Barcelona School of Economics

**Master´s Degree in Economics and Finance**
**Specialization in Finance**

# Forecasting Redenomination Risk:

## Unleashing the superior predictor via Machine Learning


Authors: Frederik Witzemann, Astthor Arnorsson

Supervisors: Filippo Ippolito; Christian Brownlees


11/06/2023

## ABSTRACT IN ENGLISH

We are the first to construct a general model for forecasting redenomination risk by using the difference in a country's EUR-denominated CDS spreads compared to Germany. The goal of this work is to further close this gap by harnessing the power of machine learning algorithms, with a particular focus on techniques such as Random Forests, XGBoosting, and Lasso regressions. The results have important implications for policymakers, financial institutions and investors. Relying mainly on daily financial market data, we find that machine learning models significantly improve forecasting accuracy in both a structural crisis and a period of economic recovery. We conclude that XGB and Lasso are the best performing forecasting models, as for the latter modeling approach we get correlations between predicted and actual values as high as 0.6 for some countries. Finally, when we look at the measures of variable importance in our machine learning models, we indeed find that Lasso selects different types of predictors with much larger macroeconomic reference as relevant, which seems to explain the difference in prediction accuracy.

## ABSTRACT IN CATALAN/ SPANISH:

Somos los primeros en construir un modelo general para pronosticar el riesgo de redenominación utilizando la diferencia entre los diferenciales de los CDS denominados en euros de un país en comparación con Alemania. El objetivo de este trabajo es cerrar aún más esta brecha aprovechando el poder de los algoritmos de aprendizaje automático, con especial atención en técnicas como Random Forests, XGBoosting y regresiones Lasso. Los resultados tienen implicaciones importantes para los responsables de la formulación de políticas, las instituciones financieras y los inversores. Basándonos principalmente en datos diarios del mercado financiero, encontramos que los modelos de aprendizaje automático mejoran significativamente la precisión de los pronósticos tanto en una crisis estructural como en un período de recuperación económica. Concluimos que XGB y Lasso son los modelos de pronóstico de mejor rendimiento, ya que para el último enfoque de modelado obtenemos correlaciones entre los valores previstos y reales de hasta 0,6 para algunos países. Finalmente, cuando observamos las medidas de importancia variable en nuestros modelos de aprendizaje automático, encontramos que Lasso selecciona diferentes tipos de predictores con una referencia macroeconómica mucho mayor como relevantes, lo que parece explicar la diferencia en la precisión de la predicción.

**KEYWORDS IN ENGLISH:**

Sovereign Risk, Credit Default Swaps, Euro Zone, Forecasting, Time Series, Machine Learning, Random Forests, XGBoosting, Lasso.

**KEYWORDS IN CATALAN/ SPANISH:**

Riesgo soberano, Swaps de incumplimiento crediticio, zona euro, previsión, series temporales, aprendizaje automático, bosques aleatorios, XGBoosting, Lasso.

# Forecasting Redenomination Risk

**Unleashing the superior predictor via Machine Learning**

Frederik Witzemann
Astthor Arnorsson

11th of June
Supervisor: Christian Brownlees

## Abstract

We are the first to construct a general model for forecasting redenomination risk by using the difference in a country's EUR-denominated CDS spreads compared to Germany. The goal of this work is to further close this gap by harnessing the power of machine learning algorithms, with a particular focus on techniques such as Random Forests, XGBoosting, and Lasso regressions. The results have important implications for policymakers, financial institutions and investors. Relying mainly on daily financial market data, we find that machine learning models significantly improve forecasting accuracy in both a structural crisis and a period of economic recovery. We conclude that XGB and Lasso are the best performing forecasting models, as for the latter modeling approach we get correlations between predicted and actual values as high as 0.6 for some countries. Finally, when we look at the measures of variable importance in our machine learning models, we indeed find that Lasso selects different types of predictors with much larger macroeconomic reference as relevant, which seems to explain the difference in prediction accuracy.

**Keywords**: Sovereign Risk, Credit Default Swaps, Euro Zone, Forecasting, Time Series, Machine Learning, Random Forests, XGBoosting, Lasso.

# Contents

**1 Introduction** **5**

**2 CDS spreads and Redenomination Risk - A Literature Review** **5**

**3 Data, methodology and introductory analyses** **8**

3.1 Data . . . . . 8

3.2 Methodology . . . . . 8

3.3 Introductory Analyses . . . . . 9

**4 Finding the Baseline model** **10**

4.1 Full Sample . . . . . 11

4.2 Debt Crisis . . . . . 11

**5 Adding Machine Learning** **12**

5.1 Full Sample . . . . . 14

5.2 Debt Crisis . . . . . 16

**6 Extending Machine Learning** **19**

6.1 Full sample . . . . . 19

6.2 Debt Crisis . . . . . 21

6.3 Optimal h-step ahead forecasts . . . . . 22

**7 Discussion** **24**

**8 Conclusion** **27**

**9 Appendix** **29**

9.1 Graphs . . . . . 29

9.2 Tests and Robustness Checks . . . . . 39

9.3 Additional Information . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

# 1 Introduction

The fiscal developments during the European debt crisis and the pandemic have shown that the continued existence of the monetary union is anything but certain. In fact, the past has shown that the political instability of just one European economy can jeopardize the monetary union to a significant extent, even outside of a crisis. Therefore, the risk of a country leaving the euro and devaluing into its national currency (e.g. Greece or Italy) has always been a constant threat. This risk can be measured in a number of ways, such as the huge capital outflows in the case of Greece in 2010 [Hetzel, 2015, p. 7]. There are also other financial or capital market instruments that measure this specific component of sovereign risk of a country, such as credit default swaps (CDS). Investors associate credit default swaps with higher risk than ratings from rating agencies because these market-based instruments represent the cost of hedging for a given country and appear therefore to be more accurate in predicting sovereign risks [Aizenman et al., 2016].
To measure this risk, one requires the difference between the insurance price, also known as the "CDS spread", and a spread measure of the same unit, which is considered risk-free. The term introduced by [De Santis, 2019] also refers to the risk of redenomination, which will be the focus of the paper.
Although redenomination risk has generally been an important explanatory factor in previous works, e.g. [De Santis, 2019] and [Corradin et al., 2021], so far there is no literature that constructs a general model that explains redenomination risk or examines its forecasting behavior. In fact, it is of interest to investors and policymakers to closely follow the trends of a measure of redenomination risk such as CDS spreads. To optimize this approach, we also use non-linear machine learning models, which have also recently been applied in related topics such as forecasting eurozone government bond spreads [Belly et al., 2021].
By basing our work on the cited literature, we divide our analyses into two main exercises, namely predicting the redenomination risk of a given country based only on that country's past and predicting the future based on a model covering all countries. Finally, once we have identified the pattern across countries, models, and sub-samples, we determine the optimal H-step-ahead forecast for a given country.

# 2 CDS spreads and Redenomination Risk - A Literature Review

We base our idea and the main explanatory variables on the work of De Santis: "Measure of Redenomination risk" [De Santis, 2019]. The work focuses on quanto CDS spreads, which measure the difference between USD and EUR-denominated CDS spreads. Assuming the same reference entity and ensuring comparability of both instruments, the difference accurately reflects the risk that, for example, Italy would devalue into its national currency. Logically, this value would be positive, especially during the debt crisis, since the payout in a EUR-denominated credit risk is less attractive [De Santis, 2019, p.2]. Instead of creating a general model by examining the impact of each variable, de Santis created a shocked

SVAR model that aims to examine the relationship between the market's perception of redenomination risk and redenomination risk itself.

In addition, [Corradin et al., 2021] use a redenomination risk analysis by decomposing government bond yields into various sub-components, including redenomination risk. Although this exercise was more accounting-related, it revealed valuable links between redenomination risk and fiscal and monetary policy. Putting a stronger focus on the European debt crisis, [Klose and Weigert, 2014] show that the persistence of redenomination risk was a quantitatively significant component of government bond yields for nine eurozone countries during this particular period. Furthermore, they note that in the event of a eurozone breakup, the newly introduced local currencies of countries such as Portugal, Ireland, Spain and Italy were to be expected to depreciate, while the currencies of the remaining countries [Klose and Weigert, 2014] would appreciate. Finally, Lando and Nielsen, in an elaborate paper entitled Quanto CDS Spreads, show that such spreads also depend on the risk of an exchange rate jump. Comparing the impact on the government bonds of four eurozone sovereigns, they find that the bond returns of all currency denominations using standard FX forward hedges are actually missing a key quanto effect component [Lando and Bang Nielsen, 2018].

In fact, most of the work that use CDS spreads to measure fiscal or sovereign risk examine relationships at the monthly level. For example, a publication by the IMF shows that the fundamentals affecting CDS spreads as a measure of sovereign risk change over time. While weak economic fundamentals were responsible for the rise in CDS spreads during the 2008/2009 crisis, the declines in CDS spreads after the European debt crisis in 2012 were mainly due to falling risk aversion [Heinz and Sun, 2014]. In the IMF working paper, for example, the authors use explanatory variables such as the budget deficit or the current account deficit, which is of great importance for the observations during the European debt crisis and thus also for our own modeling approach. While entirely basing our paper on macroeconomic variables of monthly nature could provide important insights, such as the relationship between redenomination risk and debt to GDP levels, it could also cause the issue of having too little observations which could degrade the results of our "data-hungry" machine learning models. Therefore, we continue to consider these monthly macroeconomic variables, but only as an additional component, as we consider daily financial market variables to be our main predictors.

There is little research on sovereign risk in the eurozone and the interface to machine learning. However, noteworthy is the work of Arakelian, which applied Random Forest-based modeling approaches to create a European country risk stratification [Arakelian et al., 2019]. Using CDS spreads, they identify specific risk zones ranging from safe to risky based on important, in contrast to our approach, macroeconomic predictors using decision trees. In contrast, the methodology of [Belly et al., 2021] is strongly linked to our work and our approaches. Using data on monthly government bond spreads, they forecast government bond spreads for all euro area countries using forecasting techniques such as rolling windows. Furthermore, [Belly et al., 2021] also exploit the risk of redenomination by using quanto CDS spreads as an explanatory variable for their forecasting model.

To find the desired measure of redenomination risk, we could simply rely on De Santis' quanto CDS spreads, or obtain our measure of redenomination risk as the difference to the German CDS spread for the relevant country. We choose the latter for several reasons.

First, this measurement not only satisfies the “risk-free” property already emphasized in the introduction, but also ensures the absence of currency risk. Second, and most importantly, it also provides greater consistency and transparency in terms of liquidity (since CDS vary in their liquidity across currencies).

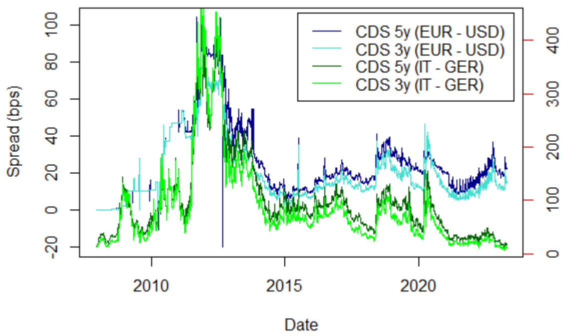
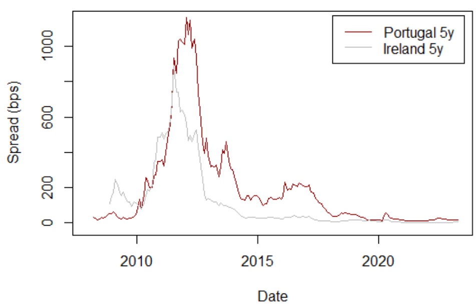
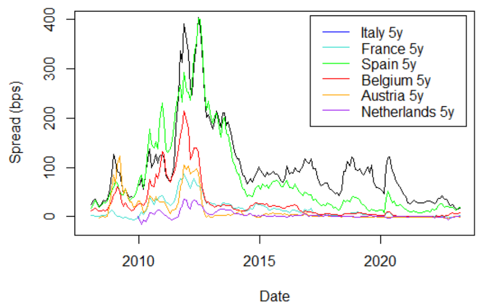



Figure 1: Different measures of redenomination risk, derived from Datastream (daily)

Comparing both approaches, it is clear that both follow an almost identical pattern, with small deviations during the European debt crisis and quanto CDS spreads following slightly higher levels in relative terms during the subsequent economic recovery.




(a) Countries of "higher" redenomination Risk, derived from Datastream (monthly)




(b) Countries of "lower" redenomination Risk, derived from Datastream (monthly)

By comparing our measurement of redenomination risk between our selected eurozone countries, we are able to rank our economies based on their magnitude of CDS spreads compared to Germany. This will also be useful for subsequent analyses:
1) High spread countries: Italy, Spain, Portugal and Ireland
2) Low spread countries: France, Belgium, Austria and the Netherlands
However, throughout the economic recovery from the debt crisis, only Portugal and Italy remain at significant redenomination risk.

# 3 Data, methodology and introductory analyses

## 3.1 Data

***Capital Market Variables***
We source both our CDS spreads and almost all of our explanatory variables from Datastream. However, we derive a small set of variables from FRED St. Louis. Our analyses are performed daily, with our total sample consisting of approximately 35,000 observations in the combined panel and approximately 3,800 observations for each country. We use a total of 20 predictors, with only 4 variables being country-specific: government bond spreads, USD-denominated CDS spreads, our measure of redenomination risk (the difference in 5-year CDS spreads compared to Germany) and country-specific equity returns. All other variables such as breakeven inflation rate or volatility premium refer either to the eurozone as a whole or to the US. By lagging each of these variables by 1 to 5 days (corresponding to one week) as well as including factor dummy variables for each country, we end up with 95 different predictors in our combined panel set. Our sample is of course limited to the period December 2007 to present, as this is the earliest sample size available for CDS spreads. Therefore, since all remaining variables are accessible earlier, we do not have to deal with missing observations, making our analyses simpler and our predictions more robust.
***Macroeconomic Variables***
In addition to our key capital market predictors of daily frequency, we also consider a number of macroeconomic variables proposed by other authors as [De Santis, 2019]. For example, [Montes et al., 2022] use debt to GDP ratio, real interest rate and foreign exchange reserves to predict CDS spreads on a quarterly basis. We further extend the [De Santis, 2019] model with these variables, albeit on a monthly basis, by linearly interpolating our quarterly observations for debt-to-GDP. This should allow us to gain further insight into how our daily measurement of CDS spreads responds to information at the monthly level. In addition, we also consider survey-based indicators such as each country's economic sentiment, as [Montes et al., 2022] show the relationship between CDS spreads and sentiment towards a country's macroeconomic condition. Finally, as the plots in our literature review show, the risk of redenomination in the eurozone was enormous during the European debt crisis, when the public had serious concerns about a possible breakup of the monetary union. That's why we include the Euro Zone Break-up Index, which measures the probability of this exact situation occurring. However, because this index was created in 2012, during the peak of the debt crisis, we will only use this variable for our full sample predictions.

## 3.2 Methodology

Our main sub-samples or periods of interest are the European debt crisis (test years 2011 and 2012) and the overall sample, which we call the "full sample", which does not extend to our last observation in the data, but uses the years from 2007 to 2017 as training - and

validation window and tests the models from 2017 to 2019. We do this for several reasons. First, since our goal is to compare the forecasting accuracy of redenomination risk between countries and models between an economic crisis and a period of economic stability, we refrain from stretching our "overall sample" to the last observation because we would test our models against very volatile data from the pandemic. Second, we also want to include and evaluate the prediction accuracy of jumps in CDS spreads during the French (2017) and Italian (2018) elections [Balduzzi et al., 2020].
We follow the [Belly et al., 2021] methodology, but use daily rather than monthly data. We include further variations in our analysis by distinguishing between static time-series forecasts and rolling H-step-ahead forecasts. For example, if we aim to predict differences in CDS movements during the pandemic and therefore rely on the entire sample as training data, we also rely on a validation sample for our optimization parameters that is the same length as the testing sample. On the other hand, when we make predictions based on a rolling window, we assess H-step-ahead performance because we are always moving our training window one observation into the future.
Finally, after first implementing the machine learning capabilities by running both the training and testing analysis for the country of interest, we add a forecasting exercise where we train our machine learning models for all countries simultaneously, but only predict or test for our country of interest. This is intended not only to optimize our forecasting results, but also to provide a general overview of which hyperparameters provide the best prediction and which predictors appear to be most important overall.

## 3.3 Introductory Analyses

As far as the preliminary analyses are concerned, that of stationarity is the most important. When variables, and especially the dependent variable, are stationary, it means that our variable of interest does not follow any particular trend and has a constant mean, variance, and autocorrelation over time. To ensure this, we perform an Augmented Dicker Fuller (ADF) test. Especially since our thesis focuses so much on forecasts, it is crucial to ensure that our time series are stationary as they prevent our forecasts from deteriorating. Table 9 in the appendix provides an overview of the ADF test for all of our variables. The ADF test rejects the null of non-stationarity and therefore assumes that the series is stationary if the p-value is significantly small (less than 0.1 or 0.05). Therefore, if we look at the p-values of the ADF test in the appendix, we find that no variables have significantly low p-values that would reject the non-stationarity hypothesis. The variables with very low p-values are variables such as stock market volatility in the US or the eurozone, which were converted to a stationary series just prior to analysis because both result from a GARCH(1,1). To eliminate non-stationarity across all variables, we put our entire sample in first differences, since by subtracting the observation of t-1 from t, we remove any possible trends that the variables are subject to. Therefore, we estimate both our actual prediction models and our hyperparameter tuning for our machine learning models in first differences. Both predictions and evaluations are also made in first differences. Therefore, we forecast the change in redenomination risk in all our forecast analyses.

# 4 Finding the Baseline model

In order to categorize the performance of the machine learning models in the subsequent analyses, we perform introductory forecasting analyses to help us identify our baseline model and to give us an initial insight into the underlying pattern that determines redenomination risk across countries.
We validate our baseline models out of sample over two time periods, the overall sample and the debt crisis. By applying the same ratio between test and training data (70-30), we predict redenomination risk out of sample for each country.

We compare the performance of a simple linear model with that of an ARMA (5,5) model. We choose these two modeling approaches as possible candidates for our baseline model for several reasons:
First, the output of a linear model can be of great value in assessing the effect of the explanatory variables on our variable of interest. Because of its simple construction, we can ensure a very easy interpretation as we also present a linear machine learning model in the next section. Second, the ARMA model itself not only includes the lag values of our target variable, but also accounts for the error term of the MA model by adjusting each forecast by the error term of the previous forecast, thus accounting for recent deviations from the expected value.

$$Y_t = \omega + \sum_{t=1}^{5}(\phi_{lag_t} + \Theta_{lag_t} + \sum_{k=1}^{i} X_{i,lag_t}) + \sum_{k=1}^{i} Z_{i,t-1} + \epsilon \tag{1}$$

$\sum_{t=1}^{5} \phi_{lag_t}$, the lag error terms $\sum_{t=1}^{5} \Theta_{lag_t}$ but also of a set of daily $\sum_{k=1}^{i} X_i$ and monthly $\sum_{k=1}^{i} Z_i$ external regressors, determined based on information criteria such as AIC. As noted by $\sum_{t=1}^{5}$, our daily capital market variables range from lag 1 to 5 (corresponding to one week) whereas our monthly variables contain only the value of the previous month.

Even if the construction of the baseline models is very simple in itself, we still have to consider important causalities and procedures. First, we examine our full model (including all variables) for the persistence of multicollinearity. So, by testing each variable of the full model for its variance inflation factor (VIF) and setting the threshold to 5, we are already adjusting our predictors of both models for one important factor. Second, we try to further optimize our model fit using information criteria such as Akaike (AIC) and Bayesian Schwarz (BIC)[1]. This means that we end up with a unique set of optimal explanatory variables for each country, aiming to optimize our linear and ARMA models to the fullest.

[1]Performing the selection of predictors based on these information criteria is beneficial because they control the amount of predictors in the existing model, as opposed to traditional "goodness of fit" such as $R^2$, which increases with construction by the number of variables we include in our regression.

## 4.1 Full Sample

In our full sample analysis, our training sample ranges from December 2007 to January 2017 and the testing sample from 2017 to 2019. We assess the performance of each out-of-sample forecasting model using different metrics, such as root-mean-squared-error (RMSE) or mean-absolute-error (MAE) as well as a quality measure where we choose the correlation coefficient (CORR) between the predictions and the actual values of our test data. Therefore, if we look at Table 1, which compares our different measurements across countries, we see that both modeling approaches already show significant differences in how well explanatory variables of [De Santis, 2019] work across countries. Finally, to make our results easy to interpret, we measure and forecast our redenomination risk in basis points (bps). In other words, since all of our analyses are done in first differences, we forecast the change in redenomination risk for a given day, assessing, for example, how many basis points our forecast error differs, on average, from the actual change. This is consistent across all of our sections.

| Metric | IT | ES | FR | PT | GR | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.24 | 0.21 | 0.17 | 0.07 | 0.00 | 0.07 | 0.16 | 0.26 | 0.09 | 0.16 |
| CORR (LM) | 0.23 | 0.21 | 0.18 | 0.07 | 0.00 | 0.07 | 0.16 | 0.24 | 0.09 | 0.16 |
| MAE (ARMA) | 2.17 | 1.56 | 0.54 | 2.38 | 49.85 | 1.26 | 0.66 | 0.47 | 0.75 | 1.22 |
| MAE (LM) | 2.17 | 1.56 | 0.55 | 2.42 | 50.74 | 1.26 | 0.65 | 0.48 | 0.75 | 1.23 |
| RMSE (ARMA) | 3.16 | 2.95 | 0.87 | 3.23 | 636.33 | 1.70 | 0.89 | 0.70 | 1.11 | 1.83 |
| RMSE (LM) | 3.16 | 2.95 | 0.87 | 3.28 | 638.96 | 1.70 | 0.89 | 0.71 | 1.11 | 1.83 |

Table 1: Quality and accuracy metrics, baseline model estimation (Full Sample). Values for the columns EZ (Euro Zone) are computed by averaging all countries except for Greece.

For example, if we look at the accuracy metrics across countries, we find that both of our baseline models behave very similarly when we test our models out of sample in 2017-2019. Both forecasting models appear to be quite similar in terms of quality and accuracy measurements across countries, predicting a RMSE of 3.16 basis points in the change in redenomination risk. For now, this leads us to conclude that neither model really seems to outperform the other, but that they are in fact very identical across the sample. Furthermore, by considering the correlation between the predicted and actual values, we can additionally observe how much of a country's change in redenomination risk can be explained by [De Santis, 2019] variables. It appears that the model performs relatively well across the sample for Italy, Spain, the Netherlands and France, and less accurately for the remaining countries, with Greece being the only country where our selected predictors appear to have no effect at all.

## 4.2 Debt Crisis

Next, we consider the out of sample performance for our baseline models during the peak of the European Debt Crisis. In order to keep the training/testing set ratio consistent we train our model until May 2011 and test it for the following year until May 2012.

| Metric | IT | ES | FR | PT | GR | AT | NL | BG | IE | EZ |
|---|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.30 | 0.16 | 0.31 | 0.12 | 0.12 | 0.16 | 0.34 | 0.12 | 0.22 | 0.22 |
| CORR (LM) | 0.30 | 0.16 | 0.31 | 0.12 | 0.12 | 0.16 | 0.34 | 0.12 | 0.21 | 0.22 |
| MAE (ARMA) | 11.00 | 10.33 | 4.11 | 30.69 | 364.38 | 5.06 | 4.96 | 6.82 | 14.56 | 10.94 |
| MAE (LM) | 11.00 | 10.33 | 4.11 | 30.77 | 362.87 | 5.06 | 4.96 | 6.83 | 14.58 | 10.95 |
| RMSE (ARMA) | 15.23 | 13.88 | 5.83 | 45.11 | 835.44 | 7.79 | 9.64 | 9.53 | 22.80 | 16.23 |
| RMSE (LM) | 15.23 | 13.87 | 5.83 | 45.23 | 830.81 | 7.79 | 9.64 | 9.54 | 22.90 | 16.25 |

Table 2: Quality and accuracy metrics, baseline model estimation (Debt Crisis). Values for the columns EZ (Euro Zone) are computed by averaging all countries except for Greece.

As we test our models during the European debt crisis, the period with the highest CDS spreads in any country relative to Germany, we see a significant increase in our accuracy metrics as a result. In the case of Greece, for example, our measure of change in redenomination risk during the European debt crisis exceeds the five figures mark, making a out of sample RMSE of about 830 basis points more understandable. In addition to the increase in our accuracy metrics due to a significantly higher risk of redenomination in general, we also see an improvement in most of our correlation coefficients for all countries except Spain. Ireland also stands out, as the RMSEs tend to increase proportionally in most countries compared to the overall sample. Comparing a RMSE of 1.7 basis points across the sample to one of 22.9 basis points during the debt crisis, the volatility of redenomination risk over time becomes more tangible, as illustrated in our review of the literature when dividing countries into high- and low-spread countries. Most importantly, this analysis not only shows that redenomination risk in general during the European debt crisis appears to be better explained across countries and that the patterns in the two samples may differ across countries. In addition, it also suggests the following consideration: Predicting redenomination risk out of sample may not be ideal, since the impact of the drivers and the drivers themselves appear to be very volatile, suggesting that forecasts based on rolling windows are likely to be more accurate.

# 5 Adding Machine Learning

In the previous section, it was shown that our in-sample-estimated linear and ARMA models perform very similarly in both sub-samples, meaning that neither model has resulted in significant improvements for now. Therefore, based on the methodology of [Belly et al., 2021], we expand our selection of modeling approaches to those with a non-linear structure in order to optimize our country-specific forecasts.

Although Lasso is not a non-linear modeling technique, we present it in comparison to other machine learning techniques because it also requires cross-validation. In fact, the resampling technique ensures that a weighted balance between bias and variance is found by choosing the best parameter value that minimizes the estimator's error rate. The so-called "shrinkage methods" aim to reduce the coefficients of the estimators to 0, which means that unimportant predictors are ultimately introduced into the model with a coefficient of 0 [Chan-Lau, 2017]. While the optimization problem in case of least squares (LS) is defined as

$$\min_{\beta_0,\beta_j} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_j x_{ij})^2 \tag{2}$$

the Lasso regression is extended by the additional hyperparameter $\lambda$ which controls the strength of the regularization penalty and defines the following Lagrangian formulation of the Lasso regression[2].

$$\min_{\beta_0,\beta_j} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_j x_{ij})^2 + \lambda||\beta||_1 \tag{3}$$

.

Our first nonlinear method of interest is Random Forest, a machine learning method that can be used for both classification tasks and regressions [Breiman, 2001]. It is defined as an ensemble learning method, which means that it combines multiple models or multiple decision trees, with each model trained on a different subset of the training data [Xu et al., 2019, p.4]. On the other hand, although setting up a random forest requires a large number of decision trees, it prevents overfitting because it always selects a random subset of input features [Tang et al., 2018, p.4]. In the appendix, we list a table of hyperparameters that we used to optimize our machine learning models. Although the literature suggests that an increasing tree size does not necessarily lead to better performance and accuracy (while the optimal number of trees is often around 100), when we performed our cross-validation we found that the accuracy of our model actually improves up to a certain tree size. For example, we find that a tree size of 150 is optimal for each country model, and 250 is the optimal number if we train our models together for all countries [Tyralis and Papacharalampous, 2017, p.6]. Therefore, given the tree size, we estimate our remaining hyperparameters using a cross-validation, namely mtry (number of variables randomly chosen at each split) and nodesize (minimum number of observations required to form a node) for that. We attach an example decision tree generated in R using the caret and randomforest packages.

Unlike Random Forest (a bagging-type machine learning method), we also rely on boosting models, which differ in that these models are always trained using the same training data. Therefore, the choice of hyperparameters becomes more crucial than in bagging-based models like Random Forest [Chen and Guestrin, 2016, p.786]. XGBoost is based on a simple decision tree, but gradually adds more and more decision trees to optimize the overall performance by taking into account the errors of the previous tree. The final model is therefore a model with a weighted sum of all decision trees [Ramraj et al., 2016, p.652]. Similar to Random Forest, we generate our set of optimal hyperparameters by applying cross-validation functions and optimal grid search using packages caret and XGBoost in R. Although we use all hyperparameters listed in the appendix, we find that the number of boosting rounds, maximum depth of a tree and the shrinkage factor $\eta$ were the most important hyperparameters we use to train our XGB model.

[2]We train our Lasso forecasts by making use of the package glmnet in R.

### 5.1 Full Sample

Similar to the previous exercises, we start integrating the machine learning approaches by retesting our model using samples for the years 2017-2019. However, using our machine learning methods requires that this time we split our data prior to 2017 into training and validation samples, the latter being the same size as our testing sample. As mentioned before, we first train and forecast our models for each country separately before we train our models together in the next chapter. Therefore, this also requires careful tuning of the hyperparameters depending on each country sample combination, which means that if we look at 9 countries across two different samples, we end up with 18 different hyperparameter combinations. This is particularly important for our XGB models. Finally, assuming that our machine learning models do actually differ in accuracy, we further test our machine learning models for their difference in predictive accuracy by performing a Diebold-Mariano test. The reason it is essential to perform this test is that even if for two different forecasts from two different models (A and B) for the same target variable we obtain the result $\widehat{MAE}(X_t^A) < \widehat{MAE}(X_t^B)$ does not necessarily mean that forecast A wins over forecast B [Diebold, 2015]. In fact, it is imperative to ensure that the difference in, for example, the predicted RMSE is significant and not just due to chance. We also list the results of these tests in the appendix. We list the respective test statistics and corresponding p-values of the prediction difference of our error measure (MAE in our case) of the respective machine learning model to one of our baseline models (e.g. ARMA)[3]. Therefore, we interpret the test output as follows: if the p-value does not exceed the chosen significance level (typically 0.05), we reject the null hypothesis, which means that the difference in forecasting accuracy between our respective ML models and our baseline model is indeed statistically significant.

| Metric | IT | ES | FR | PT | GR | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.24 | -0.21 | 0.17 | 0.07 | -0.00 | -0.07 | 0.16 | 0.26 | 0.09 | 0.16 |
| CORR (LASSO) | 0.35 | -0.27 | 0.25 | 0.08 | -0.00 | -0.04 | 0.23 | 0.26 | 0.01 | 0.19 |
| CORR (LM) | 0.23 | -0.21 | 0.18 | 0.07 | -0.00 | -0.07 | 0.16 | 0.24 | 0.09 | 0.16 |
| CORR (RF) | 0.30 | -0.22 | 0.21 | 0.11 | 0.00 | 0.04 | 0.06 | 0.19 | 0.20 | 0.17 |
| CORR (XGB) | 0.24 | -0.35 | 0.24 | 0.08 | 0.00 | 0.02 | 0.06 | 0.23 | 0.24 | 0.18 |
| MAE (ARMA) | 2.17 | 1.56 | 0.54 | 2.38 | 49.85 | 1.26 | 0.66 | 0.47 | 0.75 | 1.22 |
| MAE (LASSO) | 1.74 | 1.20 | 0.47 | 1.56 | 49.92 | 0.56 | 0.40 | 0.40 | 0.56 | 0.86 |
| MAE (LM) | 2.17 | 1.56 | 0.55 | 2.42 | 50.74 | 1.26 | 0.65 | 0.48 | 0.75 | 1.23 |
| MAE (RF) | 1.77 | 1.16 | 0.48 | 1.61 | 31.29 | 0.52 | 0.44 | 0.40 | 0.56 | 0.87 |
| MAE (XGB) | 1.82 | 1.21 | 0.49 | 1.66 | 31.92 | 0.52 | 0.42 | 0.39 | 0.56 | 0.89 |
| RMSE (ARMA) | 3.16 | 2.95 | 0.87 | 3.23 | 636.33 | 1.70 | 0.89 | 0.70 | 1.11 | 1.83 |
| RMSE (LASSO) | 2.80 | 2.61 | 0.83 | 2.45 | 636.78 | 0.91 | 0.67 | 0.66 | 0.97 | 1.49 |
| RMSE (LM) | 3.16 | 2.95 | 0.87 | 3.28 | 638.96 | 1.70 | 0.89 | 0.71 | 1.11 | 1.83 |
| RMSE (RF) | 2.89 | 2.60 | 0.84 | 2.47 | 616.12 | 0.89 | 0.79 | 0.66 | 0.94 | 1.51 |
| RMSE (XGB) | 3.09 | 2.78 | 0.84 | 2.53 | 616.84 | 0.87 | 0.77 | 0.65 | 0.93 | 1.56 |

Table 3: Quality and accuracy metrics, isolated country estimation (Full sample). Values for the columns EZ (Euro Zone) are computed by averaging all countries except for Greece.

[3]Since we are using the MAE as our loss function, we also need to conduct our test with power = 1 since the MAE loss function is untailed.

The results indicate that indeed both non-linear models, namely Random Forest and XG-Boost, improve our forecast accuracy and in most cases appear to have a higher correlation with the actual values of the test sample. By comparing accuracy and quality measures across countries, we cannot conclusively say whether Random Forest or XGB performs better overall. For example, the latter technique seems to predict the change in redenomination risk better than Random Forest for almost all countries, whereas in the case of Italy, Random Forest seems to correlate better with actual values. In contrast, when looking at Random Forest performance, we observe a divide between "Southern" and "Nordic" countries in terms of our accuracy metrics. For the countries of Italy, Spain and Portugal, Random Forest predicts the change in redenomination risk with a RMSE of significantly lower basis points, for example for Italy a RMSE of 2.89 (compared to 3.09 for XGB and 3.16 for our baseline models). Still, it is the Lasso regression that works surprisingly well, and in most cases correlates better with real values than even our nonlinear models. Similar conclusions can be drawn by looking at our accuracy metrics such as RMSE and MAE. We also conclude that extending our approaches with nonlinear models does not improve Greece's forecasting performance across the sample.
At the country level, we still find that the change in redenomination risk seems to be explained even better with our choice of predictors for the countries Italy, Spain, France and the Netherlands. This also applies to the two non-linear models, which allow us to further reinforce the pattern observed at the country level in the previous section. The change in redenomination risk for Austria seems to be better explained by linear models (also including Lasso), while the opposite is true for Belgium. The correlation coefficients of countries like Portugal and Ireland still appear to be significantly low. This is not surprising at this point as the previous section showed that these two countries in particular were at significantly higher levels of redenomination risk during the debt crisis that is part of our training window. Finally, if we consider the Diebold-Mariano test results presented in the appendix, we can further confirm these results, with the test statistic measuring the magnitude and direction of the difference in forecast accuracy relative to the ARMA as the baseline model being the lowest in case of the Lasso regression. By also considering our p-values, we see that our test rejects the hypothesis of equal predictive accuracy for all of our listed models, and instead applies the alternative hypothesis that the two models actually have different predictive accuracies.

If we look at the variable importance of our XGBoost and Random Forest models in the appendix, we can already see significant differences between countries. Unlike measures such as $R^2$, the XGB variable importance does not measure how much of the variance a certain variable explains but rather how much a particular variable reduces the loss function on average over all trees. Similarly, RF variable importance measures the decrease in model performance when a feature or predictor is removed. Therefore, by comparing the variable importance measures from both nonlinear models across our sample, we can already conclude that two variables always rank highest in all countries. Both the lags of our target variable (difference of CDS spreads to Germany) and lags of the USD-denominated CDS spreads reduce the loss function on average more than other predictors, for example by 0.2 in the case of the Netherlands. In addition, lags in other country-specific variables such as government bond spreads or the returns of the respective stock index appear to

be of further importance. Occasionally, non-country variables such as US volatility from a GARCH(1,1), eurozone break-even inflation or the US Investment Grade are included in the top 15 most important variables. It is noticeable that the distribution and magnitude of the most important variables vary from country to country. For example, Greece, the Netherlands and Ireland are the countries with the largest absolute gains in importance, with an average reduction in the loss function of around 0.25 when the USD-denominated CDS spread lag is implemented in the Greece-specific model. Ireland, for example, is the only country where the lag of government bond spreads appears to be key to reducing the loss function. However, for the rest of the countries, the distribution and magnitude of the importance of the XGB and RF variables appear to be relatively similar, with each country's lags in redenomination risk, USD-denominated CDS spreads, government bond spreads and equity returns driving the model performance the most.
Finally, we also consider the importance measures of our linear machine learning model. As mentioned earlier, the lasso regression penalizes the relatively low-magnitude coefficients with 0, so only higher-magnitude coefficients are included in the model. In fact, due to the linear nature of the model, it tends to consider variables as important that appear to have a strong linear relationship with the dependent variable [Fonti and Belitser, 2017, p.4]. This reasoning is consistent with the observations we make by considering the Lasso variable importance in different countries, which differ markedly from the nonlinear ones. First, the penalization of the Lasso coefficients is so severe for most countries that fewer than 15 variables are selected as important overall. Second, some of the variables that were classified as very important during by RF and XGB (e.g. the lag in redenomination risk and the lag in USD denominated CDS spreads) seem not even to be listed in any of the countries Lasso variable importance. Although it is difficult to obtain a relative measure of variable importance for our Lasso regression, making comparison between countries more difficult, it can still be concluded that countries with high volatility in redenomination risk (e.g. Greece, Spain or Italy) were less affected by the coefficient shrinking. Only the lags in government bond spreads are considered as relevant by both linear and non-linear ML models.

### 5.2 Debt Crisis

As our number of observations before the European debt crisis could be considered too small, we change our approach and implement forecasts based on a rolling window. This is particularly critical for our nonlinear machine learning models, since training these models with almost 100 features requires a sufficient amount of observations that could not otherwise be guaranteed. Therefore, forecasting redenomination risk a day ahead means that the training window (starting December 15, 2007) is shifted by one day or one row in our data frame to the last day in our window (May 1, 2012). According to the [Belly et al., 2021] methodology, the validation set equals therefore always the last observation of our training set. This approach also allows us to adjust for time-dependence as the validation window selects those parameters that minimize our accuracy metric of interest the most. Looking at the results below, we see that extending our analysis with machine learning models for a sample limited to the debt crisis optimizes the 1-step prediction forecasts for most countries.

| Metric | IT | ES | FR | PT | GR | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.05 | 0.06 | 0.14 | 0.04 | 0.06 | 0.11 | 0.03 | 0.09 | 0.08 | 0.08 |
| CORR (LASSO) | 0.28 | 0.15 | 0.28 | 0.23 | 0.08 | 0.40 | 0.43 | 0.61 | 0.14 | 0.31 |
| CORR (LM) | 0.30 | 0.16 | 0.31 | 0.12 | 0.12 | 0.21 | 0.16 | 0.34 | 0.12 | 0.22 |
| CORR (RF) | 0.20 | 0.18 | 0.02 | 0.13 | 0.09 | 0.23 | -0.01 | 0.34 | 0.16 | 0.16 |
| CORR (XGB) | 0.20 | 0.12 | 0.07 | 0.11 | 0.19 | 0.34 | -0.04 | 0.34 | 0.23 | 0.18 |
| MAE (ARMA) | 10.96 | 10.06 | 4.26 | 26.73 | 245.27 | 14.28 | 4.38 | 5.16 | 6.46 | 10.28 |
| MAE (LASSO) | 10.85 | 10.12 | 4.18 | 26.51 | 241.46 | 13.79 | 4.13 | 4.64 | 6.50 | 10.09 |
| MAE (LM) | 11.00 | 10.33 | 4.11 | 30.77 | 362.87 | 14.58 | 5.06 | 4.96 | 6.83 | 10.95 |
| MAE (RF) | 11.03 | 9.93 | 4.44 | 28.74 | 260.88 | 14.01 | 4.49 | 5.18 | 6.57 | 10.55 |
| MAE (XGB) | 10.99 | 10.05 | 4.32 | 27.42 | 242.54 | 13.83 | 4.48 | 4.99 | 6.41 | 10.31 |
| RMSE (ARMA) | 15.91 | 13.91 | 6.06 | 42.57 | 546.33 | 23.49 | 7.65 | 10.30 | 9.50 | 16.17 |
| RMSE (LASSO) | 15.48 | 13.72 | 5.98 | 41.15 | 532.39 | 22.31 | 7.08 | 8.10 | 9.34 | 15.39 |
| RMSE (LM) | 15.23 | 13.87 | 5.83 | 45.23 | 830.81 | 22.90 | 7.79 | 9.64 | 9.54 | 16.25 |
| RMSE (RF) | 15.70 | 13.64 | 6.32 | 43.19 | 543.26 | 22.74 | 7.73 | 9.48 | 9.33 | 16.02 |
| RMSE (XGB) | 15.60 | 13.77 | 6.15 | 42.22 | 523.92 | 22.27 | 7.78 | 9.48 | 9.18 | 15.81 |

Table 4: Quality and accuracy metrics, isolated country estimation (Debt Crisis). Values for the columns EZ (Euro Zone) are computed by averaging all countries except for Greece.

In fact, considering our accuracy metrics, we find that the linear model (which selects the optimal predictors based on AIC criteria) performs better for some countries. For Italy or France, our linear model appears to be more accurate than ARMA and our machine learning models appear to have lower RMSE values and correlations with the real values greater than 0.3. Note that our accuracy metrics tend to increase as we re-constrain the sample to the debt crisis. For Italy, for example, only the Lasso regression comes reasonably close in terms of quality and accuracy. In fact, we again find that the Lasso regression performs better than nonlinear modeling techniques for most countries, for example with a correlation with the real values of 0.43 in case of Ireland or even 0.6 in case of the Netherlands. Overall, when considering accuracy and quality metrics, we actually observe a head-to-head race between the Lasso model (which has the highest overall correlation coefficients) and the XGB model (which has the lowest RMSEs and MAEs). In fact, compared to the full sample analysis, we observe that the nonlinear models lose predictive power slightly in favor of our linear machine learning model, Lasso regression. Notably, for the countries Greece and Portugal, which were hardest hit during the debt crisis, the difference in RMSEs between the linear model and the machine learning models is more significant, as ML models have a RMSE that is 3 basis points lower for Portugal than the linear model.

By looking again at the variable importance plots from our XGB and RF models, we gain further insight into the different patterns during the European debt crisis. In general, we find that other variables such as national equity indices or volatility premiums appear less frequently in our country-specific charts. As a result, we find that these lists now consist primarily of two variables, namely the lag in redenomination risk itself and the government bond spread. In other words, the lagging variables of government bond spreads appear to be more important in reducing the cross-country loss function during the European debt crisis than across the sample as a whole.
Similar to the previous exercise, when performing the out-of-sample prediction, the Lasso model also identifies a similar set of variables as important when performing the rolling

time series exercise. In fact, the lag of the USD/EUR exchange rate, the respective country's government bond spread and the OIS are reduced the least by the Lasso coefficients. This is another revealing conclusion that can be drawn. Lasso regression actually considers different types of predictors as relevant compared to the nonlinear modeling approaches. While our non-linear models mainly assigns the greatest importance to the total lag distribution of redenomination risk itself, USD-denominated CDS spreads and government bond spreads, the Lasso regression considers predictors on the other hand that have a stronger macroeconomic focus. This can be explained by the fact that the Lasso model identifies variables as important that per se have a very direct relationship to the target variable [Hastie et al., 2009, p.68]. For example, by construction, it would prefer variables that have a strong individual impact on the change in redenomination risk. European-level variables such as OIS or break-even inflation show a strong direct correlation with the change in redenomination risk across multiple countries, increasing the chance of being picked by the Lasso model. Finally, if we compare the importance of the Lasso variables during the debt crisis, we observe an increasing importance of the EUR/USD exchange rate at the expense of breakeven inflation, which was one of the main predictors of the Lasso model in relation on the full sample. The fact that the EUR/USD exchange rate is gaining importance when narrowing the testing window to the debt crisis is also consistent with the literature pointing to the connection between foreign currency events and sovereign creditworthiness over time during the European debt crisis [Alsakka and ap Gwilym, 2012].

Most importantly, we conclude that at least one of our machine learning models outperforms our baseline models for all countries. In particular, our Lasso model performs very well in terms of quality measures, realizing correlation values of around 0.5 to 0.6 for some countries. So, also accounting for the increase in the correlation coefficients of our nonlinear models, restricting our sample to the debt crisis not only appears to be a factor in our improved quality measures, but the fact that we perform forecasts based on a rolling window appears to be of great importance too. Since the debt crisis is the most volatile subsample of our entire training data, shifting the training window after each observation allows our ML models to better capture the underlying trends in redenomination risk during the debt crisis. However, we also see that when we switch from the full sample to the sample restricted to the debt crisis, our nonlinear machine learning models lose accuracy compared to our cross-country lasso regression. This can have several reasons:
1) The determination of the redenomination risk, as measured by CDS spreads, is not only influenced by corresponding explanatory variables in the respective country, but is rather determined by the interaction between the countries (since all countries are members of a monetary union) and is also influenced by spillover effects from other countries. Therefore, we could solve this problem by creating a panel data set where we train all countries together.
2) On the other hand, other explanatory factors could also be the subject of political economy. For example, the level of redenomination risk during the European debt crisis was indeed strongly influenced by factors related to the fiscal environment. It is plausible that these factors were largely not captured by our financial market variables obtained from the [De Santis, 2019] approach.

# 6 Extending Machine Learning

Based on our results from the previous section and the resulting conclusions, we extend our training data in the sense that we now train our machine learning models together, including observations from all countries. As highlighted in the previous section, the idea is that by training our models together across countries, we would be able to capture cross-country dependencies that would otherwise be hidden. For example, spillover effects from Greece during the European debt crisis could be of great importance in determining the change in redenomination risk in other European countries. A higher learning rate of ML models, realized due to the interactions within the countries, could further optimize our forecast for our country of interest. Therefore, for the following two exercises, we also include a selection of dummy factor variables in our panel training set. By limiting the number of observations of our test data to our country of interest, but also expanding the amount of variables to include the same factor dummy variables, our nonlinear models should be able to identify cross-country dependencies to further optimize the predictive power of our machine learning models.

## 6.1 Full sample

By comparing our combined training analysis across the entire sample with the country-specific analysis from Section 5.1, we can already see several differences. The results suggest that training our models together for all countries does indeed result in forecast improvements for some of these countries, most notably our Southern European economies.

| Metric | IT | ES | FR | PT | IE | AT | NL | BG | EZ |
|---:|---|---:|---:|---:|---:|---:|---:|---:|---:|
| CORR (ARMA) | 0.26 | -0.17 | 0.18 | 0.08 | -0.08 | 0.15 | 0.26 | 0.10 | 0.17 |
| CORR (LASSO) | 0.33 | -0.25 | 0.04 | 0.10 | -0.02 | -0.04 | -0.08 | -0.00 | 0.12 |
| CORR (LM) | 0.32 | -0.17 | 0.01 | 0.09 | -0.03 | -0.00 | -0.04 | 0.01 | 0.09 |
| CORR (RF) | 0.28 | -0.22 | 0.03 | 0.07 | 0.03 | 0.07 | 0.05 | 0.08 | 0.11 |
| CORR (XGB) | 0.38 | -0.33 | 0.19 | 0.07 | -0.08 | -0.03 | 0.03 | 0.08 | 0.16 |
| MAE (ARMA) | 1.97 | 1.47 | 0.53 | 2.27 | 1.25 | 0.58 | 0.47 | 0.75 | 1.15 |
| MAE (LASSO) | 1.82 | 1.27 | 0.62 | 1.64 | 0.66 | 0.58 | 0.57 | 0.70 | 1.03 |
| MAE (LM) | 1.90 | 1.38 | 0.78 | 1.72 | 0.82 | 0.71 | 0.73 | 0.86 | 1.16 |
| MAE (RF) | 1.74 | 1.17 | 0.51 | 1.52 | 0.54 | 0.42 | 0.42 | 0.58 | 0.91 |
| MAE (XGB) | 1.73 | 1.14 | 0.45 | 1.49 | 0.48 | 0.39 | 0.37 | 0.54 | 0.87 |
| RMSE (ARMA) | 3.04 | 2.89 | 0.87 | 3.18 | 1.77 | 0.81 | 0.71 | 1.09 | 1.80 |
| RMSE (LASSO) | 2.85 | 2.71 | 0.97 | 2.51 | 0.99 | 0.85 | 0.83 | 1.06 | 1.68 |
| RMSE (LM) | 2.92 | 2.78 | 1.16 | 2.59 | 1.18 | 1.02 | 1.04 | 1.22 | 1.82 |
| RMSE (RF) | 2.87 | 2.57 | 0.88 | 2.43 | 0.88 | 0.70 | 0.69 | 0.96 | 1.58 |
| RMSE (XGB) | 2.77 | 2.62 | 0.84 | 2.45 | 0.86 | 0.70 | 0.67 | 0.95 | 1.57 |

Table 5: Quality and accuracy metrics, joint country estimation (Full sample). Greece not being included in the training sample.

In fact, by excluding Greece from the training sample (as suggested by the poor performance in the individual country analysis in Section 5.1), we actually observe a reduction in the RMSE in the change in redenomination risk from 3.09 to 2.77 basis points for Italy and similar reductions of 2.78 to 2.62 basis points for Spain and 2.53 to 2.45 basis points for Portugal. However, the remaining countries do not appear to be improving in either our quality or accuracy metrics, at least not across the full sample. Overall, we also reaffirm our results from Section 5.1. that our machine learning models still tend to make better predictions than our baseline models over the entire sample, even when we train our models together for all countries. However, since joint training of our models leads to predictive improvements for only three Southern European economies (Italy, Spain and Portugal), but hardly improves the forecast accuracy of the remaining countries, we consider further splitting our out of sample analysis into Southern and northern European countries. We hereby include France in our sample of “Southern” countries, as our literature review has clearly shown that the pattern of redenomination risk appears to be more similar to that of “northern” countries.

| Metric | IT | ES | FR | PT | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.32 | -0.10 | 0.18 | 0.07 | -0.05 | 0.07 | 0.07 | 0.06 | 0.12 |
| CORR (LASSO) | 0.00 | -0.55 | -0.17 | 0.04 | -0.02 | -0.03 | -0.20 | 0.07 | 0.14 |
| CORR (LM) | 0.02 | -0.12 | -0.03 | 0.01 | -0.07 | -0.04 | -0.06 | -0.01 | 0.04 |
| CORR (RF) | 0.17 | -0.40 | 0.02 | 0.06 | 0.08 | 0.16 | 0.00 | 0.19 | 0.14 |
| CORR (XGBOOST) | 0.22 | -0.34 | 0.02 | 0.12 | 0.04 | 0.11 | -0.03 | 0.17 | 0.13 |
| MAE (ARMA) | 1.88 | 1.41 | 0.52 | 2.14 | 0.97 | 0.53 | 0.48 | 0.73 | 1.08 |
| MAE (LASSO) | 1.76 | 1.17 | 0.47 | 1.51 | 0.56 | 0.46 | 0.44 | 0.59 | 0.87 |
| MAE (LM) | 2.32 | 1.90 | 1.37 | 2.12 | 1.03 | 0.89 | 0.94 | 1.03 | 1.45 |
| MAE (RF) | 1.81 | 1.27 | 0.54 | 1.56 | 0.51 | 0.40 | 0.37 | 0.55 | 0.88 |
| MAE (XGBOOST) | 1.81 | 1.19 | 0.54 | 1.54 | 0.51 | 0.40 | 0.36 | 0.56 | 0.86 |
| RMSE (ARMA) | 2.94 | 2.77 | 0.87 | 3.05 | 1.44 | 0.77 | 0.74 | 1.07 | 1.71 |
| RMSE (LASSO) | 3.03 | 2.71 | 0.89 | 2.44 | 0.91 | 0.74 | 0.66 | 0.97 | 1.54 |
| RMSE (LM) | 3.67 | 3.44 | 2.21 | 3.21 | 1.41 | 1.24 | 1.28 | 1.42 | 2.23 |
| RMSE (RF) | 3.09 | 2.77 | 0.90 | 2.50 | 0.86 | 0.68 | 0.58 | 0.94 | 1.54 |
| RMSE (XGBOOST) | 3.05 | 2.72 | 0.90 | 2.45 | 0.86 | 0.69 | 0.55 | 0.95 | 1.52 |

Table 6: Quality and accuracy metrics, joint country estimation (Full sample). Seperate training samples for Southern and Nordic countries. Greece not being included in the training sample.

The results suggest that by dividing our training model into northern and Southern European countries, we slightly improve our accuracy measures for our “northern” economies, but not for the “Southern”. In fact, Italy in particular is exposed to a sharp increase in its RMSE, which is reaching similar values as the individual country estimate. Spain and Italy share similar characteristics except for one model. By dividing our joint training exercise into Nordic and Southern European countries, Lasso regression predicts the change in redenomination risk with a similar magnitude in RMSE but with a jump in the correlation coefficient to 0.55. On the other hand, while the forecast accuracy for the Nordic countries improves slightly, the correlation coefficients do not improve. This confirms our original assumption: cross-country dependencies appear to be important for redenomination risk for Southern economies but not for Nordic ones.

By additionally considering the variable importance of both sub-samples, we obtain further revealing results. Not only do we confirm the results from the previous section, but we also find that for "Southern" countries, the redenomination risk was more dependent on the lag of the redenomination risk itself, while the lags in government bond spreads for the Nordic countries were more important. Intuitively, this also makes sense as Southern countries are generally more exposed to fiscal risk, hence the lag in the redenomination risk itself gains prognostic importance. On the other hand, since the redenomination risk itself is a component of a country's sovereign spread, it determines the change in the total sovereign spread and hence the change in the redenomination risk[45]. In other words, the government bond spread itself affects redenomination risk for Nordic and Southern countries to a similar extent, whereas the lag in redenomination risk appears to be a more important predictor for Southern European countries than vice versa [Corradin et al., 2021].
Finally, considering the Diebold-Mariano tests, we find that most of our machine learning methods indeed yield statistically significant results for most country-model combinations. Our test results show that our machine learning models significantly outperform our baseline models for all countries except Italy and France.

## 6.2 Debt Crisis

Because we are now using a panel training dataset, meaning that we are adding eight additional country-specific factor dummy variables to our training dataset, concerns about data limitations are less of an issue. For this reason, and in contrast to the forecasting exercise in section 5.2, we now again base our forecasts on an out-of-sample analysis.

| Metric | IT | ES | FR | PT | GR | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.29 | 0.10 | 0.28 | 0.20 | 0.18 | 0.27 | 0.21 | 0.34 | -0.03 | 0.21 |
| CORR (LASSO) | 0.20 | 0.14 | 0.12 | 0.15 | 0.13 | 0.29 | 0.08 | -0.03 | 0.22 | 0.15 |
| CORR (LM) | 0.23 | 0.09 | 0.12 | 0.14 | 0.16 | 0.32 | 0.14 | 0.11 | 0.14 | 0.16 |
| CORR (RF) | 0.23 | 0.21 | 0.10 | 0.15 | -0.16 | 0.31 | -0.03 | -0.25 | 0.22 | 0.19 |
| CORR (XGB) | 0.23 | 0.18 | 0.17 | 0.10 | -0.13 | 0.20 | -0.01 | -0.27 | 0.17 | 0.17 |
| MAE (ARMA) | 11.00 | 10.56 | 4.16 | 27.48 | 292.77 | 14.81 | 4.57 | 4.95 | 5.09 | 10.33 |
| MAE (LASSO) | 10.92 | 10.06 | 4.30 | 26.48 | 239.77 | 14.05 | 4.31 | 5.09 | 6.45 | 10.21 |
| MAE (LM) | 11.19 | 10.49 | 4.84 | 27.27 | 261.24 | 14.16 | 4.92 | 5.76 | 6.76 | 10.67 |
| MAE (RF) | 10.65 | 9.95 | 4.40 | 26.34 | 232.05 | 13.72 | 4.38 | 5.27 | 6.36 | 10.13 |
| MAE (XGB) | 11.48 | 10.91 | 4.63 | 28.72 | 245.27 | 15.06 | 5.03 | 5.69 | 7.06 | 11.07 |
| RMSE (ARMA) | 15.30 | 14.21 | 5.89 | 41.56 | 606.41 | 22.50 | 7.48 | 9.62 | 10.11 | 15.83 |
| RMSE (LASSO) | 15.60 | 13.72 | 6.09 | 41.72 | 529.49 | 22.51 | 7.56 | 10.11 | 9.20 | 15.81 |
| RMSE (LM) | 15.56 | 14.15 | 6.59 | 41.93 | 541.73 | 22.17 | 7.74 | 10.18 | 9.50 | 15.98 |
| RMSE (RF) | 15.47 | 13.54 | 6.16 | 41.75 | 536.37 | 22.42 | 7.74 | 10.41 | 9.20 | 15.84 |
| RMSE (XGB) | 16.45 | 14.46 | 6.45 | 43.72 | 543.58 | 24.52 | 8.38 | 10.96 | 9.72 | 16.83 |

Table 7: Quality and accuracy metrics, joint country estimation (Debt Crisis). Greece not being included in the training sample.

The results suggest that when the sample is restricted to the debt crisis, the linear model predicts the change in redenomination risk more accurately and with higher correlation

[4][Montes et al., 2022] shows the relationship between budget sentiment and quarterly CDS spreads

[5][Wisniewski and Lambe, 2015] shows the interaction between economic policy uncertainty and the cost of credit protection for the US and the Eurozone.

to real values compared to the combined country exercise over the full sample. However, the best-performing models over this period again seem to be the Lasso regression and the Random Forest model, and most surprisingly the ARMA model (which includes the AIC-based optimal external regressors). By averaging our accuracy and quality metrics across all eurozone countries (except Greece), we find Random Forest and Lasso to predict the lowest RMSE in the change in redenomination risk. On the other hand, our ARMA model and again Random Forest show the largest correlation coefficients with the real values. Comparing the out-of-sample forecasts in our joint analysis during the debt crisis in different countries, we find that the forecasts are improving for most countries, with the exception of Spain and Italy. This already leads us to an interesting observation. Both countries, Spain and Italy, whose levels of redenomination risk have varied greatly over time, can be predicted more accurately over the full sample than over the debt crisis, as demonstrated in the previous chapter when obtaining correlation coefficients of 0.43 and 0.55 for Italy and Spain using Lasso regression.

Finally, we reconfirm our results by considering the results of the Diebold Mariano test listed in the appendix. The results clearly show a significant difference in the prediction accuracy of the Lasso model in relation to most countries. Furthermore, we find that our machine learning models outperform our baseline ARMA model in terms of prediction accuracy in all countries except Italy.

## 6.3 Optimal h-step ahead forecasts

To link to practice, the final part of our empirical analysis is to identify and highlight the optimal one-step-ahead forecast for each country-model combination for both sub-samples. In fact, we believe this is insightful information for investors and policymakers, as this analysis allows us to provide a descriptive view of the different patterns between a full sample and a structural crisis. With this in mind, we compare the predictive performance of the ARMA (our best performing baseline model) to the optimal machine learning model (linear or non-linear), summarized in the respective ML column. In addition, we include information as to whether our single country analysis produced the optimal forecast or whether it was necessary to train different countries together. To ensure describability, we also visualize the optimal one-step-ahead forecasts in the appendix by plotting the predicted values with the actual values for each country in both sub-samples.

| Metric | IT | ES | FR | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.24 | -0.21 | 0.17 | 0.16 | 0.26 | 0.09 | 0.19 |
| CORR (ML) | 0.43 | 0.55 | 0.24 | 0.23 | 0.26 | 0.24 | 0.33 |
| MAE (ARMA) | 2.17 | 1.56 | 0.54 | 0.66 | 0.47 | 0.75 | 1.02 |
| MAE (ML) | 1.80 | 1.14 | 0.49 | 0.41 | 0.40 | 0.56 | 0.80 |
| RMSE (ARMA) | 3.16 | 2.95 | 0.87 | 0.89 | 0.70 | 1.11 | 1.61 |
| RMSE (ML) | 2.80 | 2.63 | 0.84 | 0.67 | 0.66 | 0.93 | 1.42 |
| Model | XGB | XGB | Lasso | Lasso | Lasso | XGB | |
| Sample | JS | JS | Single | Single | Single | Single | |

Table 8: Optimal H-step ahead Forecasts (Full sample).JS = Joint training sample consisting only of South European economies.

The results clearly show that none of the machine learning models is absolutely superior in terms of both accuracy and quality of predictions. As shown in the previous sections, we observe a head-to-head race between XGBoost and Lasso when we restrict the analysis to the entire sample. In addition, we exclude Portugal and Ireland from the table because we cannot significantly improve the predictive ability of these two countries through machine learning. Indeed, considering the EZ column, which in turn represents the average of all included countries, we observe significant improvements in terms of accuracy (drop in RMSE from 1.61 to 1.42 bps) and quality (increase in correlation with the real data of 0.19 to 0.33) compared to the baseline. Furthermore, we find that the predictive improvements are particularly strong for Italy and Spain, which are the only countries that provide optimal forecasts through a joint cross-country training analysis, while the improvements for the rest of the countries are rather moderate. Our analyses to date have shown that simply by restricting the period of interest to the debt crisis, we were able to improve our forecasts for a broader range of countries.

| Metric | IT | ES | FR | PT | IE | AT | NL | BG | EZ |
|---|---|---|---|---|---|---|---|---|---|
| CORR (ARMA) | 0.05 | 0.06 | 0.14 | 0.04 | 0.11 | 0.03 | 0.09 | 0.08 | 0.08 |
| CORR (ML) | 0.35 | 0.28 | 0.39 | 0.32 | 0.45 | 0.56 | 0.62 | 0.23 | 0.40 |
| MAE (ARMA) | 10.96 | 10.06 | 4.26 | 26.73 | 14.28 | 4.38 | 5.16 | 6.46 | 10.28 |
| MAE (ML) | 10.64 | 9.82 | 4.05 | 26.12 | 13.34 | 4.04 | 4.64 | 6.41 | 9.88 |
| RMSE (ARMA) | 15.91 | 13.91 | 6.06 | 42.57 | 23.49 | 7.65 | 10.30 | 9.50 | 16.17 |
| RMSE (ML) | 15.05 | 13.39 | 5.74 | 40.25 | 21.18 | 6.42 | 8.02 | 9.18 | 14.90 |
| Model | Lasso | Lasso | Lasso | Lasso | Lasso | Lasso | Lasso | XGB | |
| Sample | Single | Single | Single | Single | Single | Single | Single | Single | |

Table 9: Optimal H-step ahead Forecasts (Debt Crisis).

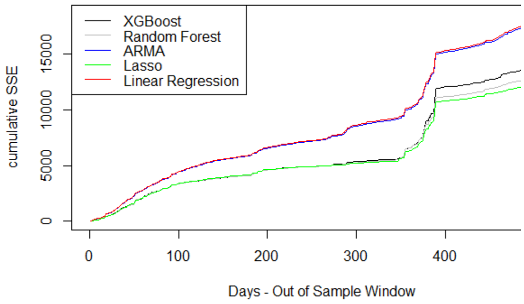
Indeed, by comparing the optimal ML forecasts to the ARMA model during the debt crisis, we find a broader increase in forecasting accuracy and quality across countries. Looking at the EZ column, we observe an increase in the correlation with the real values from 0.08 to 0.4 when we rely on the optimal ML of our choice. We believe this is indeed a strong rise as, as mentioned, not only is it sustained across countries, but is further confirmed by our decline in MAE and RMSE. Furthermore, we believe that the Lasso model is significantly superior to the other machine learning models because it was selected as the optimal forecasting model for all countries except Belgium. Finally, we also reiterate our findings from Section 6.2, where we believe that joint training exercises are unlikely to be of much use during the debt crisis. In fact, all optimal one-step-ahead forecasts are based on an

estimate for a single country.
Therefore, we conclude that the pattern of our optimal H-step-ahead forecasts is indeed different when differentiating between the two sub-samples, the overall sample and the debt crisis. Across the sample, predictive improvements are more limited and model superiority less clear. In contrast, the addition of machine learning models to our analysis in times of structural crises leads to huge forecast improvements and shows a clear superiority of the model over Lasso regression.

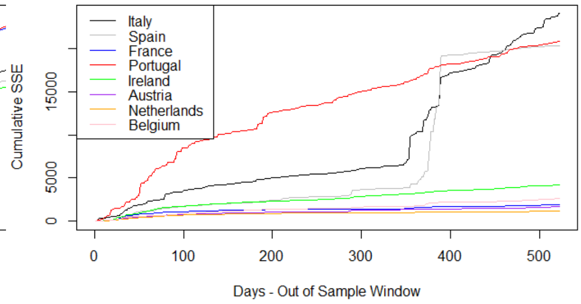
## 7 Discussion

Our introductory exercise aimed at finding our baseline model allowed us to identify important differences in terms of the forecasting applicability of [De Santis, 2019] predictors. We conclude that traditional forecasting models (linear regression and ARMA) have limited predictability for countries with high volatility (Portugal and Ireland) over the full sample. Restricting our test sample to the European debt crisis increases, on average, the overall correlation of forecasts with actual values for all eurozone countries, while our accuracy metrics tend to spike over this period due to an overall higher risk of redenomination.

Both nonlinear models (Random Forest and XGB) improve our forecast accuracy and correlation coefficients across the entire sample. In addition, we observe different performances between "Southern" and "Northern" countries. For the countries Italy, Spain, Portugal and Greece, Random Forest predicts the change in redenomination risk with a significantly lower RMSE, backed by the Diebold Mariano tests for forecast accuracy. Still, it is the Lasso regression that works surprisingly well and, in most cases, correlates better with the real values than our nonlinear models. At the country level, we find that the change in redenomination risk still seems to be better explained with our choice of predictors for the countries Italy, Spain, France and the Netherlands. Lags in redenomination risk itself, as well as lags in USD-denominated CDS spreads, on average reduce the loss function more than other predictors and are therefore more important in predicting the change in redenomination risk. In contrast, macroeconomic factors such as lags in breakeven inflation and overnight index swaps (OIS) appear to be the most important predictors of redenomination risk using Lasso. Countries with high volatility of redenomination risk over the full sample (e.g. Greece, Spain or Italy) were less affected by the coefficient shrinking by the Lasso model compared to the other countries.




(a) Across models



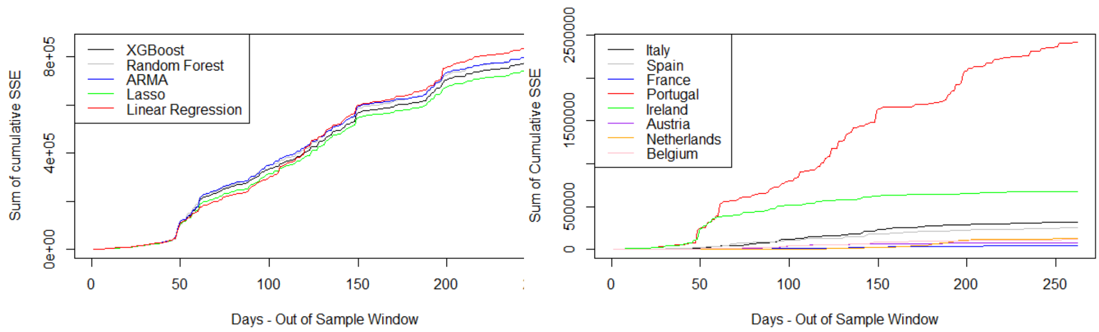

(b) Across countries

Figure 3: Cumulative SSEs across models and countries, single country estimation (Full Sample)

To summarize our results and also to illustrate them in a time perspective, we introduce the cumulative sum of squared errors (SSE) across models and countries. We therefore simply sum the squared errors of all countries (models) at each point in time to visualize the performance of the models (countries). Figure 3a) reaffirms our results by showing that the machine learning models indeed outperform the traditional forecasting models across the sample. Additionally, looking at the same chart, we notice that all models show a sharp increase in their accumulated SSE around trading day 320. Considering Figure 3b), where we sum across countries (rather than models), we find that this sharp increase can be mainly attributed to the trend of two countries, Spain and Italy, where the cumulative SSE of the latter country rose first. Interestingly, we observe that the cumulative SSE for Italy increases around the same time as the tense political situation after the 2018 elections took place. The failure of the March 2018 elections to produce a clear winner resulted in a hung-parliament and a three-month period of political uncertainty before the Conte government was formed in June 2018 [Chiaramonte et al., 2018, p.496]. This is consistent with the literature we referred to in the first few chapters, where political uncertainties during the French and Italian elections led to jumps in CDS spreads. Furthermore, with regard to the jump for the country of Spain, we argue that in our analyses in this paper, there continues to be a close connection between Italian and Spanish CDS spreads. For example, in Section 6.1 we show that by creating a joint training sample of Southern European countries, the correlation coefficients of the Lasso model for Spain increase to 0.55, indicating a huge increase in predictability for Spain once we add Italy and Portugal to the model. Therefore, we conclude that machine learning models significantly outperform our traditional forecasting models in terms of forecast accuracy over the full sample. Furthermore, we note that the increase in cumulative SSEs in all models is due to political tensions after the 2018 elections in Italy and that the close connection of Spanish CDS spreads also appears to be driving up the respective cumulative SSEs for Spain. So, if we disregard the increase in RMSEs for Italy and France, we see that Portugal is the only country to have a high cumulative SSE from the start of our out-of-sample window.

By applying the same analysis to the sample consisting of the European debt crisis, we find that the linear model increases in predictive accuracy and quality for some countries. In fact, we again find that Lasso regression performs better than nonlinear modeling techniques for most countries, for example with a correlation with real values of 0.43 in the case of Ireland or even 0.6 in the case of the Netherlands. The nonlinear models just mentioned lose predictive power slightly in favor of our linear machine learning models. Non-linear ML models still tend to work better for Portugal and Ireland, considering the debt crisis as a test window. The lag in redenomination risk itself as well as the government bond spread seem to be the main drivers in relation to our non-linear models. USD-denominated CDS spreads, on the other hand, are declining, which is understandable given the increasing interdependence between the eurozone and the US during this period.

In contrast to the full sample, the Lasso variable importance now assigns the greatest relevance to the USD/EUR exchange rate (at the expense of break-even inflation). This result is not surprising since the literature shows the connection between measuring fiscal and sovereign risk (in this case sovereign creditworthiness) and exchange rate events, including the European debt crisis [Alsakka and ap Gwilym, 2012]. This strong correlation with the exchange rate therefore applies not only to CDS spreads in general, but logically speaking also to redenomination risk. Although using the other measure of redenomination risk, namely quanto CDS spreads ($CDS^{IT}_{USD} - CDS^{IT}_{EUR}$), Nielsen and Lando show the persistence of such a strong relation across four euro zone sovereigns where they find that when comparing bond yields across currencies using standard FX forward hedges, an important component of the quanto effect is overlooked [Lando and Bang Nielsen, 2018].




(a) Across models (b) Across countries

Figure 4: Cumulative SSEs across models and countries, single country estimation (Debt Crisis)

The superior performance of our machine learning models decreases as soon as we limit the test sample to the debt crisis. In fact, in the first quarter of the year, we even observe that our linear model slightly outperforms our ML models. However, once the ML models appear to be trained on the debt crisis data (since we performed this exercise based on a rolling window), we eventually observe a significant decrease in cumulative SSEs compared to our baseline models at the end of the test window. Similar to the accumulated SSEs in the different countries throughout the sample, we find that Portugal accounts for the largest proportion of all countries, while SSEs for Ireland were accumulated mainly in the

first quarter of the debt crisis.
Finally, the results of the H-step-ahead forecasts suggest that the pattern appears to be different for both subsamples. Over the full sample, predictive improvements are more limited and model superiority less clear. In contrast, adding machine learning models to our analysis in a period of structural crisis with Lasso as the superior modeling approach results in huge forecast improvements.

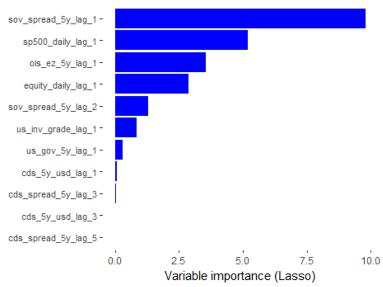
A final question that needs to be answered is the rationale for the superiority of the different models. Overall, we find that XGB and Lasso significantly outperform both our baseline models and Random Forest. Furthermore, we show that the Lasso model is clearly superior to all other models in the case of a debt crisis, especially in terms of correlation coefficients, raising the question of how a linear technique can outperform two very different modeling techniques (XGB and Random Forest), who are able to capture complex non-linear relationships and interactions? We argue that because Lasso is a linear model, it places more emphasis on macroeconomic variables due to their direct impact on redenomination risk. Previously, we showed that, depending on the period of interest, the nonlinear models mainly consider a combination of lags in redenomination risk itself, USD-denominated CDS spreads and government bond spreads as being particularly relevant, while the Lasso regression considers predictors such as the USD/EUR exchange rate, OIS or breakeven inflation as important. We substantiate these results with insights from other authors, such as the close relationship between redenomination risk and exchange rates. If we also take into account the number of predictors actually considered important, we find a major difference between Lasso regression and the two nonlinear approaches. In fact, by reducing less important features to zero, Lasso's automatic feature selection process excludes many more predictors from the model than vice versa. It therefore seems that this mechanism not only filters out noisy variables, but also prevents overfitting of the models, which is not the case with the nonlinear models.

# 8 Conclusion

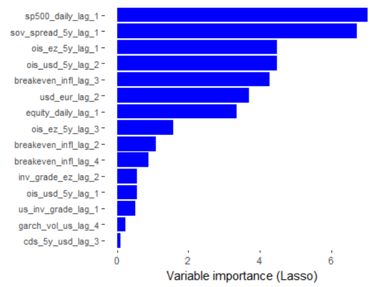
This thesis is the first ever to construct a general model that predicts the risk of a eurozone country devalueing into its national currency (redenomination risk) and makes valuable contributions to the literature. We find that traditional forecasting models struggle to predict eurozone redenomination risk. Machine learning models, especially Lasso and XG-Boost, actually improve prediction quality and accuracy across countries. By subdividing our data into a sample covering a structural crisis (European debt crisis), we find that our linear machine learning model, Lasso regression, predicts the redenomination risk for resilient countries with a correlation of 0.5 (Austria) to 0.6 (Netherlands) between predicted and real values. Furthermore, by further switching the methodology from a single estimate to a joint country estimate, we can even shift the correlation coefficients of highly volatile countries across the entire sample, for example to 0.44 (Italy) or 0.55 (Spain) if we restrict the training example to Southern European countries. Finally, considering the variable importance between models, we find that Lasso regression is superior to nonlinear machine learning models overall and is related to how Lasso assigns relevance or not to predictors, namely through coefficient shrinking. The Lasso model not only selects far fewer predictors, thus limiting the likelihood of overfitting, but also considers different types of features as relevant in the model. In contrast to the two non-linear models, the Lasso regression forecasts the redenomination risk using a selection of macroeconomic predictors consisting mainly of lags in the USD/EUR exchange rate, US and EZ overnight index swaps, government bond spreads and EZ break-even inflation. We find that these observations are consistent with the literature. Overall, we therefore consider Lasso regression to be the best performing model to forecast redenomination risk in the euro zone.

# 9 Appendix

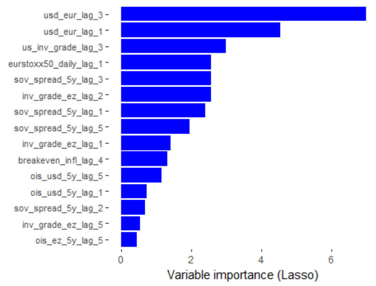
## 9.1 Graphs



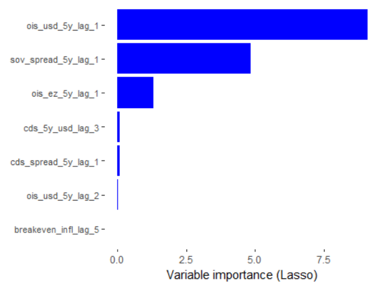

Italy



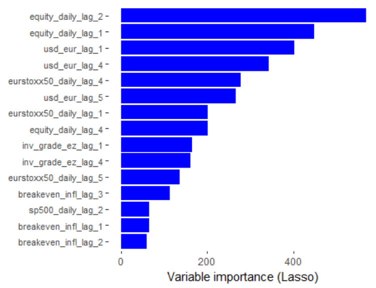

Spain



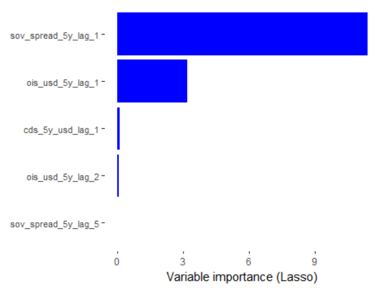

France



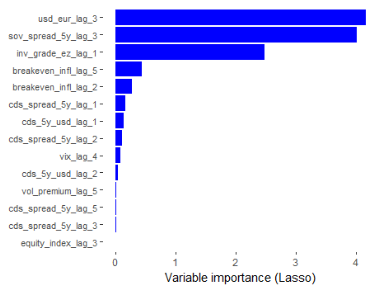

Portugal



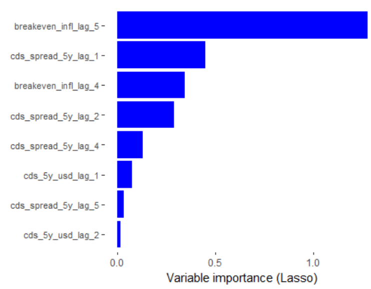

Greece



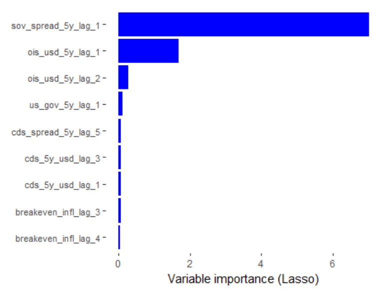

Ireland




Austria

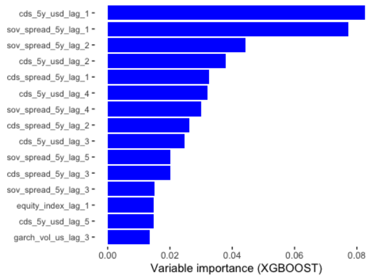



Netherlands

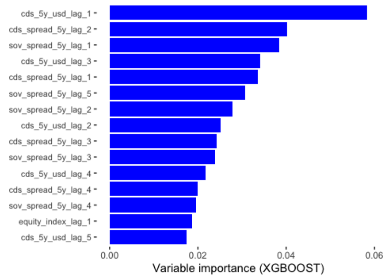



Belgium

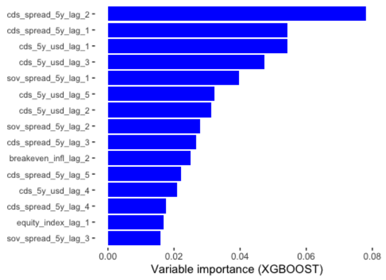
Figure 5: Lasso variable importance (Top 15) - Single Country Estimation, Out of Sample, (Full sample)



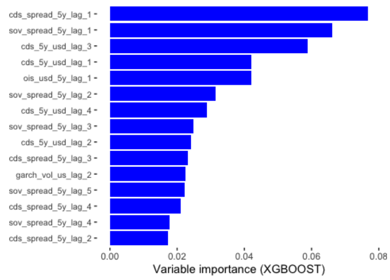

Italy



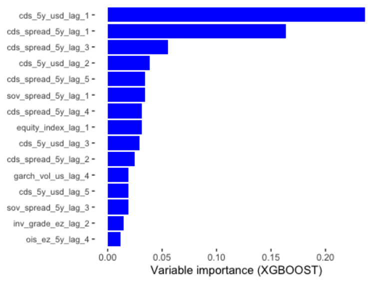

Spain



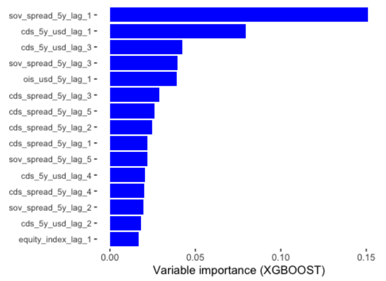

France



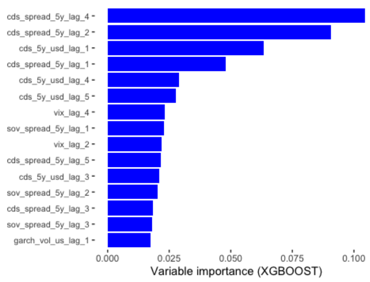

Portugal



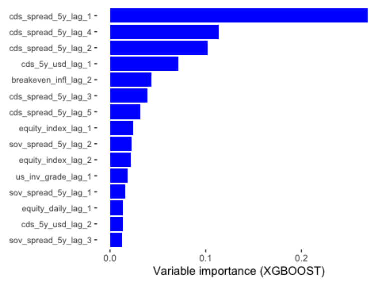

Greece



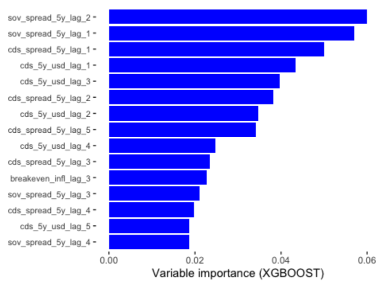

Ireland




Austria




Netherlands




Belgium

Figure 6: XGB variable importance (Top 15) - Single Country Estimation, Out of Sample, (Full sample)

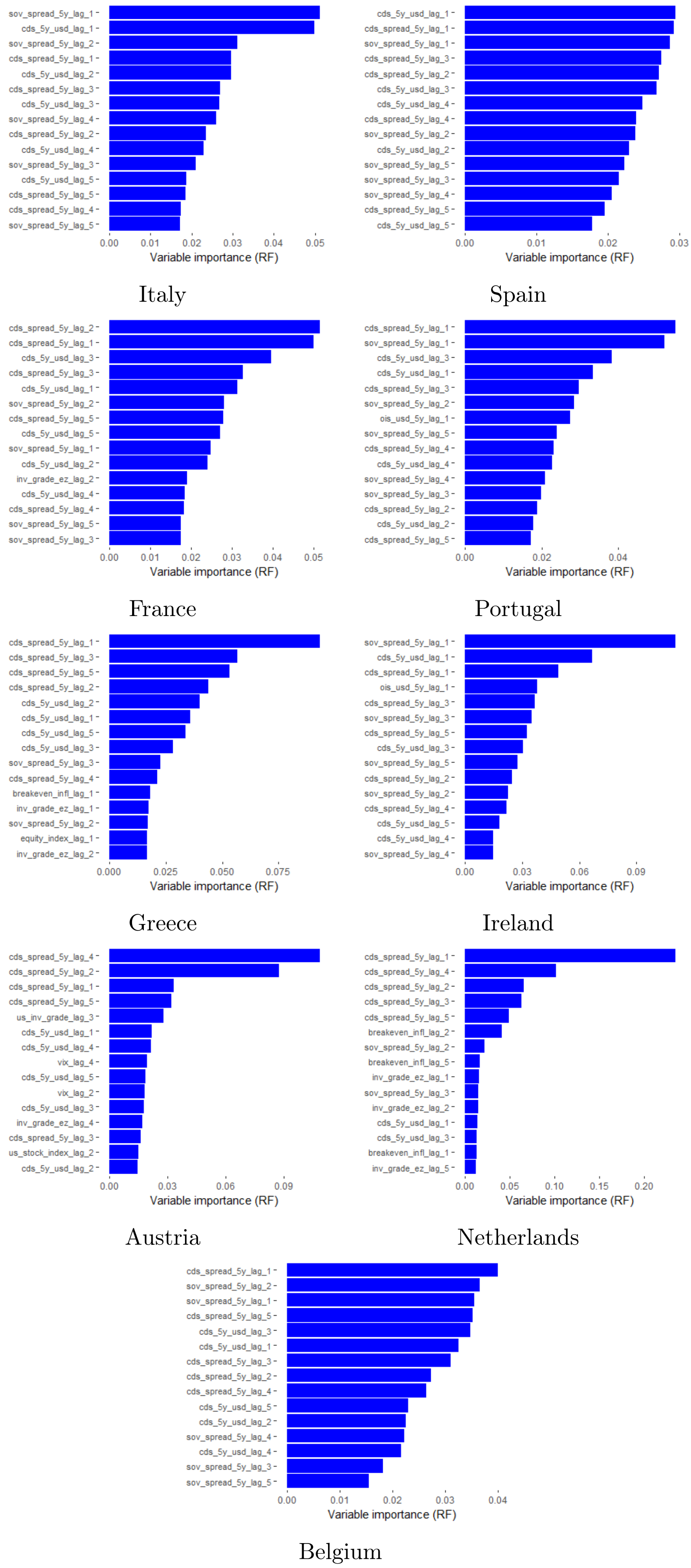



Figure 7: Random Forest variable importance (Top 15) - Single Country Estimation, Out of Sample, (Full sample)

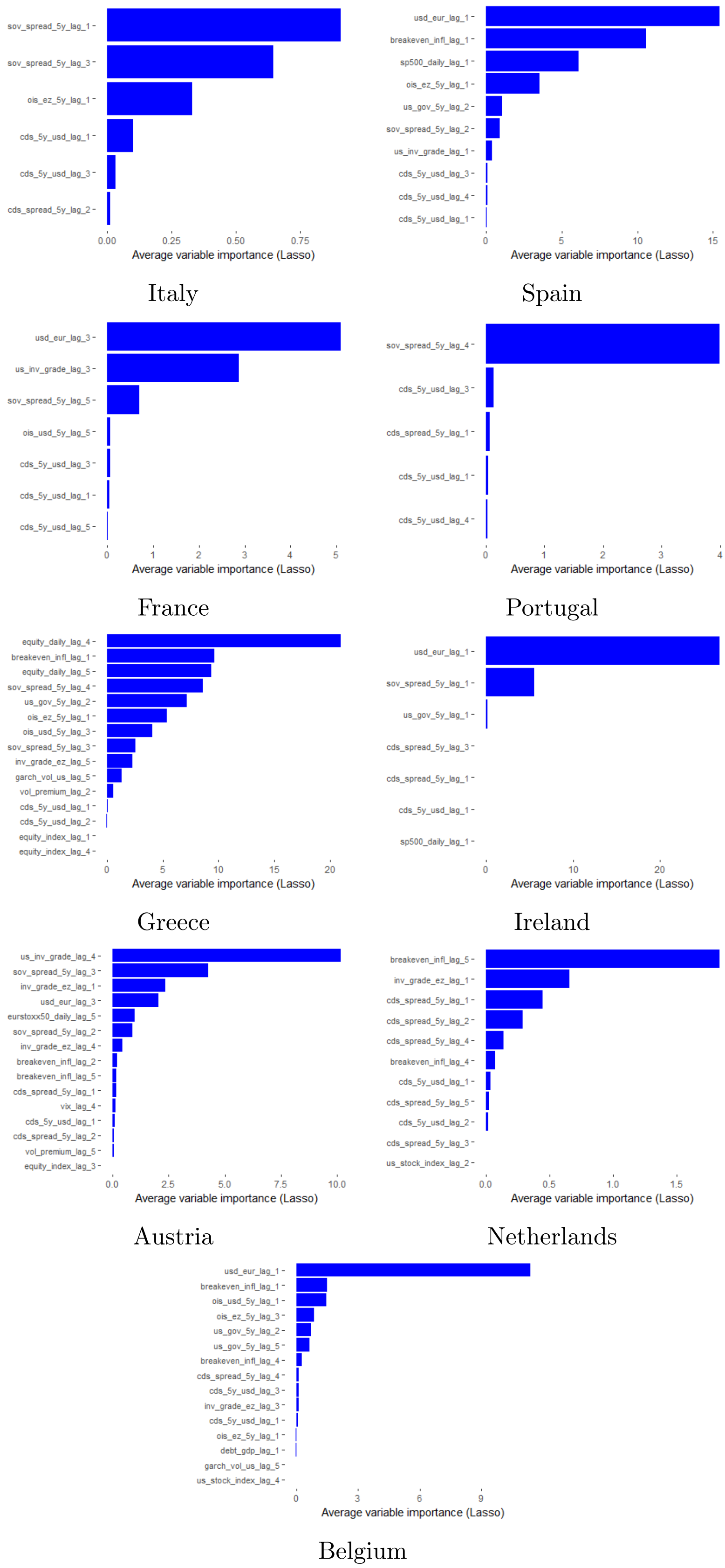



Figure 8: Lasso average variable importance (Top 15) - Rolling Time Series - Single Country Estimation (Debt Crisis)

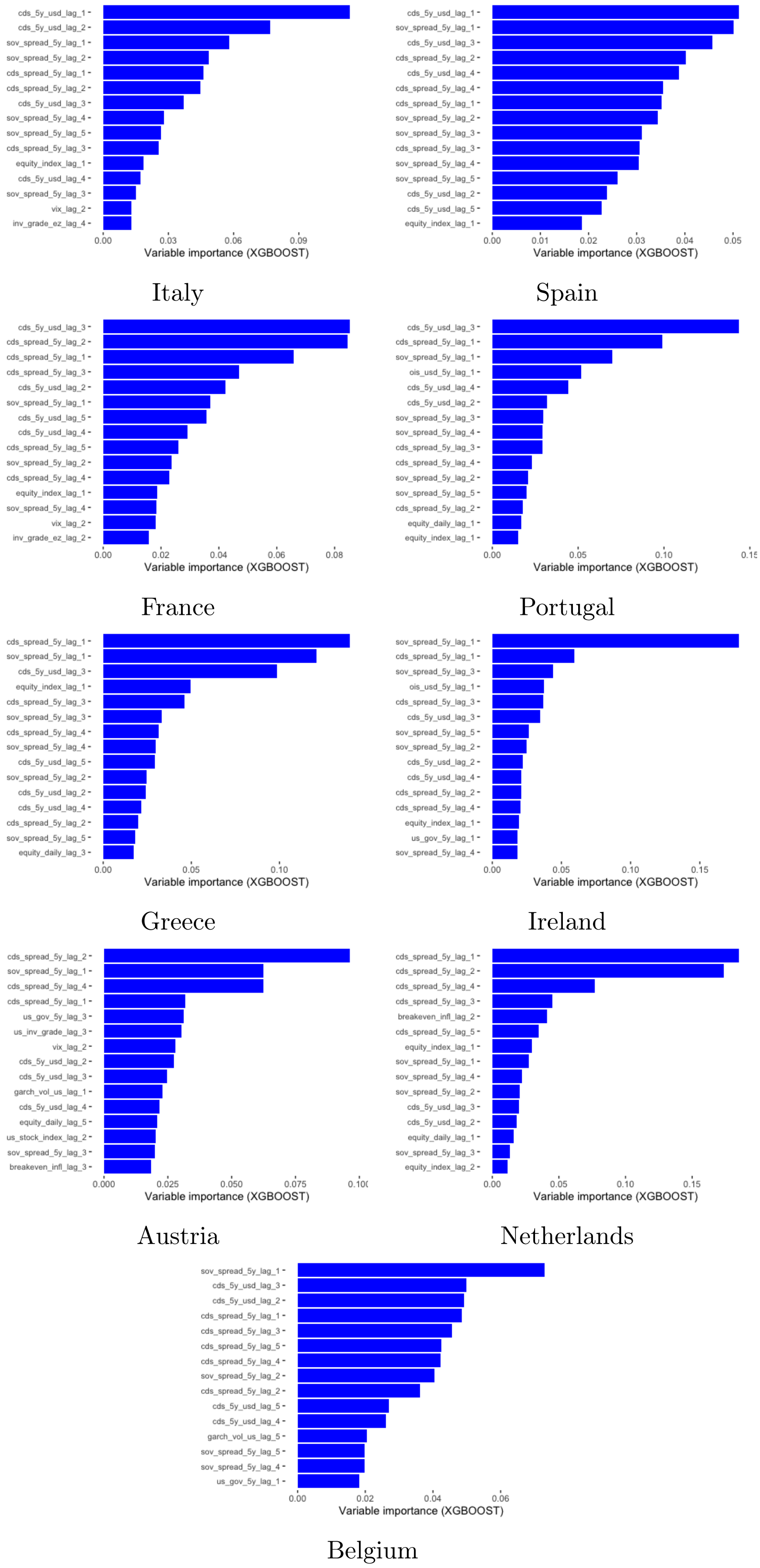



Figure 9: XGB average variable importance (Top 15) - Rolling Time Series - Single Country Estimation (Debt Crisis)

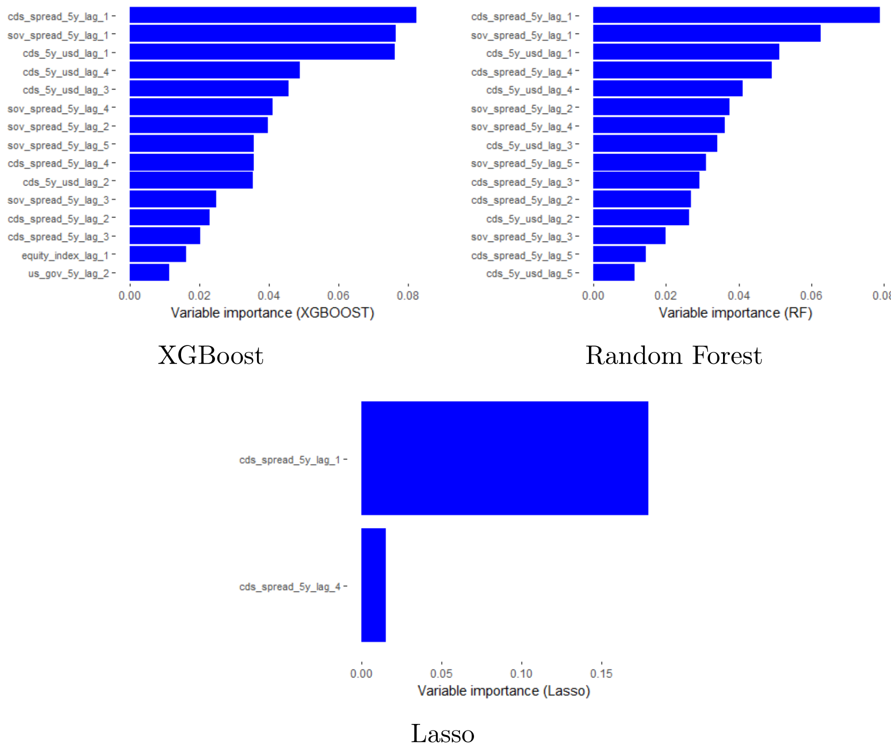



Figure 10: Southern Countries Variable importance (Top 15) - Out of Sample - Joint Country Estimation (Full sample)

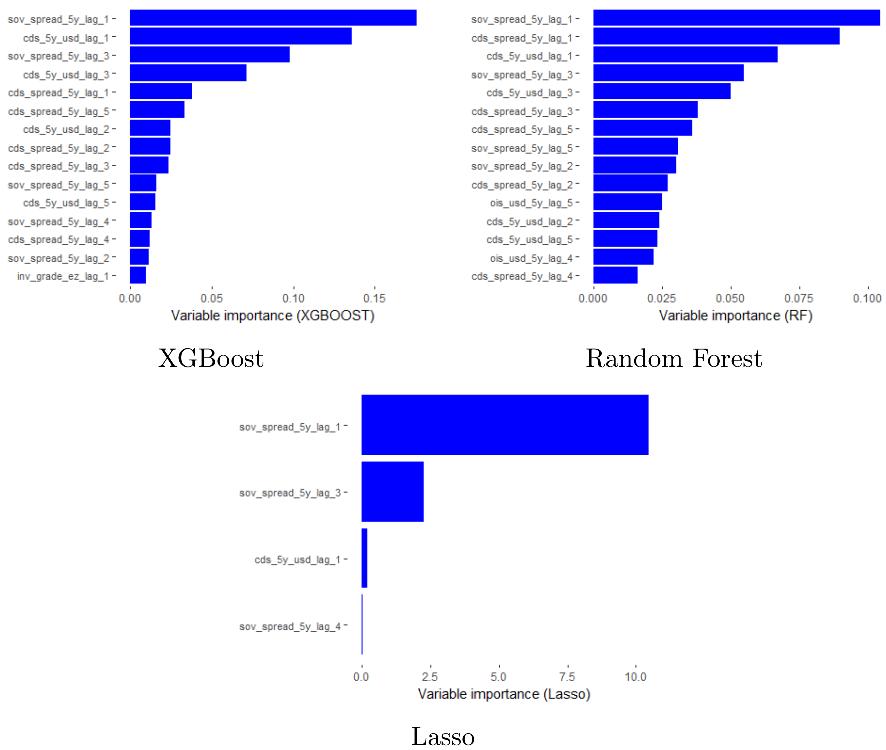
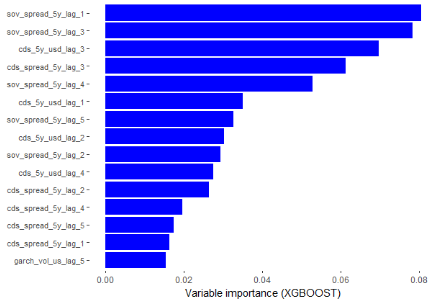
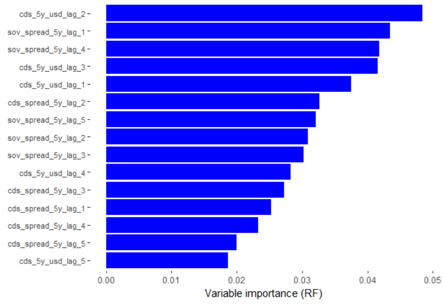


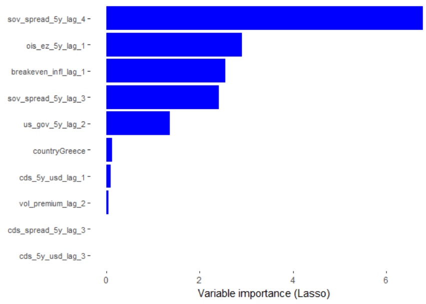

Figure 11: Nordic Countries Variable importance (Top 15) - Out of Sample - Joint Country Estimation (Full sample)




XGBoost

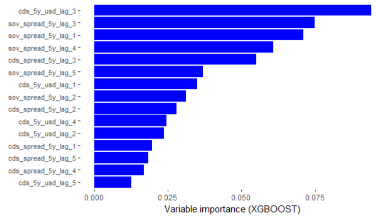



Random Forest

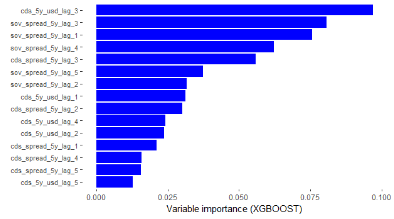



Lasso

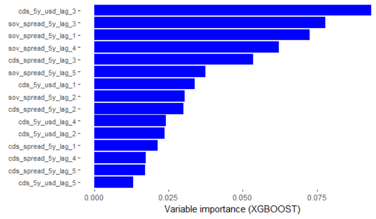
Figure 12: Variable importance (Top 15) - Out of Sample - Joint Country Estimation (Debt Crisis)



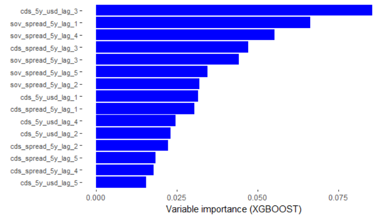

Italy



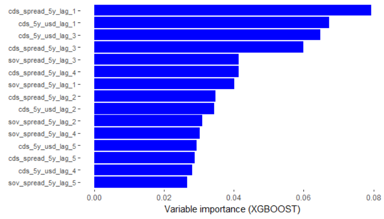

Spain



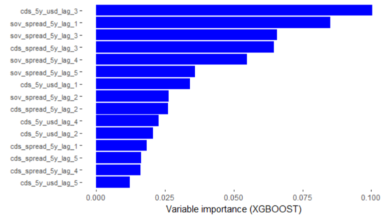

France



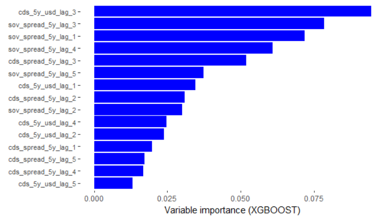

Portugal



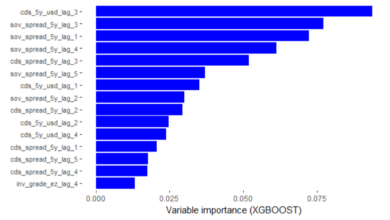

Greece



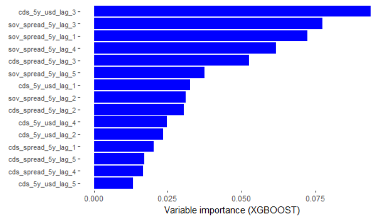

Ireland




Austria




Netherlands




Belgium

Figure 13: Average variable importance (Top 15) - Rolling Time Series - Joint Country Estimation (Debt Crisis)

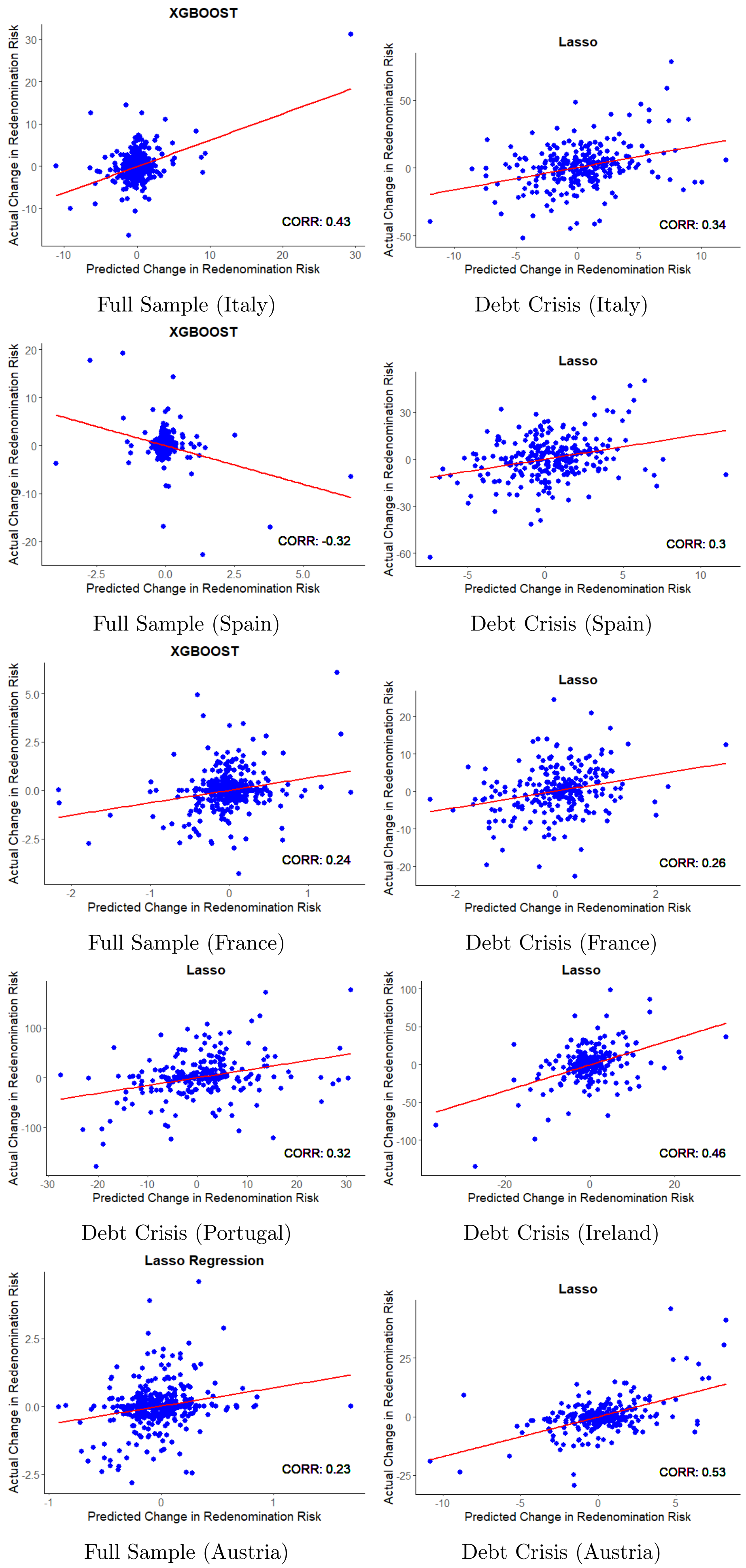
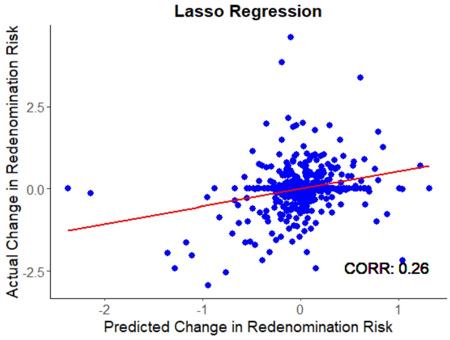
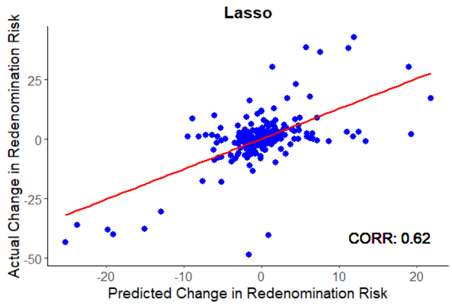


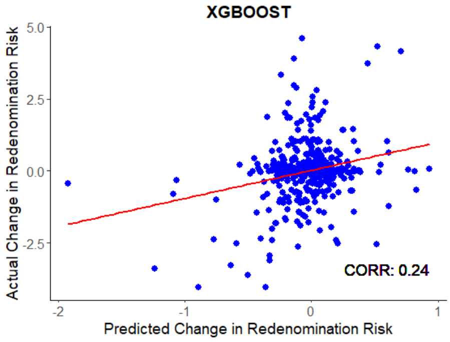

Figure 14: Optimal h-step ahead forecasts, Predicted vs Actual values



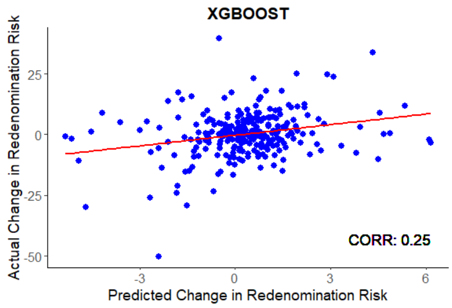

Full Sample (Netherlands)




Debt Crisis (Netherlands)




Full Sample (Belgium)




Debt Crisis (Belgium)

Figure 15: Optimal h-step ahead forecasts, Predicted vs Actual values

## 9.2 Tests and Robustness Checks

| Variable | p-Value |
|---|---|
| CDS Spread 5y | 0.20 |
| CDS Spread 5y (t-1) | 0.20 |
| CDS 5y USD (t-1) | 0.40 |
| National Sovereign Spread 5y (t-1) | 0.15 |
| National Equity Index (t-1) | 0.48 |
| National Equity Index Daily Returns (t-1) | 0.01 |
| US Sovereign Spread 5y (t-1) | 0.47 |
| Global Equity Index (t-1) | 0.64 |
| VIX volatility Index (t-1) | 0.51 |
| OIS USD 5y (t-1) | 0.59 |
| OIS EZ 5y (t-1) | 0.98 |
| USD/EUR (t-1) | 0.74 |
| Break even inflation rate (t-1) | 0.82 |
| Investment grade EZ (t-1) | 0.78 |
| S&P 500 daily returns (t-1) | 0.01 |
| US stock market volatility from a GARCH (1,1) (t-1) | 0.03 |
| Volatility premium (t-1) | 0.51 |
| Investment grade US (t-1) | 0.69 |
| EZ stock market volatility from a GARCH (1,1) (t-1) | 0.01 |

Table 10: ADF test results, variables are selected from the aggregated panel

| Model | IT | ES | FR | PT | GR | IE | AT | NL | BG |
|---|---|---|---|---|---|---|---|---|---|
| Lasso | -9.31 | -9.72 | -8.66 | -11.30 | 0.05 | -16.42 | -12.06 | -8.27 | -8.03 |
| | 0.00 | 0.00 | 0.00 | 0.00 | 0.96 | 0.00 | 0.00 | 0.00 | 0.00 |
| RF | -7.71 | -8.61 | -4.80 | -10.19 | -3.07 | -16.66 | -8.87 | -5.31 | -7.95 |
| | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| XGB | -5.32 | -7.23 | -4.01 | -9.67 | -3.24 | -16.99 | -9.36 | -5.71 | -7.77 |
| | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 11: Diebold Mariano test results, single country analysis (Full sample). For each Machine Learning model we plot the test statistics in row 1 and the corresponding p-value in the row below with the ARMA being the baseline model.

| Model | IT | ES | FR | PT | GR | IE | AT | NL | BG |
|---|---|---|---|---|---|---|---|---|---|
| Lasso | -0.77 | 0.36 | 0.70 | -0.44 | -0.59 | -2.43 | -2.15 | -1.67 | 0.30 |
| | 0.44 | 0.72 | 0.49 | 0.66 | 0.56 | 0.02 | 0.03 | 0.10 | 0.77 |
| RF | 0.28 | 0.41 | 1.68 | 2.09 | 1.89 | -0.31 | 1.09 | 0.35 | 1.40 |
| | 0.78 | 0.68 | 0.09 | 0.04 | 0.06 | 0.76 | 0.28 | 0.73 | 0.16 |
| XGB | 0.15 | -0.06 | 0.72 | 1.22 | -0.31 | -1.78 | 1.07 | -0.89 | -0.53 |
| | 0.88 | 0.95 | 0.47 | 0.22 | 0.76 | 0.08 | 0.29 | 0.38 | 0.60 |

Table 12: Diebold Mariano test results, single country analysis (Debt Crisis). For each Machine Learning model we plot the test statistics in row 1 and the corresponding p-value in the row below with the ARMA being the baseline model.

| Model | IT | ES | FR | PT |
|---|---|---|---|---|
| Lasso | -1.83 | -5.98 | -3.04 | -8.93 |
| | 0.07 | 0.00 | 0.00 | 0.00 |
| RF | -1.33 | -3.36 | 1.23 | -8.26 |
| | 0.18 | 0.00 | 0.22 | 0.00 |
| XGB | -1.34 | -5.24 | 1.24 | -8.69 |
| | 0.18 | 0.00 | 0.21 | 0.00 |

(a) Diebold Mariano test results, joint analysis (Full sample, Southern countries.)

| IE | AT | NL | BG |
|---|---|---|---|
| -10.64 | -4.76 | 0.01 | -6.96 |
| 0.00 | 0.00 | 0.99 | 0.00 |
| -11.86 | -10.37 | -4.45 | -9.55 |
| 0.00 | 0.00 | 0.00 | 0.00 |
| -11.50 | -9.27 | -5.66 | -8.87 |
| 0.00 | 0.00 | 0.00 | 0.00 |

(b) Diebold Mariano test results, joint analysis (Full sample, Nordic countries.)

Table 13: For each Machine Learning model we plot the test statistics in row 1 and the corresponding p-value in the row below with the ARMA being the baseline model.

| Model | IT | ES | FR | PT | GR | IE | AT | NL | BG |
|---|---|---|---|---|---|---|---|---|---|
| Lasso | -0.27 | -2.30 | 1.52 | -1.76 | -2.48 | -1.79 | -1.83 | 1.30 | -2.18 |
| | 0.78 | 0.02 | 0.13 | 0.08 | 0.01 | 0.07 | 0.07 | 0.19 | 0.03 |
| Random Forest | -1.24 | -2.67 | 1.96 | -1.75 | -2.47 | -2.72 | -1.20 | 2.40 | -2.45 |
| | 0.21 | 0.01 | 0.05 | 0.08 | 0.01 | 0.01 | 0.23 | 0.02 | 0.01 |
| XGB | 1.09 | 1.03 | 2.54 | 1.40 | -1.95 | 0.36 | 2.40 | 4.06 | 1.34 |
| | 0.28 | 0.30 | 0.01 | 0.16 | 0.05 | 0.72 | 0.02 | 0.00 | 0.18 |

Table 14: For each Machine Learning model we plot the test statistics in row 1 and the corresponding p-value in the row below with the ARMA being the baseline model.

## 9.3 Additional Information

| Model | Parameter |
|---|---|
| Random Forest | **n** (number of trees) |
| Random Forest | **depth** (maximum depth of trees) |
| Random Forest | **mtry** (number of variables randomly selected at each split point) |
| XGBOOST | **n** (boosting rounds) |
| XGBOOST | **max depth** (maximum depth of a tree) |
| XGBOOST | **eta** (the learning rate or shrinkage factor) |
| XGBOOST | **subsample** (the fraction of observations to be randomly sampled for each tree) |
| XGBOOST | **colsample** (the fraction of columns to be randomly sampled for each tree) |
| XGBOOST | **min child weight** (the minimum sum of instance weight needed in a child) |
| XGBOOST | **gamma** (the minimum loss reduction required to make a split) |

Table 15: Selection of hyperparameters for each model, in reference to [Belly et al., 2021]

| | TS | IT | ES | FR | PT | GR | IE | AT | NL | BL |
|---|---|---|---|---|---|---|---|---|---|---|
| **n** | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| **max depth** | 12 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| **eta** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| **subsample** | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| **colsample** | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| **gamma** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 16: XGBOOST hyperparameters by country, TS=Total Sample of all countries. Hyperparameter tuning over full sample.

| | TS | IT | ES | FR | PT | GR | IE | AT | NL | BL |
|---|---|---|---|---|---|---|---|---|---|---|
| **n** | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| **max depth** | 6 | 16 | 16 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| **eta** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| **subsample** | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| **colsample** | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| **gamma** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 17: XGBOOST hyperparameters by country, TS=Total Sample of all countries. Hyperparameter tuning during Debt Crisis.

Table 18: Explanatory variables in reference to [De Santis, 2019] and [Montes et al., 2022]

| **A measure for..** | **Description** | **Source** |
|---|---|---|
| US government bond yields | US government bond yield at 5-year maturity | Datastream |
| Sovereign spreads for each Euro Zone country | Sovereign yield 5y - 5y OIS | Datastream |
| US stock market index | Data Stream Global Equity Index | Datastream |
| US volatility premium | VIX index - realised volatility for the US stock market (resulting from a GARCH(1,1) on the daily SP500 returns.) | Datastream |
| US investment grade | Spread between US corporate BBB and AAA 7-10-year (USD). | FRED St.Louis |
| EUR depreciation vs USD 5-year forward. | Difference between euro area and the US OIS risk free rate at 5y | Datastream |
| EUR/USD depreciation | | Datastream |
| Breakeven inflation rate | five-year forward break-even inflation rate five years ahead provided by the ECB. | Datastream |
| EA 5-yr OIS. | Euro OIS rate at 5-year | Datastream |
| EA stock market volat | realised volatility obtained as a GARCH(1,1) on the daily euro area stock market returns . | Datastream |
| EA investment grade | spread between EZ corporate BBB and AAA 7-10-year (Euro). | FRED St. Louis |

| | | |
|---|---|---|
| Greek sovereing spread | | Datastream |
| Sovereign yield bid ask spread | Difference between 5-year bid and ask EUR-denominated sovereign yields | Datastream |
| Market perception of redenomination risk | Difference between the quanto CDS of Italy, Spain or France and the quanto CDS of Germany | Datastream |
| Debt to GDP (monthly) | | Datastream |
| Change in international Reserves (monthly) | | Datastream |
| Real Interest Rate (monthly) | | Datastream |
| Economic Sentiment Index (monthly) | | Datastream |
| Euro Zone Break Up Index (monthly) | | Datastream |

# References

[Aizenman et al., 2016] Aizenman, J., Jinjarak, Y., and Park, D. (2016). Fundamentals and sovereign risk of emerging markets. *Pacific Economic Review*, 21(2):151–177.

[Alsakka and ap Gwilym, 2012] Alsakka, R. and ap Gwilym, O. (2012). Foreign exchange market reactions to sovereign credit news. *Journal of International Money and Finance*, 31(4):845–864.

[Arakelian et al., 2019] Arakelian, V., Dellaportas, P., Savona, R., and Vezzoli, M. (2019). Sovereign risk zones in europe during and after the debt crisis. *Quantitative Finance*, 19(6):961–980.

[Balduzzi et al., 2020] Balduzzi, P., Brancati, E., Brianti, M., and Schiantarelli, F. (2020). Populism, political risk and the economy: Lessons from italy.

[Belly et al., 2021] Belly, G., Boeckelmann, L., Caicedo Graciano, C. M., Di Iorio, A., Istrefi, K., Siakoulis, V., and Stalla-Bourdillon, A. (2021). Forecasting sovereign risk in the euro area via machine learning. *Journal of Forecasting*.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

[Chan-Lau, 2017] Chan-Lau, M. J. A. (2017). *Lasso regressions and forecasting models in applied stress testing*. International Monetary Fund.

[Chen and Guestrin, 2016] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

[Chiaramonte et al., 2018] Chiaramonte, A., Emanuele, V., Maggini, N., and Paparo, A. (2018). Populist success in a hung parliament: The 2018 general election in italy. *South European Society and Politics*, 23(4):479–501.

[Corradin et al., 2021] Corradin, S., Grimm, N., and Schwaab, B. (2021). Euro area sovereign bond risk premia during the covid-19 pandemic.

[De Santis, 2019] De Santis, R. A. (2019). Redenomination risk. *Journal of Money, Credit and Banking*, 51(8):2173–2206.

[Diebold, 2015] Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1.

[Fonti and Belitser, 2017] Fonti, V. and Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam research paper in business analytics*, 30:1–25.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

[Heinz and Sun, 2014] Heinz, M. F. F. and Sun, M. Y. (2014). *Sovereign CDS spreads in Europe: The role of global risk aversion, economic fundamentals, liquidity, and spillovers.* International Monetary Fund.

[Hetzel, 2015] Hetzel, R. L. (2015). Should greece remain in the eurozone?

[Klose and Weigert, 2014] Klose, J. and Weigert, B. (2014). Sovereign yield spreads during the euro crisis: Fundamental factors versus redenomination risk. *International Finance*, 17(1):25–50.

[Lando and Bang Nielsen, 2018] Lando, D. and Bang Nielsen, A. (2018). Quanto cds spreads. *Available at SSRN 3268890*.

[Montes et al., 2022] Montes, G. C., Nicolay, R., and Pereira, F. (2022). Does fiscal sentiment matter for sovereign risk? *The Quarterly Review of Economics and Finance*, 86:18–30.

[Ramraj et al., 2016] Ramraj, S., Uzir, N., Sunil, R., and Banerjee, S. (2016). Experimenting xgboost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40):651–662.

[Tang et al., 2018] Tang, C., Garreau, D., and von Luxburg, U. (2018). When do random forests fail? *Advances in neural information processing systems*, 31.

[Tyralis and Papacharalampous, 2017] Tyralis, H. and Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms*, 10(4):114.

[Wisniewski and Lambe, 2015] Wisniewski, T. P. and Lambe, B. J. (2015). Does economic policy uncertainty drive cds spreads? *International Review of Financial Analysis*, 42:447–458.

[Xu et al., 2019] Xu, G., Liu, M., Jiang, Z., Söffker, D., and Shen, W. (2019). Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors*, 19(5):1088.