.432 |

Source: Own calculations

Table 4.2 represents the MSE values for all three models across all test years. The RF model shows relatively stable performance across the years. The MSE values stay between 0.053 and 0.095, indicating that RF maintains consistency in predictions. The SVR (RBF) model has slightly higher MSE values in 2020 (0.139) and 2021 (0.166) but performs better than RF in 2022 and 2023. The SVR linear model has the highest MSE values across all years. Especially in 2023, there is a big gap between the linear model (0.432) and the ML models (0.075, 0.037). However, in 2021 it even outperforms the ML models and also shows in 2022 good performance. This shows it can accurately capture the data in certain years. Overall, RF has the most consistent performance across all years. However, the SVR (RBF) model shows superior results in recent years and the SVR linear model outperforms both ML models in 2021.

Figure 4.1 presents the actual vs predicted values for both ML models. On the left is the RF model and on the right the SVR (RBF) model. In the RF plot, there is a good alignment along the 45-degree line. This is especially the case for lower values (0-1%). At higher spreads (1.5-3%), there is some scatter around the line. The model performs better on lower values than for higher spreads. In the SVR (RBF) plot, the pattern is similar to the RF model. However, there appears to be slightly more scatter than the RF plot. The SVR (RBF) also performs better at lower spreads than on higher ones. Some outliers at higher values are noticeable. Altogether, both models show reasonable prediction performance, with the RF model showing slightly better predictions. This aligns with the results, where RF has the lowest MSE, RMSE, and MAE values.

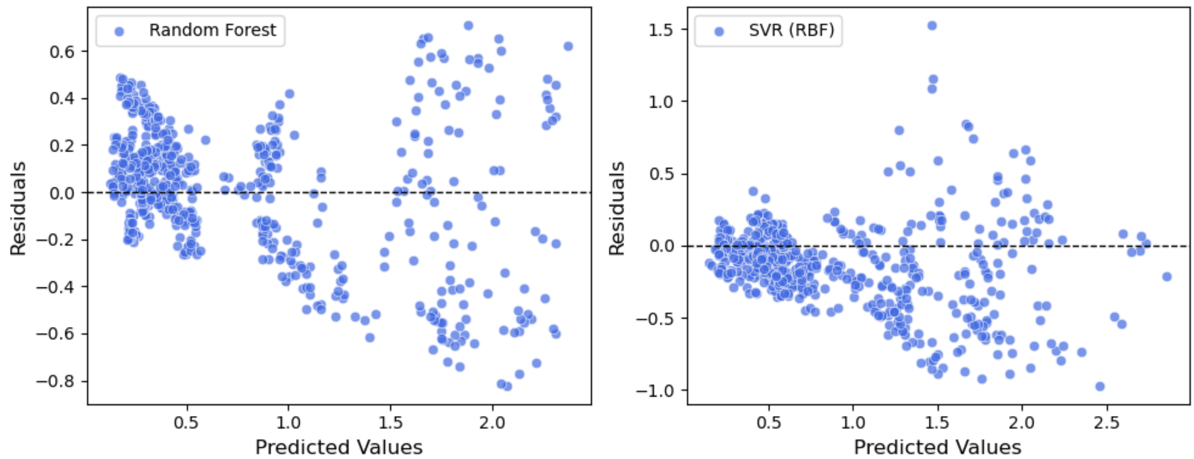
Figure 4.1: Actual vs predicted values for sovereign bond spreads




Source: Own illustration

Figure 4.2 shows the residual plots for the RF and SVR (RBF) models. In the RF plot on the left, the residuals are in a range from -0.8 to 0.6. For lower values, there is a mix of positive and negative residuals around zero. For higher spreads, there appears to be a pattern. The higher the actual spread values, the higher the residuals. This suggests some heteroscedasticity in the data. In the SVR (RBF) plot, the range is approximately from -1.0 to 1.5. This is a wider range than for the RF model. Like in the RF plot, for lower values, the residuals are clustered more tightly around zero. For higher predicted values, the residuals trend more negative. Both models show heteroscedasticity in the data. This suggests both models might be missing some structure in the data.

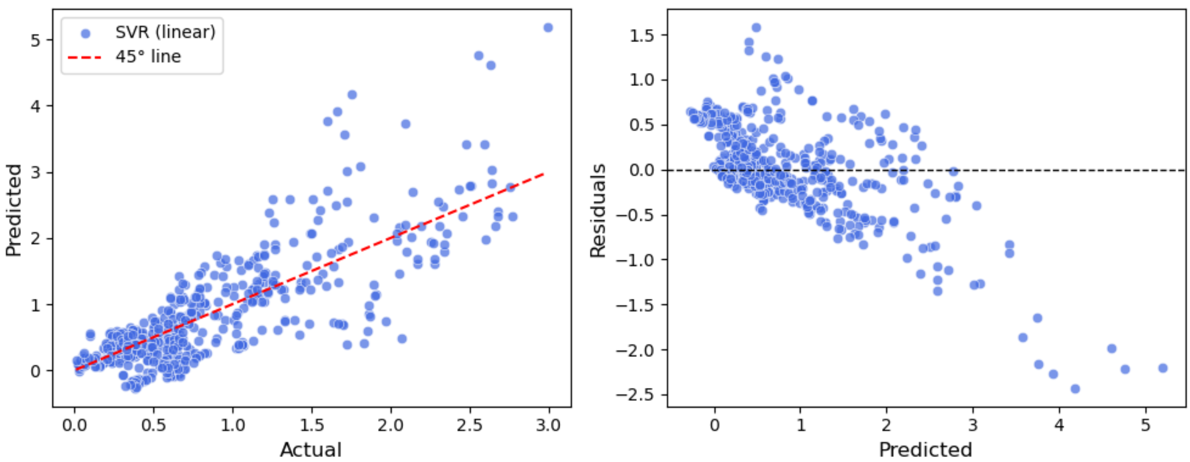
Figure 4.2: Residual Plot for Random Forest and SVR Radial




Source: Own illustration

Figure 4.3 shows the performance of the linear SVR model. On the left side are the actual vs. predicted values and on the right the residual plot. The actual vs. predicted plot shows that the points are more scattered. For higher values (2-3%), the model tends to overestimate the actual values. The alignment to the 45-degree line is much weaker compared to both ML models. Also, the residual plot on the right shows a wider range (-2.5 to 1.5). There is a clear downward trend in the residuals, showing strong heteroskedasticity. The variance isn't constant across the prediction range. For higher predicted values (3-5), residuals become more negative. The model overestimates these values. Overall, there is much less random scatter around zero compared to the ML models.

Figure 4.3: SVR Linear Model Performance



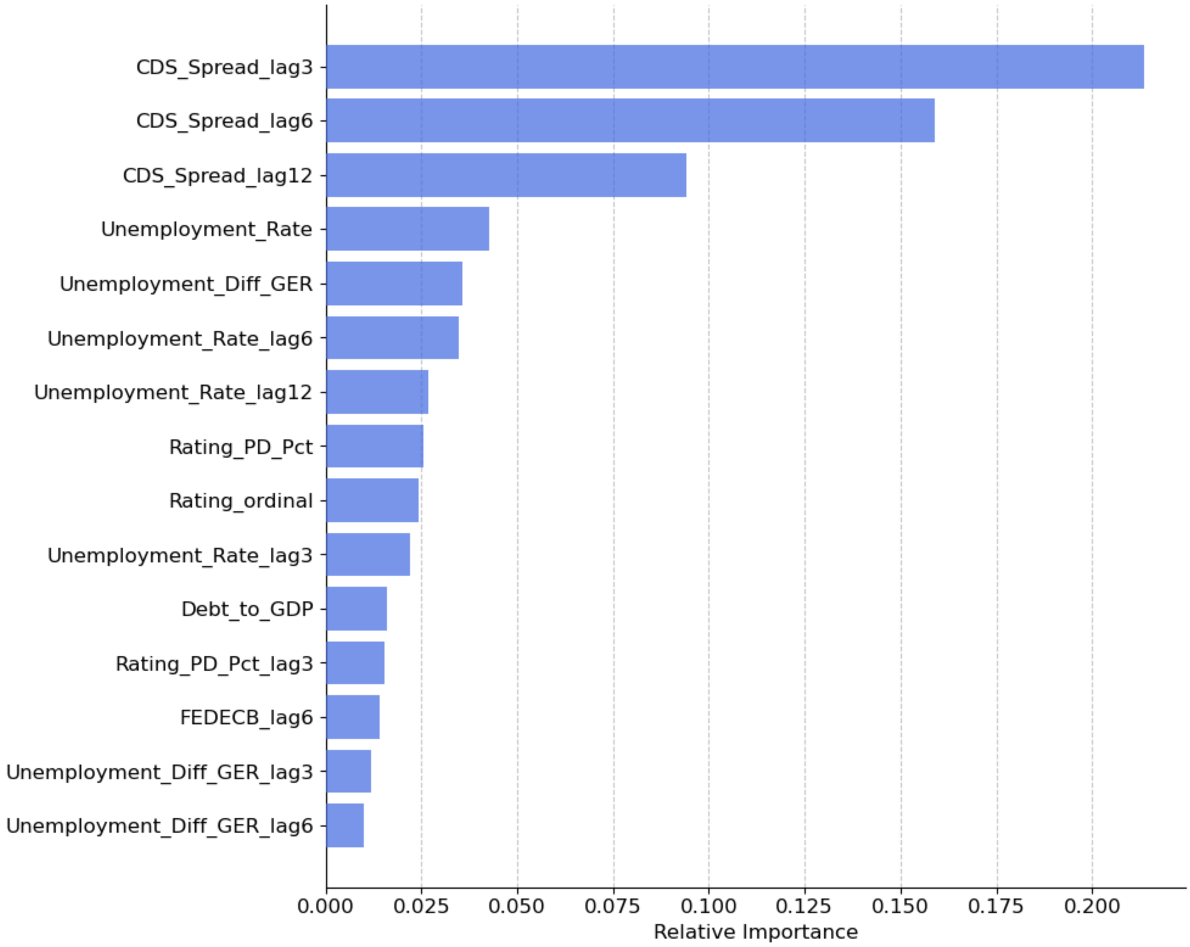

Source: Own illustration

My empirical findings strongly support **H1**: RF and SVR (RBF) predict sovereign bond spreads more accurately than the linear SVR model. The RF model has the best performance with the lowest error metrics and the highest $R^2$. The SVR (RBF) model performs better than the linear benchmark. Making it the second best model. The SVR linear model performs worst and confirms that sovereign bond spreads have non-linear relationships with the determinants.

## 4.2 Variable Importance and Key Drivers of Spreads

Understanding which factors influence sovereign bond spreads helps explain how the models make their predictions. I show for the RF and SVR (RBF) models the 15 most important variables ranked in descending order. The RF model ranks variables by how much they improve predictions, as discussed earlier. Figure 4.4 shows the most important variables in the RF model for the 2007 to 2023 period.

Figure 4.4: Random Forest variables importance

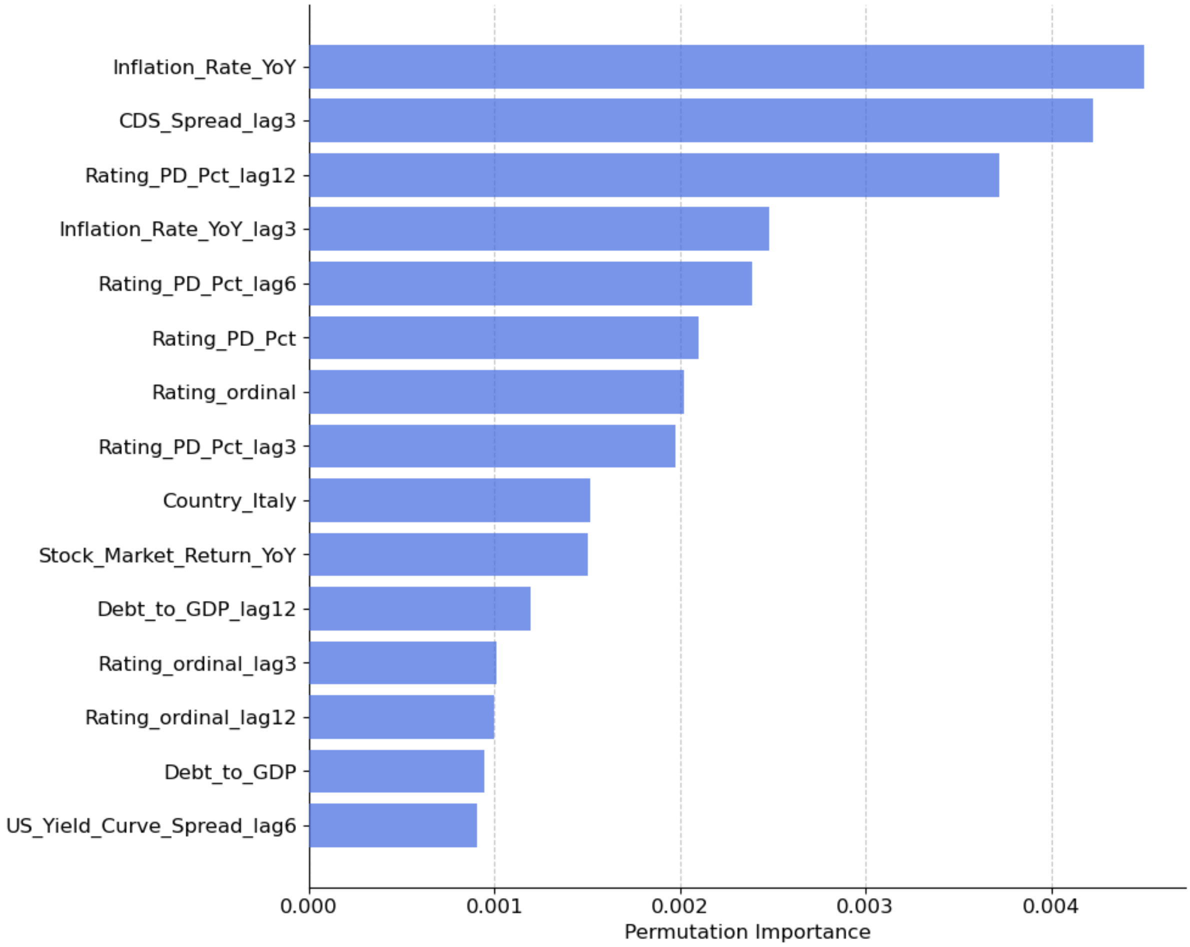



Source: Own illustration

In this model the most important factor for bond spreads is credit risk, especially the lagged CDS spreads. The top CDS spread variable (CDS_Spread_lag3) shows a relative importance of more than 0.2. This is approximately four times larger than most of the lower ranked variables in the plot. This shows that CDS spreads are closely linked to sovereign bond spreads. Which is not surprising due to the existing literature. Sovereign credit ratings also play a role in this model and appear in the top 15 variables. However, they are less important than CDS spreads, which suggests that markets react faster to CDS data than to rating changes. Macroeconomic factors help explain bond spreads as well. The unemployment rate, its difference from Germany (ending Diff_GER), and its past values are the second most important factors. This could mean that higher unemployment rates raises sovereign spreads. Debt-to-GDP is also a factor, but it has less influence than credit risk and the unemployment rate. While high debt can affect bond spreads in this model, credit risk and labor market conditions are more important. Monetary policy also plays a small role in this RF model. The interest rate difference between the Federal Reserve and the ECB, with a six-month lag, helps explain bond spreads. This suggests that differences in interest rates between the two central banks could influence bond spreads in the euro area. Overall, the RF model gives more importance to credit risk like CDS spreads and credit ratings rather than traditional macroeconomic factors.

Figure 4.5 shows the most important variables in the SVR (RBF) model for predicting bond spreads. The model ranks inflation and credit risk as the key drivers.

Figure 4.5: Support Vector Regression (RBF) variable importance




Source: Own illustration

The year-over-year inflation rate is the most important factor, suggesting that higher inflation could lead to higher bond spreads. The 3-month lagged inflation variable is also important, highlighting the strong role of inflation in this model. Credit risk factors also play a big role. The 3-month lagged CDS spread is highly influential. Sovereign credit ratings, both measured in PD and ordinal rating, are also ranked high. This aligns with the RF model, which also finds credit risk to be a key factor of spreads. Other macroeconomic and market factors also affect spreads. Stock market returns (year-over-year) appear in the rankings, suggesting that equity market performance is linked to bond spreads. Debt-to-GDP and its lagged values are included but have less importance. This means that debt ratios influence spreads, similar to the RF model, but not as much as inflation and credit risk. A country-specific factor is also relevant in the SVR (RBF) model. The dummy variable for Italy appears in the rankings, showing that Italy's unique economic and financial situation affects spreads differently compared to other euro area countries. Lastly, monetary policy conditions also play a role. This time, the U.S. Yield Curve Spread (10-year minus 1-year) with a six-month lag is among the important variables. This matches the six-month lag found in the RF model for the FED-ECB interest rate spread. Suggesting that monetary policy effects may influence bond spreads with a six-month delay. Overall, Figure 4.5 shows that the SVR (RBF) model is more balanced and places more importance on inflation and credit ratings than the RF model, which highly prioritized CDS spreads. This suggests that SVR (RBF) captures different relationships in the data, focusing more on macroeconomic conditions and sovereign ratings as key drivers of bond spreads.

These results show that the RF and SVR models focus on different aspects of credit risk and macroeconomic conditions when predicting bond spreads. The RF model relies more on market-based indicators like CDS spreads. The SVR model gives more weight to credit ratings and inflation. This means that both models see credit risk as important but measure it differently. Both models agree that credit risk, macroeconomic conditions, and monetary policy affect bond spreads. Other factors, like interest rate differences, stock market returns, and country-specific risks, also play a small role. How these findings fit in the existing literature, I will discuss in the next chapter.

## 4.3 Impact of Greece on Model Performance

Table 4.3 presents the evaluation metrics for the models after removing Greece from the dataset. Comparing these results with the original metrics in Table 4.1 shows how Greece's bond spreads affected model performance.

Table 4.3: Evaluation Metrics for Sovereign Bond Spread (excluding Greece)

| Model | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|
| Random Forest | 0.061 | 0.247 | 0.193 | 0.709 |
| SVR Radial | 0.053 | 0.231 | 0.168 | 0.746 |
| SVR Linear | 0.077 | 0.277 | 0.203 | 0.633 |

Source: Own calculations

After excluding Greece, all models improved in accuracy, as seen by the lower MSE, RMSE, and MAE values. The RF model now has an MSE of 0.061, compared to 0.079 before. This suggests that Greece's extreme bond spread movements may have added noise. This makes the predictions harder in the full dataset. The SVR (RBF) model also shows a notable improvement, with MSE dropping from 0.098 to 0.053 and $R^2$ increasing from 0.729 to 0.746. Now it explains more of the variance in bond spreads. The linear SVR model improves significantly, but it still remains less accurate than the ML models. The higher $R^2$ values for the SVR models confirm that removing Greece improved their performance. The biggest overall improvement is in the SVR linear model. The $R^2$ significantly increased from 0.395 to 0.633. However, the RF model performed slightly worse in $R^2$ statistic. $R^2$ decreased from 0.780 to 0.709. This is because Greece had very high bond spreads, which added more variance to the data. Without Greece, the total variance in bond spreads was smaller. The $R^2$ statistic measures the variance explained by the model, as discussed in chapter 3. As a result, the RF model explains less of the variance, which causes a slight drop in $R^2$. Therefore, the RF model including Greece seems to predict these high values relative well.

Table 4.4 presents the MSE for each year after removing Greece from the dataset. Comparing these values to the original results provides insights into how Greece's exclusion affected model performance over time.

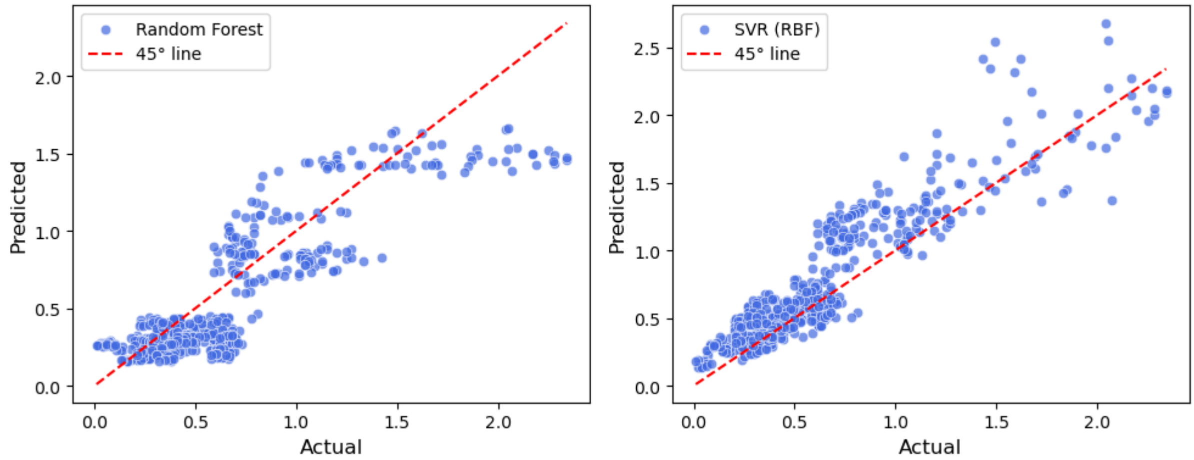
Table 4.4: MSE for Sovereign Bond Spreads by Year (excluding Greece)

| Model | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| Random Forest | 0.028 | 0.050 | 0.025 | 0.104 | 0.095 |
| SVR Radial | 0.072 | 0.089 | 0.059 | 0.022 | 0.024 |
| SVR Linear | 0.014 | 0.123 | 0.053 | 0.045 | 0.145 |

Source: Own calculations

The RF model shows lower MSE values in most years, especially in 2019 (0.028) and 2021 (0.025). However, in 2022 and 2023, the MSE increased to 0.104 and 0.095. The RF model without Greece struggles more in these years. Overall, RF remains the most stable model across the years. The SVR (RBF) model also improves in most years. The largest drop in MSE is in 2022 and 2023, where MSE decreased from 0.052 and 0.037 to 0.022 and 0.024. SVR (RBF) benefits especially in the last years, when excluding Greece as an outlier. The linear SVR model also performs better without Greece in the dataset. It has the overall lowest MSE in 2019 (0.014), showing significant improvements. However, in 2020 and 2023 the MSE remains higher than the ML models. The SVR linear model is after excluding Greece more balanced, but still less accurate than the other models. Overall, the results suggest that removing Greece leads to lower errors across all models. However, the impact is not the same for every year. SVR (RBF) benefits the most, especially in later years, while RF remains the most stable model over time. Figures 4.6 and 4.7 show the actual vs. predicted values and residual plots for the RF and SVR (RBF) models after removing Greece from the dataset.

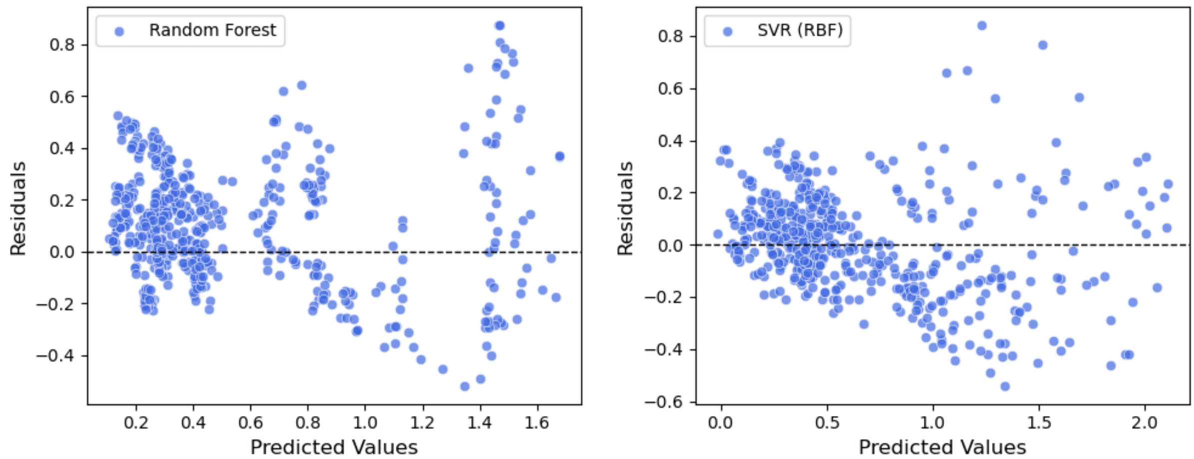
Figure 4.6: Actual vs predicted values for sovereign bond spreads (excluding Greece)




Source: Own illustration

Both models now align better with actual values. The RF model on the left remains stable, but the pattern of its predictions looks slightly unusual. Some points seem grouped in specific areas rather than being evenly spread. For lower values, the model underestimates the actual values. The SVR (RBF) model on the right shows the biggest improvement, with fewer large errors and better predictions across different spread levels. This suggests that Greece's extreme bond spreads previously made it harder for the SVR (RBF) model to find patterns.

Figure 4.7: Residual Plot for Random Forest and SVR Radial (excluding Greece)

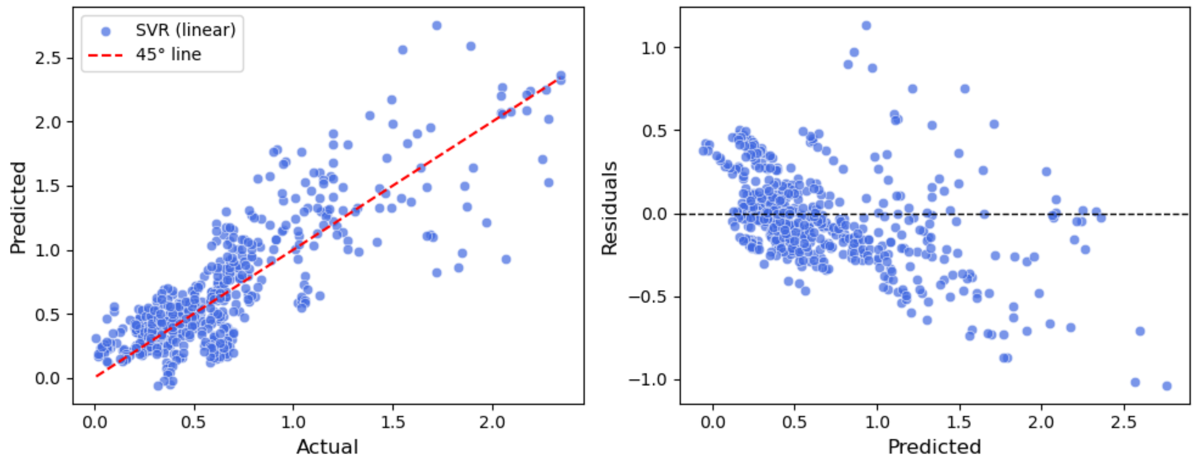



Source: Own illustration

Residuals are more uneven for the RF model. As before, the residuals are not randomly spread. Instead, they have clusters and increase at larger values. One possible reason for that could be that the RF model strongly relies on CDS spreads. Since CDS data were the most important predictors in RF, the model may focus too much on credit risk and miss other factors. The SVR (RBF) model considers a wider range of variables. This makes its predictions more balanced. The residuals, after excluding Greece, are more randomly spread. Even if the errors still increase with higher spread values. This suggests that heteroskedasticity is still the case.

Figure 4.8 shows the performance of the linear SVR model after removing Greece. The actual vs. predicted plot on the left looks better than before, with more points aligning with the 45-degree line. This suggests that removing Greece helps the model make better predictions and is supported by the lower error metrics. However, the linear model still struggles with higher spread values, as seen in the plot.

Figure 4.8: SVR Linear Model Performance (excluding Greece)

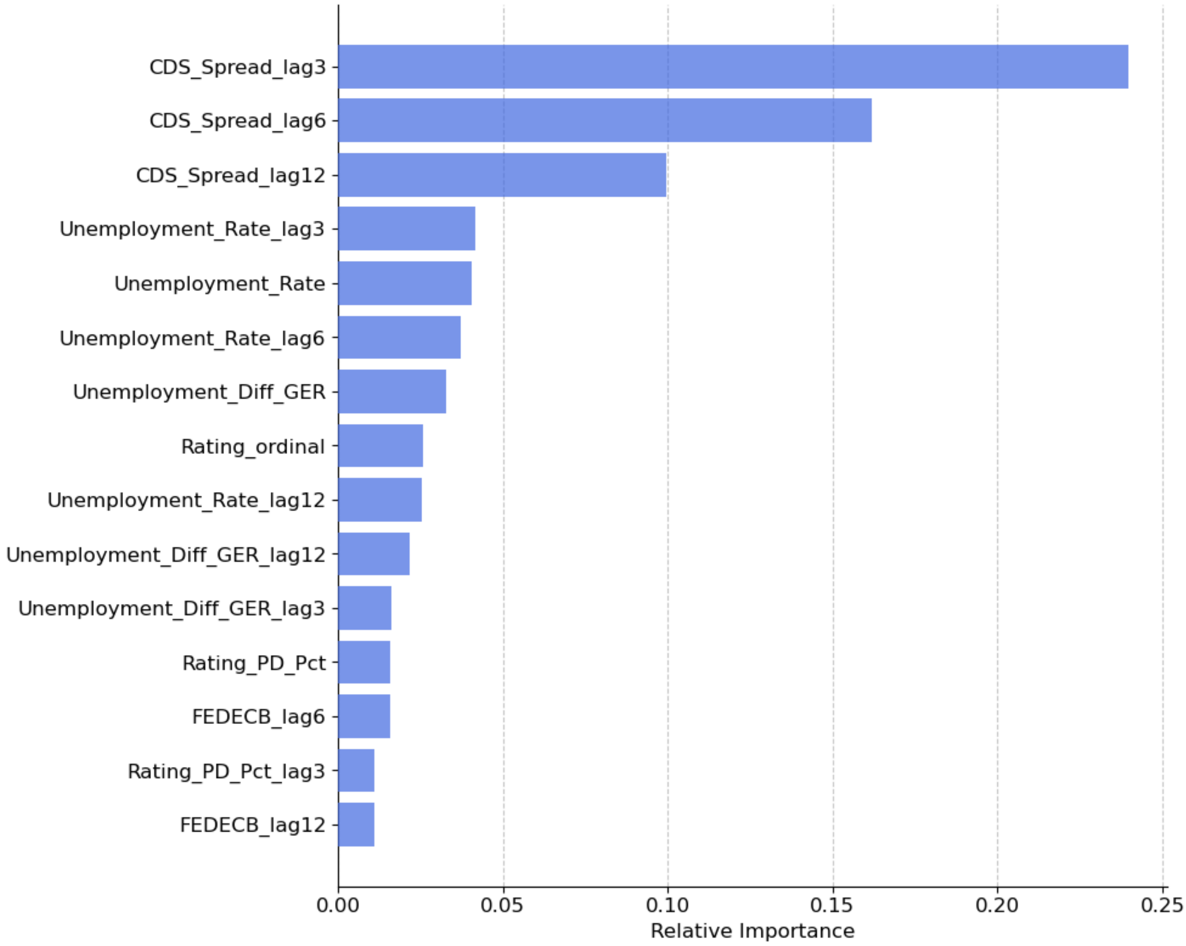



Source: Own illustration

The residual plot on the right also improves after removing Greece. The residuals are more evenly spread around zero, but a slight downward pattern remains. This supports the idea that here heteroskedasticity is still a problem. However, compared to the dataset including Greece, it improved significantly. This suggests that Greece's high bond spreads had a large effect on how well the model could predict values across different spread levels. Without these extreme values, the model can make more consistent predictions.

Figure 4.9 shows the updated feature importance in the RF model after removing Greece. The CDS spreads variables remain the most important predictors. Their influence even increased slightly. This could also explains the more unusual pattern in the predicted vs. actual plot mentioned earlier. Unemployment variables have also gain importance, especially the 3-month lag, highlighting labor market conditions as key drivers. In contrast, credit ratings become slightly less important, meaning the model now focuses more on CDS spreads and unemployment. Monetary policy factors remain in the top variables but with lower influence. These changes indicate that without Greece, the model's key drivers of spreads become more focused on credit risk and labor market conditions.

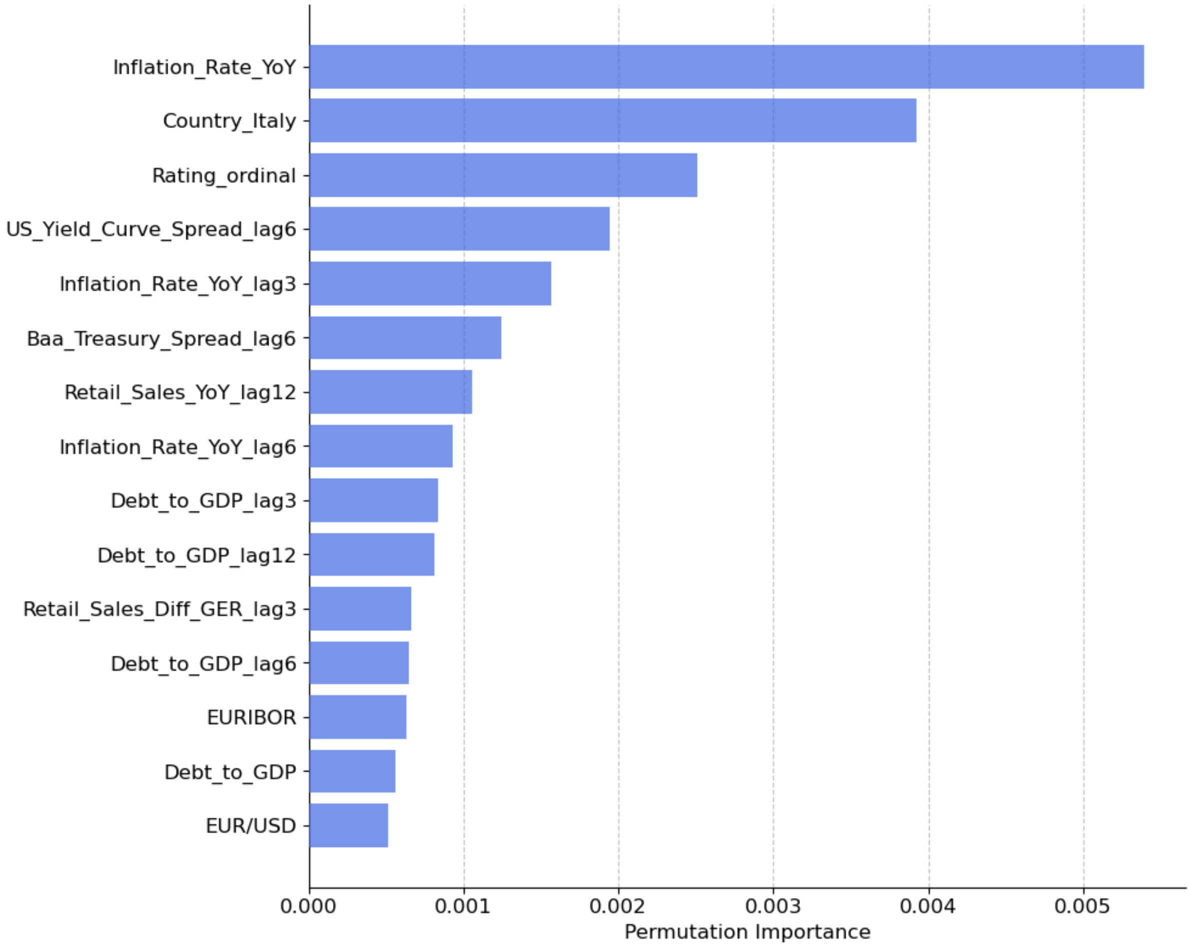
Figure 4.9: Random Forest variables importance (excluding Greece)




Source: Own illustration

Figure 4.10 shows the updated feature importance in the SVR (RBF) model after removing Greece. Inflation remains the most important factor, with an even stronger influence than before. The dummy variable for Italy (Country_Italy) is now the second most important feature. Therefore, country-specific factors become more important. Surprisingly, the CDS spread variables disappear from the top variables. This suggests that CDS spreads were only important for Greece's situation and not the other countries. Other market-based variables like the U.S. Yield Curve spread and the Baa treasury spread increase in importance. Also, more debt-to-GDP variables gain importance, even if they are still less relevant. Overall, the SVR (RBF) model without Greece shows more diverse economic indicators, which now even include retail sales data.

Figure 4.10: Support Vector Regression (RBF) variable importance (excluding Greece)




Source: Own illustration

All in all, removing Greece improved model accuracy. Error rates dropped for all models, and $R^2$ increased for SVR. RF had lower errors but a small drop in $R^2$. In variable importance, RF still relied mostly on CDS spreads, but unemployment factors became slightly more important. SVR (RBF) focused more on inflation and gave more weight to Italy's specific risks. Global financial indicators, like the U.S. yield curve spread, also became more relevant. Without Greece, the models respond more to broader economic trends instead of extreme country-specific risks. This shows that Greece's debt crisis had a strong influence on bond spreads, and removing it helps the models work better for other euro area countries.

# 5 Discussion

This study analyzed the prediction of 10-year sovereign bond spreads in the euro area with RF and SVR models. The results showed that RF achieved the highest accuracy with the lowest error metrics and highest $R^2$ statistic overall. SVR (RBF) performed also well and way better than the linear SVR model. This confirms that bond spreads have non-linear relationships with its predictors. Removing Greece from the dataset improved model accuracy for all models, in particular for SVR (RBF). This suggests that Greece's high bond spreads added noise to the models. The results confirm important findings about the models and their ability to predict bond spreads. **H1** stated that RF and SVR (RBF) would perform better than a linear model because they can capture non-linear patterns. The results clearly support this. Both RF and SVR (RBF) had lower prediction errors than the linear SVR model. **H2** suggested that macroeconomic factors like debt-to-GDP, GDP growth, and inflation are the main drivers of bond spreads. The results partly support this. Credit risk, especially CDS spreads and credit ratings, played a big role too. Inflation and unemployment were the most important macroeconomic factors. In the RF model, CDS spreads were the top factor, but unemployment also ranked high. The SVR (RBF) model focused more on inflation and credit ratings. This highlights that each model had different factors in their prediction model. Debt-to-GDP had only a minor role in the SVR model, while GDP growth had no influence at all. **H3** proposed that removing Greece would improve model accuracy and change the importance of key predictors. The results strongly support this. Excluding Greece reduced errors in all models. The biggest improvements had the SVR (RBF) and the linear model. Both of these models had struggled with Greece's extreme bond spreads. RF error metrics improved, but its $R^2$ dropped slightly. Likely because Greece contributed to overall spread variance. The importance of key features also shifted. RF continued to rely on CDS spreads, while SVR (RBF) gave more weight to inflation and other global financial indicators. These results suggest that Greece's unique debt crisis made predictions harder. Removing it allowed the models to better capture broader economic and financial trends. The findings in my analysis align with Belly et al. (2022), who also show that ML models predict sovereign bond spreads in Europe better than linear models. Similar to my results, where the tree-based model performed best, they find that boosting and bagging models like RF and XGBoost perform relatively well compared to linear models.[1] However, Eijffinger and Pieterse-Bloem (2023) used traditional regression methods rather than ML techniques. Their findings explained sovereign bond spreads also well, which means ML can also explain spreads accurately. They also find that excluding Greece helped to explain spreads better and noticed the special role of Greece as an outlier.[2] This aligns with the findings of De Grauew et al. (2017), they confirm Greece as an outlier and show that excluding Greece reduces the variance in the data.[3] When it comes to the most important factors influencing sovereign bond spreads in Europe, other literature finds that spreads are primarily driven by market-based indicators like credit risk rather than macroeconomic factors. The paper by Guirola and Pérez (2023) suggests that the relationship between public debt fundamentals and sovereign bond spreads weakened after the 2010–2012 European sovereign debt crisis. Their findings indicate that after 2012, bond spreads were more influenced by factors such as market sentiment and policy interventions.[4] This connects to my analysis because credit risk, measured by CDS spreads and credit ratings, plays a significant role in my models besides macroeconomic factors. Inflation rate and the unemployment rate rank also high, but credit risk remains the most influential factor in determining bond spreads. Eijffinger and Pieterse-Bloem (2023) confirm the importance of CDS spreads and other market-based indicators like liquidity risk as the most important drivers of spreads. Additionally, they highlight the importance of ECB policies, particularly asset purchases by the ECB.[5] My models were only slightly influenced by monetary policy data, with the interest rate difference between Federal Reserve and ECB playing a small role. However, other literature confirms the importance of monetary policy in determining bond spreads, including studies by De Grauwe et al. (2017), Afonso et al. (2020), and Belly et al. (2022). In particular, monetary announcements from the ECB can have a significant impact on spreads. Belly et al. (2022) demonstrate this through a novel measure capturing the sentiment of ECB members.[6] Similarly, Afonso et al. (2020) find that key ECB policy rate announcements strongly influence bond spreads.[7] In my analysis, I included a variety of monetary policy variables, such as the Euroibor, the interest rate difference between Federal Reserve and ECB, net asset purchases by the ECB, and dummy and interaction terms to capture the QE regime. However, my models were not significantly influenced by these variables, which contrasts with findings in the existing literature. Macroeconomic data played a significant role in both of my models, with inflation and unemployment being particularly influential. This contrasts somewhat with Eijffinger and Pieterse-Bloem (2023),

[1] cf. Belly et al., 2022, p. 24.
[2] cf. Eijffinger and Pieterse-Bloem, 2023, p. 9.
[3] cf. De Grauwe, Ji, and Macchiarelli, 2017, Chapter 2.
[4] cf. Guirola and Pérez, 2023, pp. 3976–3977.
[5] cf. Eijffinger and Pieterse-Bloem, 2023, p. 14.
[6] cf. Belly et al., 2022, p. 30.
[7] cf. António Afonso, Jalles, and Kazemi, 2020, p. 11.

who find that while these macroeconomic indicators have some impact, their influence is weaker than in my models.[8] However, Belly et al. (2022) also identify unemployment as a key factor in their cross-sectional analysis, alongside redenomination risk.[9] The discussed studies here in this chapter use similar data, time periods, and frequencies. But they apply different methods, making direct comparisons hard. In my analysis, I used the IQR method to remove outliers, limiting spreads to the 0 to 3 range. Higher spreads were not included, which may explain some differences. Because of this, comparing evaluation metrics with Belly et al. (2022) could lead to wrong conclusions. The linear SVR model should be viewed with caution. Even though I reduced multicollinearity, it is only a basic benchmark model. It is not a fully optimized model. There are also ways to improve my study. A longer time period and adding ECB speech data could help. Future research could also test other ML models, like XGBoost, Artificial Neural Networks (ANN), and alternative linear models. Furthermore, there is still an open question about using CDS data. Different studies use it in different ways. There is no clear answer on how to handle it. Also, the differences in variable importance across models mean that my results must be interpreted carefully. Feature importance is not absolute and should be seen as a rough guide and not as a final answer.

[8] cf. Eijffinger and Pieterse-Bloem, 2023, p. 9.
[9] cf. Belly et al., 2022, p. 29.

# 6 Conclusion

In this study, I investigated how well ML can predict 10-year sovereign bond spreads in the euro area. I used RF, SVR with a radial kernel, and a simple linear SVR model to compare these models with each other. My goal was to see if ML models can better predict sovereign bond spreads than linear models. I also analyzed which predictors are most important for bond spreads in Europe. To see how Greece's sovereign debt crisis affects the results, I compared my findings with and without Greece as an outlier. First, the RF model worked the best. It had the smallest evaluation metrics and explained the most with an $R^2$ score of 0.78. The SVR model with a radial kernel was also good in predicting spreads, with an $R^2$ of 0.73. The linear model didn't do well in the exercise including Greece, with an $R^2$ of just 0.40. This first part showed that bond spreads don't follow a linear relationship and are more complex. ML techniques like RF and SVR can handle that non-linearity better than the linear model. Second, I analyzed which factors matter the most for bond spreads. Credit risk was the most important factor. In the RF model, CDS spreads were the top factor for this. Unemployment was also highly important in that model, showing an unstable job market can increase spreads. In the SVR (RBF) model, inflation was most important in determining spreads. Credit risk in the form of sovereign credit ratings was the second most influential in this model. Debt-to-GDP was less influential, but not as much as I thought it might. Variables like interest rate differences between the U.S. and Europe also had a small influence, hinting that monetary policy data played a role. Overall, spreads are determined by a mix of credit risk, macroeconomic indicators, and monetary policy conditions. When I removed Greece from the dataset, all models showed better predictive accuracy. Greece had very high bond spreads during the debt crisis, which made errors larger. For example, the MSE for RF dropped from 0.079 to 0.061, and for SVR (RBF) it went down from 0.098 to 0.053. The SVR (RBF) model improved the most, with its $R^2$ increasing to 0.746. These changes suggest that Greece's extreme bond spreads were outliers that hid the real patterns in the data. Without Greece, the RF model focused more on CDS spreads and unemployment, while SVR (RBF) put more weight on inflation and even showed Italy's unique risk. Overall, ML and especially RF, is really good at predicting bond spreads. Its performance is way better than the linear model.

# Data and Code Availability

The dataset and the Jupyter Notebook containing the code used in this analysis are publicly available on GitHub. They can be accessed at:

`https://github.com/dpruess/sovereign-bond-spreads-ml`

# Bibliography

Afonso, Antonio, Michael Arghyrou, and Alexandros Kontonikas (2012). “The Determinants of Sovereign Bond Yield Spreads in the EMU”. en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2223140. URL: http://www.ssrn.com/abstract=2223140 (visited on 01/17/2025).

Afonso, António, João Tovar Jalles, and Mina Kazemi (Sept. 2020). “The effects of macroeconomic, fiscal and monetary policy announcements on sovereign bond spreads”. en. In: *International Review of Law and Economics* 63, p. 105924. ISSN: 01448188. DOI: 10.1016/j.irle.2020.105924. URL: https://linkinghub.elsevier.com/retrieve/pii/S014481882030137X (visited on 01/20/2025).

Arakelian, Veni et al. (June 2019). “Sovereign risk zones in Europe during and after the debt crisis”. en. In: *Quantitative Finance* 19.6, pp. 961–980. ISSN: 1469-7688, 1469-7696. DOI: 10.1080/14697688.2018.1562197. URL: https://www.tandfonline.com/doi/full/10.1080/14697688.2018.1562197 (visited on 01/11/2025).

Awad, Mariette and Rahul Khanna (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. en. Berkeley, CA: Apress. ISBN: 978-1-4302-5989-3. DOI: 10.1007/978-1-4302-5990-9. URL: http://link.springer.com/10.1007/978-1-4302-5990-9 (visited on 01/21/2025).

Beber, Alessandro, Michael W. Brandt, and Kenneth A. Kavajecz (Mar. 2009). “Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market”. en. In: *Review of Financial Studies* 22.3, pp. 925–957. ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/hhm088. URL: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhm088 (visited on 02/11/2025).

Belly, Guillaume et al. (2022). “Forecasting Sovereign Risk in the Euro Area via Machine Learning”. en. In: DOI: http://dx.doi.org/10.2139/ssrn.3974515.

Bernoth, Kerstin, Jürgen Von Hagen, and Ludger Schuknecht (Sept. 2012). “Sovereign risk premiums in the European government bond market”. en. In: *Journal of International Money and Finance* 31.5, pp. 975–995. ISSN: 02615606. DOI: 10.1016/j.jimonfin.2011.12.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0261560611001914 (visited on 01/28/2025).

Costantini, Mauro, Matteo Fragetta, and Giovanni Melina (Aug. 2014). “Determinants of sovereign bond yield spreads in the EMU: An optimal currency area perspective”.

en. In: *European Economic Review* 70, pp. 337–349. ISSN: 00142921. DOI: 10.1016/j.euroecorev.2014.06.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0014292114000919 (visited on 01/07/2025).

Credit Benchmark (2016). *Measurement of Sovereign Credit Quality*. Tech. rep. 6. Credit Benchmark. URL: https://www.creditbenchmark.com/wp-content/uploads/2016/09/Sovereign-Credit-Quality-Sep-2016.pdf.

De Grauwe, Paul, Yuemei Ji, and Corrado Macchiarelli (2017). "Fundamentals versus market sentiments in the euro bond markets: implications for QE". In: *London School of Economics* Systemic Risk Center Special Paper No 12. ISSN: 2055-0375.

Efron, Bradley and Trevor J. Hastie (2016). *Computer age statistical inference: algorithms, evidence, and data science*. eng. Institute of mathematical statistics monographs 5. Cambridge, NY: Cambridge university press. ISBN: 978-1-107-14989-2.

Eijffinger, Sylvester C.W. and Mary Pieterse-Bloem (Dec. 2023). "Eurozone government bond spreads: A tale of different ECB policy regimes". en. In: *Journal of International Money and Finance* 139, p. 102965. ISSN: 02615606. DOI: 10.1016/j.jimonfin.2023.102965. URL: https://linkinghub.elsevier.com/retrieve/pii/S0261560623001663 (visited on 12/15/2024).

European Central Bank (2013). *Reference interest rates: role, challenges and outlook*. Tech. rep. Monthly Bulletin, October 2013. European Central Bank (ECB), pp. 69–84. URL: https://www.ecb.europa.eu/pub/pdf/other/art2_mb201310en_pp69-84en.pdf.

Favero, Carlo and Alessandro Missale (Apr. 2012). "Sovereign spreads in the eurozone: which prospects for a Eurobond?: SOVEREIGN SPREADS". en. In: *Economic Policy* 27.70, pp. 231–273. ISSN: 02664658. DOI: 10.1111/j.1468-0327.2012.00282.x. URL: https://academic.oup.com/economicpolicy/article-lookup/doi/10.1111/j.1468-0327.2012.00282.x (visited on 01/17/2025).

Feurer, Matthias and Frank Hutter (2019). "Hyperparameter Optimization". en. In: *Automated Machine Learning*. Ed. by Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. Series Title: The Springer Series on Challenges in Machine Learning. Cham: Springer International Publishing, pp. 3–33. ISBN: 978-3-030-05317-8. DOI: 10.1007/978-3-030-05318-5_1. URL: http://link.springer.com/10.1007/978-3-030-05318-5_1 (visited on 02/19/2025).

Fontana, Alessandro and Martin Scheicher (Jan. 2016). "An analysis of euro area sovereign CDS and their relation with government bonds". en. In: *Journal of Banking & Finance* 62, pp. 126–140. ISSN: 03784266. DOI: 10.1016/j.jbankfin.2015.10.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378426615002988 (visited on 02/11/2025).

Frey, Eric and Ben Seimon (June 2024). *panelsplit*. DOI: 10.5281/ZENODO.11583542. URL: https://zenodo.org/doi/10.5281/zenodo.11583542 (visited on 02/20/2025).

Georgoutsos, Dimitris A. and Petros M. Migiakis (Nov. 2013). “Heterogeneity of the determinants of euro-area sovereign bond spreads; what does it tell us about financial stability?” en. In: *Journal of Banking & Finance* 37.11, pp. 4650–4664. ISSN: 03784266. DOI: 10.1016/j.jbankfin.2013.07.025. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378426613002951 (visited on 01/20/2025).

Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu (2018). “Empirical Asset Pricing Via Machine Learning”. en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3159577. URL: https://www.ssrn.com/abstract=3159577 (visited on 01/28/2025).

Guirola, Luis and Javier J. Pérez (July 2023). “The decoupling between public debt fundamentals and bond spreads after the European sovereign debt crisis”. en. In: *Applied Economics* 55.34, pp. 3971–3979. ISSN: 0003-6846, 1466-4283. DOI: 10.1080/00036846.2022.2120959. URL: https://www.tandfonline.com/doi/full/10.1080/00036846.2022.2120959 (visited on 01/25/2025).

Han, Jiawei, Jian Pei, and Micheline Kamber (2012). *Data Mining*. en. Elsevier. ISBN: 978-0-12-381479-1. DOI: 10.1016/C2009-0-61819-5. URL: https://linkinghub.elsevier.com/retrieve/pii/C20090618195 (visited on 02/19/2025).

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7. URL: http://link.springer.com/10.1007/978-0-387-84858-7 (visited on 02/12/2025).

James, Gareth et al. (2023). *An Introduction to Statistical Learning: with Applications in Python*. en. Springer Texts in Statistics. Cham: Springer International Publishing. ISBN: 978-3-031-38746-3. DOI: 10.1007/978-3-031-38747-0. URL: https://link.springer.com/10.1007/978-3-031-38747-0 (visited on 01/10/2025).

Kilponen, Juha, Helinä Laakkonen, and Jouko Vilmunen (2012). “Sovereign Risk, European Crisis Resolution Policies and Bond Yields”. en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2102461. URL: http://www.ssrn.com/abstract=2102461 (visited on 01/17/2025).

Kursa, Miron B. and Witold R. Rudnicki (2010). “Feature Selection with the **Boruta** Package”. en. In: *Journal of Statistical Software* 36.11. ISSN: 1548-7660. DOI: 10.18637/jss.v036.i11. URL: http://www.jstatsoft.org/v36/i11/ (visited on 02/18/2025).

Matei, Iuliana and Angela Cheptea (Jan. 2012). “Sovereign Bond Spread Drivers in the EU Market in the Aftermath of the Global Financial Crisis”. en. In: *Advances in Econometrics*. Ed. by Badi H. Baltagi et al. Vol. 29. Emerald Group Publishing Limited, pp. 327–352. ISBN: 978-1-78190-307-0. DOI: 10.1108/S0731-9053(2012)0000029017. URL: https://www.emerald.com/insight/content/doi/10.1108/S0731-9053(2012)0000029017/full/html (visited on 01/20/2025).

Pagano, M. (Dec. 2004). "The European Bond Markets under EMU". en. In: *Oxford Review of Economic Policy* 20.4, pp. 531–554. ISSN: 1460-2121. DOI: 10.1093/oxrep/grh031. URL: https://academic.oup.com/oxrep/article-lookup/doi/10.1093/oxrep/grh031 (visited on 01/27/2025).

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine learning in python". In: *Journal of Machine Learning Research* 12.85, pp. 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

Willmott, Cort J. and Kenji Matsuura (2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". In: *Climate Research* 30.1. Publisher: Inter-Research Science Center, pp. 79–82. ISSN: 0936577X, 16161572. URL: http://www.jstor.org/stable/24869236 (visited on 02/20/2025).

# Appendix

Table A.1: List of variables used in the models

| Name | Description | Frequency | Source |
|---|---|---|---|
| Sovereign Bond Spreads | The difference between a country's long-term interest rate on government bonds maturing in ten years and the corresponding rate in Germany. | Monthly | OECD |
| Inflation Rate | The annual percentage change in the Consumer Price Index. | Monthly | S&P Capital IQ |
| Unemployment Rate | The percentage of the total labor force that is actively seeking employment. | Monthly | S&P Capital IQ |
| Retail Sales | A monthly measure of retail performance based on a sample of retail stores of various sizes and types. | Monthly interannual rate of growth | S&P Capital IQ |
| GDP Real Growth Rate | The annual percentage change in a country's Real GDP. | Monthly (interpolated from quarterly observations) | S&P Capital IQ |

*Continued on next page*

Table A.1: List of variables used in the models

| Name | Description | Frequency | Source |
|---|---|---|---|
| Consumer Confidence Index | Measures household sentiment on financial conditions, economic outlook, unemployment, and savings. Values above 100 indicate optimism, while values below 100 indicate pessimism. | Monthly | OECD |
| Current Account Balance to GDP | A measure of a country's net international transactions relative to its GDP. | Monthly (interpolated from quarterly observations) | OECD |
| Government Debt to GDP | The total consolidated debt of the government as a percentage of GDP. | Monthly (interpolated from quarterly observations) | ECB Data Portal |
| Government Debt Growth Rate | The year-over-year percentage change in consolidated government debt. | Monthly (interpolated from quarterly observations) | ECB Data Portal |
| Stock Market Returns | The monthly and yearly percentage change in the benchmark stock index. | Monthly | S&P Capital IQ |
| CBOE Volatility Index | A measure of expected 30-day volatility of the U.S. stock market, derived from S&P 500 option prices. | Monthly average of daily data | FRED |
| US Yield Spread | The difference between the U.S. 10-year and 1-year Treasury yields at the end of each period. | Monthly | ECB Data Portal |
| Baa Corporate Bond Spread | The yield difference between Moody's Seasoned Baa Corporate Bonds and the 10-year U.S. Treasury. | Monthly average of daily data | FRED |

*Continued on next page*

Table A.1: List of variables used in the models

| Name | Description | Frequency | Source |
|---|---|---|---|
| Eurosystem Total Assets/Liabilities Change | Measures the month-to-month change in total assets or liabilities reported by the Eurosystem. | Monthly | ECB Data Portal |
| Credit Default Swap Spreads | The cost of protection against the default of a borrower or bond issuer over a 5-year period. | Monthly | S&P Capital IQ |
| Credit Rating (Numeric Average) | The average credit rating from S&P, Moody's, and Fitch, converted into a numeric scale from 1 (AAA, highest) to 22 (D, default). | Monthly | World Government Bonds |
| Credit Rating (Probability of Default) | The annualized probability of default derived from S&P, Moody's, and Fitch ratings, using values from the Credit Benchmark report. | Monthly | Credit Benchmark Report |
| Political Stability and Absence of Violence/Terrorism Index | Measures perceptions of political instability, violence, and terrorism. Scores range from -2.5 (high instability) to 2.5 (high stability). | Monthly (interpolated from yearly observations) | World Bank |
| Real Effective Exchange Rate | Measures a country's inflation-adjusted currency value relative to its trading partners, using trade-weighted averages. | Monthly | BIS Data Portal |
| Euro Interbank Offered Rate | The average interest rate at which leading European banks lend unsecured funds to one another in the interbank market (3-month term). | Monthly | S&P Capital IQ |

*Continued on next page*

Table A.1: List of variables used in the models

| Name | Description | Frequency | Source |
|---|---|---|---|
| FED-ECB Interest Rate Differential | The difference between the key interest rates set by the Federal Reserve (U.S.) and the European Central Bank (ECB). | Monthly | BIS Data Portal |
| Total ECB Asset Purchases | The combined value of securities purchased under the ECB's Pandemic Emergency Purchase Programme (PEPP) and Public Sector Purchase Programme (PSPP). | Monthly | ECB |

Table A.2: Optimal Hyperparameters from GridSearchCV

| Model | With Greece | Without Greece |
|---|---|---|
| Random Forest | max_depth: 10<br>max_features: 0.2<br>n_estimators: 500 | max_depth: 5<br>max_features: 0.2<br>n_estimators: 500 |
| SVR Radial | C: 1.2<br>epsilon: 0.08 | C: 7<br>epsilon: 0.05 |
| SVR Linear | C: 12<br>epsilon: 0.03 | C: 8<br>epsilon: 0.005 |

Table A.3: Credit ratings mapped to numeric ratings and annualized probabilities of default (PD)

| **Credit Rating** | **Numeric Rating** | **PD (%)** |
|---|---|---|
| AAA | 1 | 0.00 |
| AA+ | 2 | 0.0125 |
| AA | 3 | 0.02 |
| AA- | 4 | 0.03 |
| A+ | 5 | 0.04 |
| A | 6 | 0.06 |
| A- | 7 | 0.08 |
| BBB+ | 8 | 0.13 |
| BBB | 9 | 0.20 |
| BBB- | 10 | 0.30 |
| BB+ | 11 | 0.48 |
| BB | 12 | 0.74 |
| BB- | 13 | 1.35 |
| B+ | 14 | 2.50 |
| B | 15 | 4.20 |
| B- | 16 | 7.50 |
| CCC+ | 17 | 12.00 |
| CCC | 18 | 18.50 |
| CCC- | 19 | 29.93 |
| CC | 20 | 48.43 |
| C | 21 | 78.36 |
| D | 22 | 100.00 |

# Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche einzeln kenntlich gemacht. Es wurden keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Osnabrück, 28.02.2025 Unterschrift: D. Prüß