BETA
Bureau d'économie théorique et appliquée

Documents de travail

« Forecasting Public Debt in the Euro Area Using Machine Learning: Decision Tools for Financial Markets»

Auteurs

Amélie Barbier-Gauchard, Emmanouil Sofianos

Document de Travail n° 2024 – 47

Novembre 2024

Bureau d'Économie Théorique et Appliquée BETA

www.beta-economics.fr

@beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Université de Strasbourg
UNIVERSITÉ DE LORRAINE
INRAE
AgroParisTech

Forecasting Public Debt in the Euro Area Using Machine Learning: Decision Tools for Financial Markets

Amelie BARBIER-GAUCHARD[1], Emmanouil SOFIANOS[2]

Abstract

The situation of public finance in the eurozone remains a burning issue for certain Euro area countries. The financial markets, the main lenders of the Member States, are more attentive than ever to any factor which could affect the trajectory of public debt in the long term. The risk of bankruptcy of a Member State and a domino effect for the entire monetary union represents the ultimate risk weighing on the Eurozone. This paper aims to forecast the public debt, with a universal model, on a national level within the Euro area. We use a dataset that includes 566 independent variables (economic, financial, institutional, political and social) for 17 Euro area countries, spanning the period from 2000 to 2022 in annual frequency. The dataset is fed to four machine learning (ML) algorithms: Decision Trees, Random Forests, XGBoost and Support Vector Machines (SVM). We also employ the Elastic-Net Regression algorithm from the area of Econometrics. The best model is an XGBoost with an out-of-sample MAPE of 8.41%. Moreover, it outperforms the projections of European Commission and IMF. According to the VIM from XGBoost, the most influential variables are the past values of public debt, the male population in the ages 50-54, the regulatory quality, the control of corruption, the female employment to population ratio for the ages over 15 and the 10 year bond spread.

JEL classification: C53; H63; H68

Keywords: Public Debt; Euro Area; Machine Learning; Forecasting;

The authors are deeply grateful to Yabre Tobignaré for his valuable suggestions and help with data.

[1]University of Strasbourg, University of Lorraine, CNRS, BETA, 67000, Strasbourg, France, abarbier@unistra.fr.
[2]University of Strasbourg, University of Lorraine, CNRS, BETA, 67000, Strasbourg, France, sofianos@unistra.fr, corresponding author.

1. Introduction

The situation of public finance in the eurozone remains a burning issue for certain Euro area countries. The financial markets, the main lenders of the Member States, are more attentive than ever to any factor which could affect the trajectory of public debt in the long term. The risk of bankruptcy of a Member State and a domino effect for the entire monetary union represents the ultimate risk weighing on the Eurozone.

Although there have been fiscal discipline rules since the birth of the eurozone on January 1, 1999, the trend of public debt continues a worrying dynamic for certain countries. These rules, introduced by the Stability and Growth Pact (SGP, 1996), which have been reformed several times, must combine the sound management of national public finance in eurozone countries on the one hand, but also leave Member States sufficiently margin to achieve their fiscal policy on the other hand. This dual objective, particularly tricky to achieve, was reaffirmed by the reform adopted in 2024 and which comes into force on January 1, 2025.

This reform, which reaffirms the thresholds of 3% for the public deficit and 60% for the public debt, leaves enough room for maneuver for Member States, through other indicators, but contribute to maintain the distrust of the financial markets towards certain Member States. With financial markets now being the number one lender to Member States, the existence of these rules is not enough to guarantee fiscal seriousness. It is in this context that forecasts of the level of future public debt become crucial, potentially leading rating agencies to review their assessment of the quality of public assets held.

According to Estefania-Flores et al. (2023), there is a significant positive forecast error in debt-to-GDP ratio projections in both advanced economies and emerging markets and developing economies. This error in public debt forecasts can have serious repercussions. High, unpredicted, levels of public debt often lead to higher interest rates on government bonds, increasing the cost of borrowing for governments. This in turn can limit their ability to finance future projects. It can also lead to low investor confidence as investors become increasingly concerned about the government's ability to repay its debt, which can lead to reduced foreign investment and economic instability. The government's fiscal flexibility may also be reduced, limiting its ability to implement expansionary fiscal policies during economic downturns.

Furthermore, the level of public debt can have a significant impact on financial stability. Inflationary pressures can also threaten currency stability, making the broader financial environment more volatile and unpredictable. Strained government finances produce excessive or unsustainable public debt, raising concerns of an eminent debt restructuring or even default. Financial market disruptions, greater borrowing costs, and increased systemic risks are just a few of the negative effects that increased public debt may have. An unpredicted, high public dept can cause cascading effects, such as in the 2009 debt crises that simultaneously hit many EU countries (Ireland, Spain, Portugal, Italy and more severely Greece). Furthermore, the efficiency of the transmission mechanisms of monetary policy might be impacted by the level of public debt. High levels of public debt can discourage private investment, raise the cost of borrowing for the private sector, and reduce the effectiveness of monetary policy actions meant to promote economic growth.

Numerous methods exist for forecasting public debt, but none of them consider institutional and political factors such as the electoral cycle, government communications or the composition of budget committees, for the Euro area. The aforementioned underscore the importance of public debt management for the stability of the Euro area.

The aim of this paper is to create a universal forecasting model that can accurately predict the level of public debt on a national level within the Euro area. The proposed arsenal to complete this goal includes 4 machine learning algorithms alongside with the elastic-net regression model, as benchmark, from the area of econometrics. In this setting, complex relationships and non-linear dynamics can be identified and captured, improving forecast accuracy and enhancing the ability for policy decision-making by utilizing advanced machine learning techniques and broad indicator sets such as, economic, financial, institutional, political and social indicators, as confirmed from Belly et al. (2023). Moreover, a critical strength of the paper is the identification and ranking, according to significance, of the most influential variables in terms of public debt forecasting and the use of the Partial Dependence Plots. These tools can contribute to governments and central banks to better formulate economic policy recommendations in favor of a better monitoring of national public debt evolution.

The paper is organized as follows: in Section 2 we will present the literature review, in Section 3 we will briefly discuss the methodologies and the dataset and in Section 4 we present our empirical results. Finally, Section 5 concludes the paper.

2. Literature Review

The literature on the determinants of public debt is abundant. There are various factors behind the fluctuations of public debt (Knapková et al., 2019) for the Euro area (Belly et al., 2023). They vary from historical factors (Stasavage, 2003; Page, 2018; D'Erasmo & Mendoza, 2018), to political (Di Bartolomeo et al., 2018; Ono & Uchida, 2018), and economic factors (Poghosyan, 2017; Chen et al., 2017).

Dawood et al., (2017) created models demonstrating that it is critical to incorporate variables that can detect the likelihood of spillover from the banking sector and the foreign exchange market when building an effective Early Warning System (EWS) for sovereign debt crises. Fioramanti (2008), also developed an ANN-based early warning system that can, under certain conditions, outperform more consolidated methods.

The links between government and private sector debt, the financial sector, and the political mechanisms controlling the resolution of fiscal and economic challenges (Kalemli-Özcan et al., 2016) demonstrate the necessity of testing not only financial and economic indicators, but also political ones.

Many authors focused primarily on explaining how macroeconomic indicators affected public debt. Nonetheless, numerous research (e.g. Vuckovic & Basarac Sertic, 2013; Eusepi & Wagner, 2017) have confirmed that non-economic factors have also a significant impact on public debt; but, yet, they still did not take into account the interaction of macroeconomic and non-economic factors.

For example, Estefania-Flores et al., (2023) discovered that the forecast error on debt is larger during periods of elections. This may be a result of governments choosing to follow different regimes, not pursuing fiscal consolidation in elections periods. All these non-economic factors can have a considerable influence on the way in which the financial markets assess a country's exposure to the risk of default, particularly the rating agencies for national public debt.

3. Methodology and Dataset

In this section, we will briefly discuss the various methodologies and techniques used in this research. In section 3.1, after a short introduction to machine learning and AI, the various machine learning methodologies are presented, namely Support Vector Machines, Decision Trees, Random Forests and XGBoost. In section 3.2 the Elastic-Net regression methodology is explained, from the area of econometrics. Section 3.3

presents the problem of overfitting and the various techniques used to overcome it. The Recursive Feature Elimination (RFE) technique used to select the most informative variables, is analyzed in section 3.4 and finally, in section 3.5 we discuss the Variable Importance Measure (VIM) that ranks the variables based on their contribution to a model. Finally, 3.6 describes the variables used in this paper.

3.1. Machine Learning methodologies

The field of machine learning was developed in the 1950s with the goal of equipping artificial intelligence (AI) systems with the "learning" component. The fundamental concept behind machine learning is that computers can learn from data, identify patterns, and make judgments with minimal human input. This capability is essential for automated analytical model construction, which is the foundation of machine learning.

Historically, large data sets have been essential for machine learning (Gogas and Papadimitriou, 2021). The availability of high-frequency data in the finance sector has led to the application of machine learning techniques to financial data. During the past few years, several techniques and methods have been used in smaller datasets as well, such as macroeconomic ones, providing results that outperform other econometric techniques (Sermpinis et al., 2014 inflation and unemployment, Gogas et al., 2022 unemployment, Sofianos et al., 2021 output gaps).

Support Vector Machines (SVM) are a well-known classification (Support Vector Classification – SVC) and regression (Support Vector Regression – SVR) technique, used particularly when non-linear patterns in the data or complex interactions are present. Due to its capacity to handle both linear and non-linear data, SVM can be utilized in a diverse range of real-world applications, including engineering, economics, finance, and more. In situations involving datasets that are contaminated by noise, outliers, or non-linear systems SVM are coupled with the so-called "kernel trick". This kernel function projects the original data from the data space to a higher-dimensional space, the feature space, where the dataset may be simpler to separate and more precisely classified or regressed (Cortes and Vapnik, 1995).

Decision Trees (DT) is a supervised machine learning technique, applied to regression and classification problems. These are top-down, node-and branch-based structures that resemble flowcharts (Figure 1). The data is divided recursively based on the most informative characteristic of the variables (Gogas et al., 2022). For regression tasks, the decision tree calculates predictions by averaging the target variable values of the training data points assigned to the same leaf node. Each node represents a criterion for splitting, whereas each branch represents an outcome. The top node, also known as the root node, represents the complete dataset, while the remaining nodes are referred to as decision nodes. The nodes that do not split any further, also known as leaves (or terminal nodes and leaf nodes), represent the final outcomes of the decision-making process (a value of public debt to GDP ratio in our case).

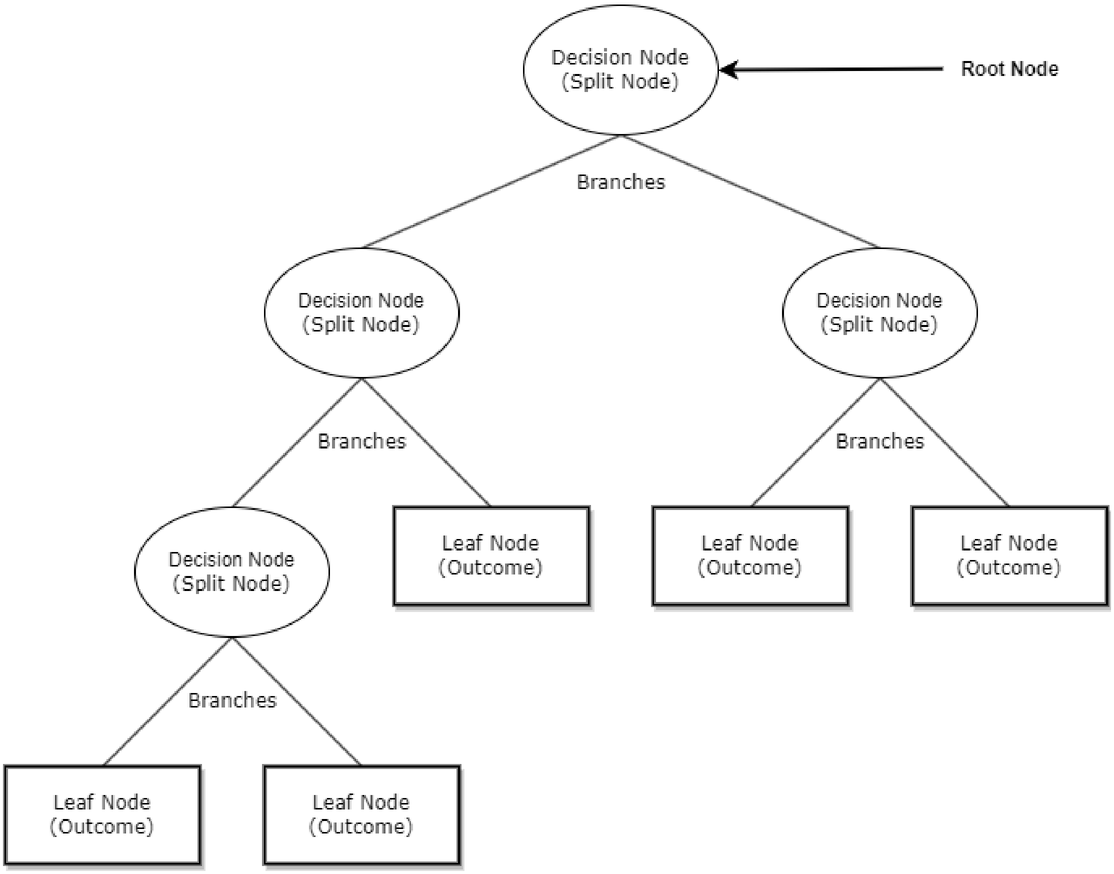


Figure 1: Example of a decision tree (Sofianos et al., 2024).

The benefits of decision trees are their ease of comprehension and good performance with training data. However, their main weakness is that they perform badly when handling out-of-sample data, thus, they have limited generalization ability. Their usual characteristics include low bias and high variance (overfit). One strategy to get around overfitting is to use ensembling techniques, where predictions from several different machine learning models are combined. The objective of ensemble learning is to combine the strengths and weaknesses of individual models to create an optimal model that performs better than any of the individual models. This procedure, which usually yields superior accuracy and robustness, entails combining the predictions from each algorithm to create a final forecast. Bagging (Breiman, 1996) and boosting (Freund and Schapire, 1997) are the two primary ensemble techniques. In bagging, each data point can be utilized to train a new decision tree, since each observation is sampled independently. Furthermore, each observation is equally important. In contrast, boosting creates decision trees sequentially, varying the weight or importance of each observation at each iteration. This allows for greater weight to be given to misclassified or hard-to-predict data points as the algorithm attempts to improve its overall performance and correct its faults (Figure 2).

In bagging, the average score of the model is obtained through the parallel and independent training of weak learners. In regression tasks, predictions are made using the average value of the target variable from the data points of each leaf node. The random forest algorithm is the most frequently used method of bagging (Breiman, 2001), where a distinct subsample of the original dataset is selected at random with replacement for every tree (bootstrapping).

Another ensemble technique is the XGBoost (eXtreme Gradient Boosting; Chen & Guestrin, 2016). In contrast to random forest, it creates an ensemble of decision trees in a stepwise manner, with each new tree attempting to improve on the mistakes made by the one before it (boosting). Gradient boosting is a framework that enables the algorithm to grow each tree with the goal of minimizing a given objective function. This function is typically a combination of a regularization term and a loss function. XGBoost is regarded as a cutting-edge algorithm that improves the precision and accuracy of the outcomes. This methodology's effectiveness has been generally acknowledged in a number of machine learning and data mining challenges (Chen and Guestrin, 2016).

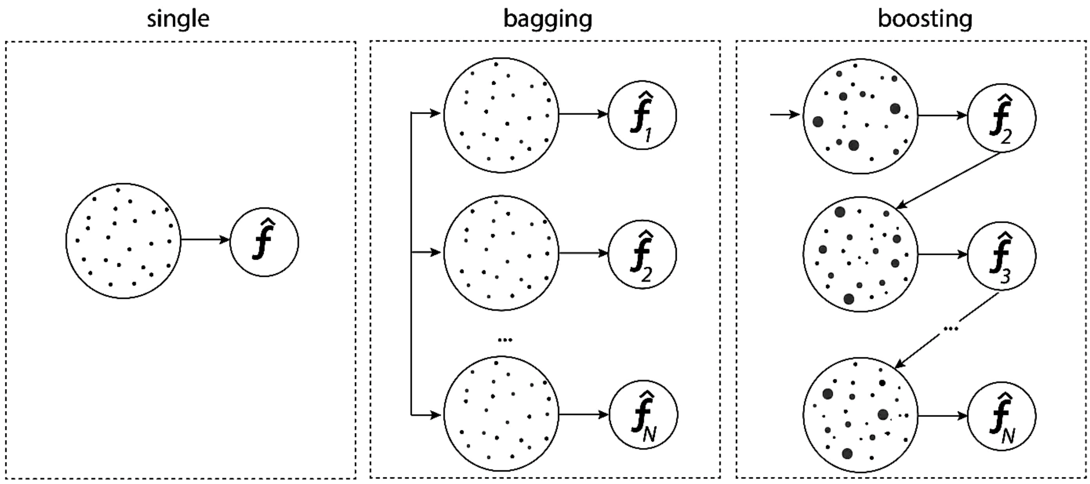


Figure 2: An illustration of the boosting and bagging processes. In that example, the size of each observation corresponds to its importance on each tree.

3.2. Elastic Net Regression

Elastic Net regression is a hybrid model that combines the strengths of two popular regularization techniques: Lasso (L1) and Ridge (L2) regression. It addresses the limitations of each method by introducing a penalty term that is a linear combination of the L1 and L2 regularization of the coefficient vector. This allows Elastic Net to select variables and handle multicollinearity more effectively than Lasso, while still encouraging sparsity in the solution like Lasso. The penalty terms L1 and L2 and their trade-off, are controlled by parameters tuned with cross validation, offering flexibility in model complexity and feature selection. By striking a balance between variable selection and regularization, Elastic Net regression provides a powerful tool for handling high-dimensional data with correlated predictors, commonly encountered in many research contexts.

3.3. Overfit

Overfitting occurs when a trained model is well-fitted to the training set but fails when applied to new and unknown data, meaning the model has low bias and high variance (Figure 3). This is also known as "low generalization ability".

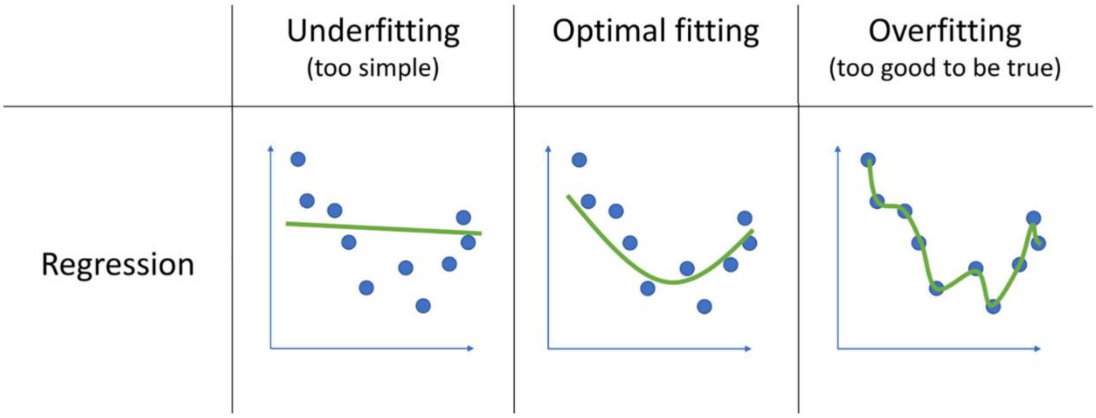


Figure 3: A visual representation of underfit, overfit and optimal fit in regression (Solanes & Radua, 2022). In underfitting the model is too simple and not able to identify the underlying pattern of the data. In overfitting, the model is too complex and fits the training data too well; thus, it is not able to identify the underlying pattern. When a model fits the data well and identifies the underlying pattern we have optimal fitting.

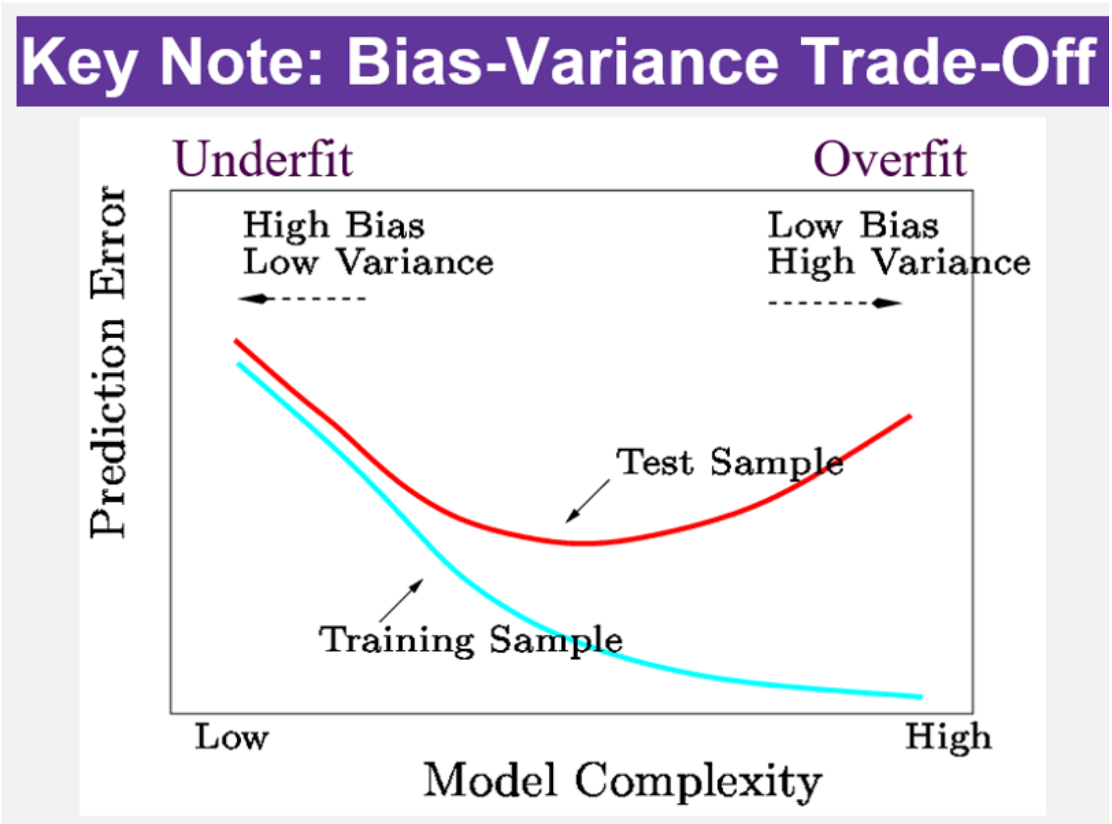


Figure 4: The Bias-Variance trade-off. The goal is to train a model that does not overfit (high variance and low bias) or underfit (high bias and low variance). The optimal model should minimize the prediction error achieving a balance between bias and variance.

To test the generalization ability of our models, the dataset was divided into two subsamples: a) the training sample (in-sample), on which the algorithms were trained, and b) the test sample (out-of-sample), where the generalization ability of the trained models was tested using "unseen" data, in order to assess the models' capacity for generalization (Figure 4). To prevent overfitting, cross-validation is employed during the training phase to identify the optimal hyperparameters for the models. The in-sample portion of the dataset is divided into k equal-sized subsets (folds) for cross-validation. For each distinct set of hyperparameters that is tested, the training/testing procedure is repeated k times. The testing of the model is conducted using a different fold for each iteration, whereas the training of the model is conducted using the remaining k-1 folds. The overall performance of each set of examined hyperparameters is calculated as the average performance over all test k folds (Figure 5). In our experiments, we employed a 5-fold configuration.

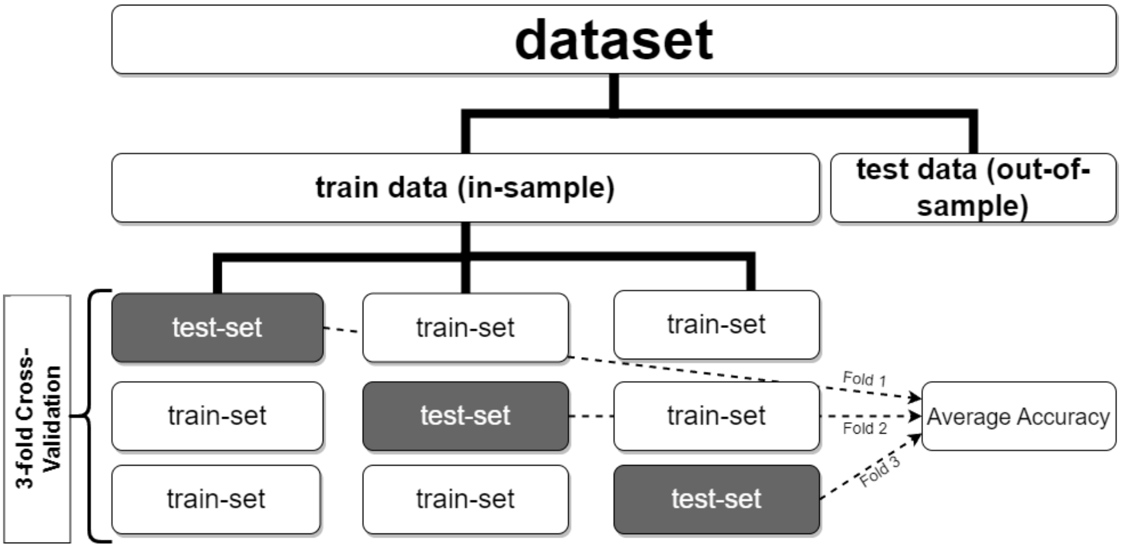


Figure 5: The illustration depicts a 3-fold cross-validation procedure. Each fold represents a test sample for each combination of hyperparameters under consideration. The remaining folds are utilized to train the model. To evaluate the model (in-sample) and test it on the out-of-sample portion of the dataset, that contains data that the model has not used during training, the average performance for each combination of parameters over the test k-folds is utilized (Gogas et al., 2019).

A variety of forecasting metrics can be calculated to assess and contrast the forecasting efficacy of our models. In this paper, we utilize the Mean Absolute Percentage Error (MAPE). This metric is the most prevalent in the relevant literature due to its simplicity and interpretability, and it is defined as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{1}$$

where y_i and $\hat{y}_i$ are the target variable's actual and predicted values, respectively, and n is the sample size. The average mean absolute difference between the predicted and actual values is measured in percentage terms by the MAPE metric.

3.4. Recursive Feature Elimination

To identify the most relevant features (variables) for the prediction problem, we employed a Recursive Feature Elimination (RFE) technique. RFE starts by fitting the model with all the features present and obtaining the importance of each feature. This can be done through a specific attribute or callable of the model (such as coefficients or feature importance). Then, the least important features are pruned from the current set of features. This process is repeated recursively on the pruned set until the desired number of features to select is finally reached. In our case, we used the Random Forest algorithm as an estimator. The Random forest ranks features based on their Gini significance, prioritizing those who are crucial for the model's performance (Guyon, Weston et al., 2002).

3.5. Variable Importance Measure

The Variable Importance Measure (VIM), ranks variables based on their contribution to the model and it can be computed in tree-based models to help us break the algorithm's black box nature. The VIM calculates the impact of each attribute split point on performance, weighted by the number of observations each node handles. VIMs are then averaged across all decision trees in the model. A higher value for this metric indicates that one feature is more important for generating a prediction than another.

3.6. The Dataset

The dataset consists of 17 Euro area countries[3] and 566 economic, financial, political, institutional and social variables (Table 3 at appendix) at an annual frequency spanning the period from 2000 to 2022 (panel data), for a total of 340 observations. A quick summary of the data can be found in Table 1. Three main categories of data are collected; a) economic and financial data, to cover macroeconomic features of each country, b) political and institutional variables, to consider institutional and political system features and c) social variables to integrate social demo features for each country. The economic and financial data mainly extracted from different sources

[3] From the 20 Euro Area countries, we excluded Cyprus, Malta and Croatia due to data availability restrictions. The economic weight of these 3 countries represents only 0.9% of the GDP of the euro zone (2023 GDP data).

(*Eurostat*, *the OECD* and *the World Bank*), cover traditional macroeconomic indicators (GDP, employment, inflation, international trade, international capital flows, activity by sector, public expenditure, public revenues, public deficit, public debt, 3-month short-term interest rate, 10-year bond yields, 10 year bond spread, …). Political and institutional data also come from different sources (*Worldwide Governance Indicators of the World Bank*, *the V-Dem Dataset*, *the International Country Risk Guide-Researcher Dataset* and the *International Institute for Democracy and Electoral Assistance Database)* and makes it possible to take into account the country's mode of government, the characteristics of the institutions (quality, control, etc.), the electoral system, the voting process, the result elections, respect for the rules of law, ... Social data from *the World Bank* and *the Standardized World Income Inequality Database* cover the socio-demographic characteristics of each country. Moreover, we added up to 3 lags for every variable, increasing the total number of variables to 1694. The dataset was split in two parts, the in-sample data (used for training the models) that includes 80% of the data, or 272 observations, and the out-of-sample (OOS) data, that includes 20% of the data, or 68 observations, to test the generalization ability of our models. According to this split, the OOS part consists of the last 4 years of the dataset (2019-2022). Some missing values (less than 3%) were filled with the mean value of each variable per country.

Table 1: Description of the data and their sources.

Category	**Sources *(Acronym)***	**Description**
Economic and financial	The World Bank Open Data *(WBOD)*	Free and open access to global development data
	Eurostat *(Eurostat)*	Statistics and data on Europe (Public Debt, inflation, deficit/surplus etc.)
	OECD *(OECD)*	Provides comprehensive and reliable economic, social, and environmental data for policy-making and analysis (long-term interest rates, short-term interest rates etc.).
Political and institutional	V-Dem database *(VDEM)*	Provides measures of democracy, governance, and political institutions (e.g. freedom of expression, civil liberties, political participation)
	International Country Risk Guide *(ICRG)*	Provides measures of democracy, governance, and political institutions (e.g. civil liberties, executive constraints)
	International Institute for Democracy and Electoral Assistance *(I-IDEA)*	Collection of voter turnout statistics from presidential, parliamentary and European Parliament elections, separately.

Social	Standardized World Income Inequality Database *(SWIID)*	Comparable Gini indices, market income inequality and information on absolute and relative redistribution.
	The World Bank, Gender Data portal *(WBGD)*	Database with social indicators and statistics (e.g. age population, gender equality, labour force participation, school enrolement, social protection system, retirement system etc.).

4. Empirical Results

4.1. Forecasting models results

For the empirical part, we first identified the 25 most informative variables, then we used them to train the models. The best models were identified through a cross validation process. Finally, we used the best model to interpret the results, break the so-called "black-box" of machine learning. This can help us to give a better economic interpretation to the results.

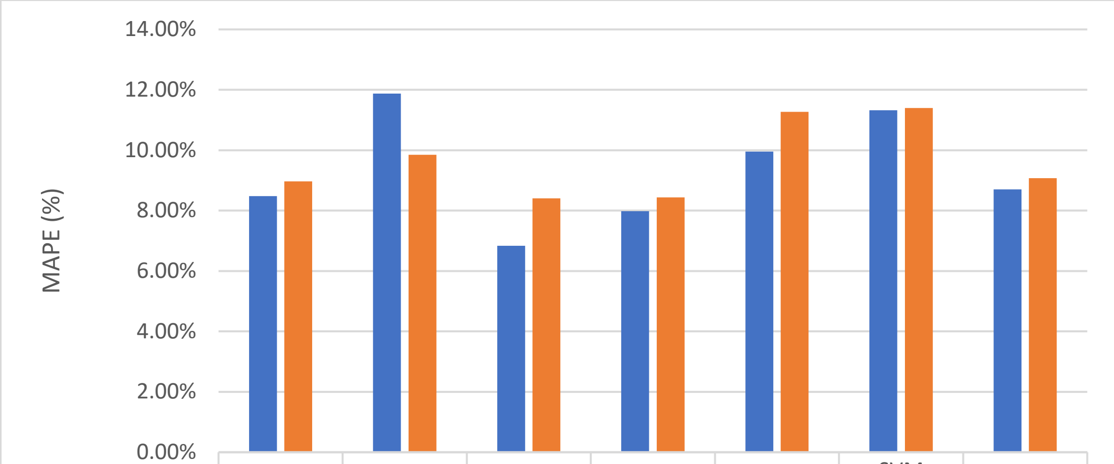
In the first step we use the RFE methodology to select the 25 most informative variables and then use them to train the machine learning algorithms. The 25 variables is a number that strikes a balance between model complexity and performance, reducing overfitting and computational time. We trained models with more than 25 variables but the results provided higher MAPE scores. The selected variables are presented in Table 2.

Table 2: The 25 selected variables based on the RFE technique.

#	Code	Type of Data	Description
1	CC.PER.RNK_lag_1	Political/Institutional	Control of Corruption, Percentile Rank
2	IP.JRN.ARTC.SC_lag_1	Social	Scientific and technical journal articles
3	NE.CON.PRVT.ZS_lag_2	Financial/Economic	Households and Non-Profit Institutions Serving Households (NPISH) final consumption expenditure (% of GDP)
4	NE.CON.TOTL.ZS_lag_2	Financial/Economic	Final consumption expenditure (% of GDP)
5	NE.GDI.FTOT.ZS_lag_3	Financial/Economic	Gross fixed capital formation (% of GDP)
6	NV.IND.MANF.ZS_lag_2	Financial/Economic	Manufacturing, value added (% of GDP)
7	NV.IND.TOTL.ZS_lag_1	Financial/Economic	Industry (including construction), value added (% of GDP)
8	NV.IND.TOTL.ZS_lag_2	Financial/Economic	Industry (including construction), value added (% of GDP)
9	NV.SRV.TOTL.ZS_lag_3	Financial/Economic	Services, value added (% of GDP)
10	NY.GDS.TOTL.ZS_lag_2	Financial/Economic	Gross domestic savings (% of GDP)
11	RQ.PER.RNK_lag_1	Political/Institutional	Regulatory Quality, Percentile Rank

12	sdg_value_lag_1	Financial/Economic	Ratio of government debt outstanding at the end of the year to gross domestic product at current market prices
13	sdg_value_lag_2	Financial/Economic	Ratio of government debt outstanding at the end of the year to gross domestic product at current market prices
14	SL.AGR.EMPL.FE.ZS_lag_1	Social	Employment in agriculture, female (% of female employment) (modeled ILO estimate)
15	SL.EMP.TOTL.SP.FE.NE.ZS_lag_1	Social	Employment to population ratio, 15+, female (%) (national estimate)
16	SL.EMP.TOTL.SP.ZS_lag_1	Social	Employment to population ratio, 15+, total (%) (modeled ILO estimate)
17	SL.EMP.VULN.MA.ZS_lag_1	Social	Vulnerable employment, male (% of male employment) (modeled ILO estimate)
18	SL.UEM.INTM.FE.ZS_lag_1	Social	Unemployment with intermediate education, female (% of female labor force with intermediate education)
19	SP.POP.5054.MA_lag_3	Social	Population ages 50-54, male
20	spread_lag_1	Financial/Economic	10 Year bond spread (Cimadomo et al., 2016, C. A. Favero, 2013, Von Hagen et al., 2011)
21	TX.VAL.MANF.ZS.UN_lag_2	Financial/Economic	Manufactures exports (% of merchandise exports)
22	TX.VAL.MANF.ZS.UN_lag_3	Financial/Economic	Manufactures exports (% of merchandise exports)
23	v2elfrcamp_lag_2	Political/Institutional	Election free campaign media
24	v2elpeace_lag_2	Political/Institutional	Election other electoral violence
25	v2xpe_exlgeo_lag_1	Political/Institutional	Exclusion by Urban-Rural Location index

In the next step, we used a 5-fold cross-validation scheme for the hyperparameters optimization of each model. Moreover, the SVM models were coupled with the linear, the RBF and the polynomial kernels. The MAPE scores for the best trained models are presented in Figure 6.



	Random Forests	Decision Trees	XGBoost	SVM Linear	SVM RBF	SVM Polynomial	Elastic Net
in-sample	8.48%	11.88%	6.83%	7.98%	9.95%	11.33%	8.70%
out-of-sample	8.97%	9.85%	8.41%	8.43%	11.27%	11.40%	9.07%

Figure 6: MAPE scores for the best models of each algorithm in-sample and out-of-sample (OOS).

According to the results, the best model, in terms of in-sample performance, is an XGBoost model, with a MAPE of 6.83% and the second-best model is an SVM model coupled with the linear kernel reaching an in-sample MAPE of 7.98%. In terms of OOS performance, the best model is also the XGBoost reaching an OOS MAPE of 8.41% and the second-best performing is the SVM coupled with the linear kernel with a MAPE of 8.43%. The decision tree model has the worst performance in-sample, while the SVM coupled with the polynomial kernel has the highest MAPE for the OOS part of the dataset. Overall, the best model is the XGBoost (Figure 7). The countries with the highest public debt to GDP ratio (in circles) are Greece and Italy and the forecasts seem to be more inaccurate for the countries with high public debt. This complies with the findings of Bachleitner & Prammer (2024). According to their results the error in the debt forecast is more pronounced for high debt countries than for low debt countries.

For each model, there is very little difference between the out-of-sample MAPE and the in-sample MAPE. This suggests that these models have comparable bias and variance, or in other words, they do not overfit and have a good generalization ability.

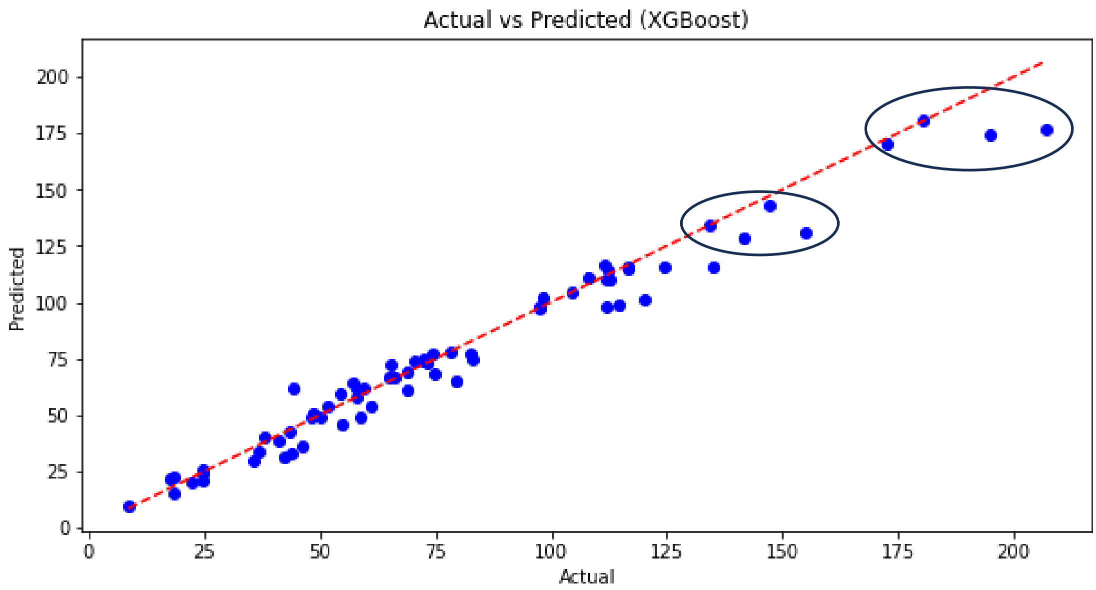


Figure 7: Scatter Plot of the True vs Predicted Values for the out-of-sample part of the dataset public debt to GDP ratio with the optimal XGBoost model. The countries with the highest public debt to GDP ratio (in circles) are Greece and Italy.

To further validate the performance of the XGBoost model, we calculated the Mean Absolute Percentage Error (MAPE) for the years 2021 and 2022[4]. This evaluation compared the predictions of the XGBoost model with the forecasts of two prominent institutions: the European Commission's European Economic Forecast (Spring 2021) and the International Monetary Fund's World Economic Outlook (April 2021: Managing Divergent Recoveries). The MAPE for XGBoost was calculated at 6.49%, significantly lower than the 10.66% for the European Commission's forecasts and the 12.06% for the IMF's forecasts. These results highlight the superior accuracy of the XGBoost model in predicting economic outcomes for these years, demonstrating its robustness and reliability when compared to established forecasting methods from leading global institutions. This performance demonstrates the potential of machine learning models such as XGBoost to improve the accuracy of economic forecasts, even in the face of uncertain and rapidly changing conditions.

4.2. Economic interpretation of the results

In this research, we used the XGBoost's (Extreme Gradient Boosting) VIM (Variable Importance Measure) since this algorithm provided the best results amongst the tree-

[4] The comparison took place only in the years 2021 and 2022 (the most recent ones in our dataset) due to data availability.

based methodologies. The most well-known VIM of XGBoost is the gain, which refers to the improvement in accuracy achieved by the addition of a particular feature to a decision tree. Specifically, it measures the utility of each feature in the construction of the boosted trees during training. The gain for a feature is calculated by comparing the loss reduction achieved when splitting a tree node based on that feature against the distribution of instances reaching that node. The idea is that before adding a new split on a feature X to the branch, there may have been some elements that were incorrectly classified. After adding the split on this feature, there are two new branches, and each of these branches is more accurate. The gain value essentially indicates the extent to which a feature contributes to enhancing the model's performance by facilitating the splitting of data. A higher gain value signifies that a feature is more pivotal for decision-making within the trees, and XGBoost frequently employs this metric to ascertain feature importance and to determine the optimal splits during the tree-building process. The VIM for the 10 most important variables[5], according to the XGBoost gain measure, is presented in Figure 8.

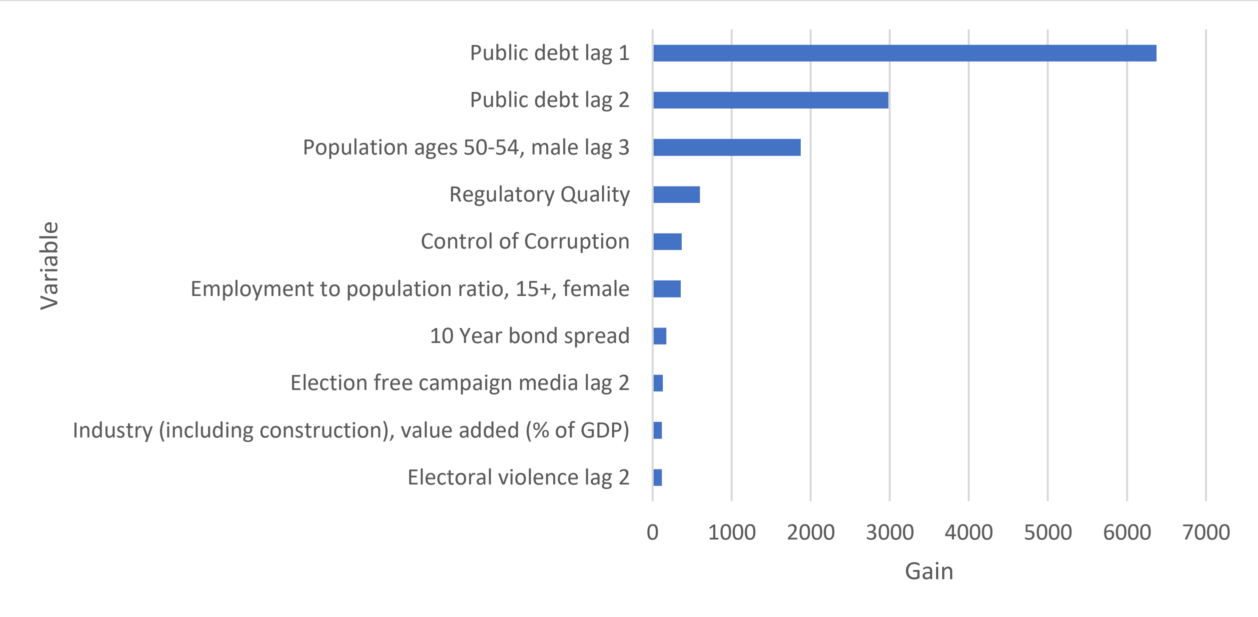


Figure 8: The 10 most important variables calculated with the Variable Importance Measure (VIM) for the best XGBoost model. When there is no lag specified the variable is in lag 1.

The past values of public debt (to GDP ratio) for lags 1 and 2 are the most influential, with the highest VIM scores. The third most important variable is the male population in the ages 50-54 lag 3.

[5] The results for all the variables can be found in Table 4 in the appendix.

The fourth most informative variable is the regulatory quality (Kaufmann et al., 2011). According to the World Bank database "*Regulatory Quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development*". The Regulatory Quality Index consists of the key steps of the law - making process and includes 48 individual indicators, focusing on two main aspects: the form (expression / clarity) and content (substance / scope) of a legal act. The overall score (ranking) is given by an aggregating formula to a 0 to 100 scale. The aim of this toolkit is to identify the main weaknesses of a regulatory environment and to operate as an indicator for comparative analysis across different countries as underlined by Saravakos et al. (2022) . In Figure 9, we can observe the regulatory quality for 4 economies of the Euro area, Germany (DE), Greece, (EL), France (FR) and Italy (IT). We can observe that the two countries with the highest public debt have also a very low rank in the regulatory quality. Thus, the policies and regulations that promote private sector development is crucial for the public debt.

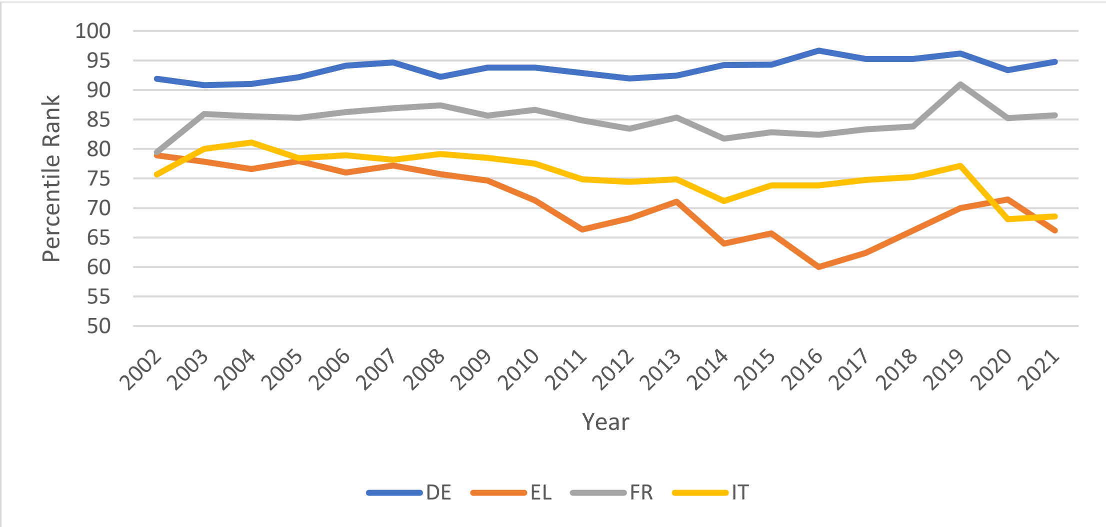


Figure 9: Regulatory Quality, Percentile Rank for 4 economies of the Euro area, Germany (DE), Greece, (EL), France (FR) and Italy (IT).

The fifth most important variable is the control of corruption (Kaufmann et al., 2011). According to the World Bank database "*Control of Corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests*". The measurement is based on a compilation of data from various sources. These sources include opinion surveys and assessments carried out by international institutions, NGOs, think tanks and risk rating agencies. These sources include Transparency International, the World Economic Forum, and local organizations. The World Bank aggregates data from these multiple sources using a statistical methodology called the Unobserved Components Model (UCM). The "Control of Corruption" score (ranking) is given by an aggregating formula to a 0 to 100 scale. In Figure 10 we can observe the control of corruption for 4 economies of the Euro area, Germany (DE), Greece, (EL), France (FR) and Italy (IT). We can observe that the two countries with the highest public debt have also a very low rank in the control of corruption.

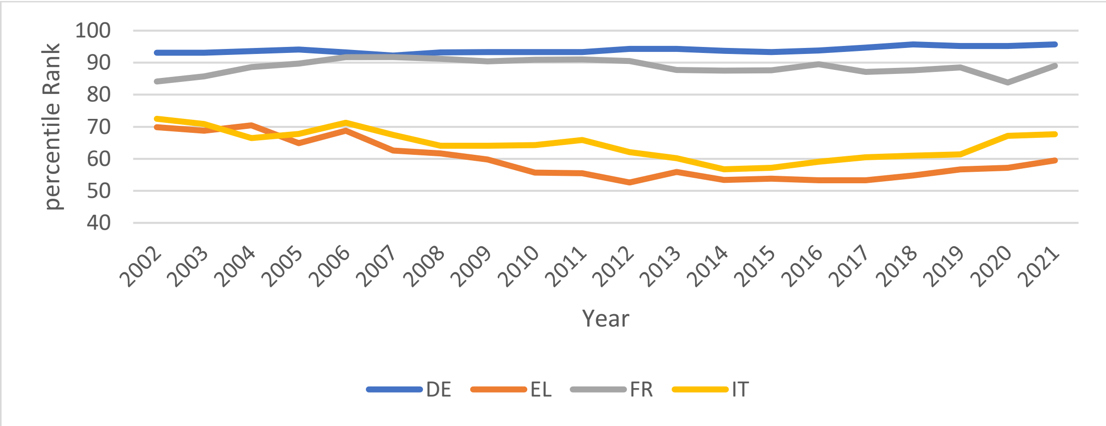


Figure 10: Control of Corruption, Percentile Rank for 4 economies of the Euro area, Germany (DE), Greece, (EL), France (FR) and Italy (IT).

There are other studies that highlight the influence of corruption on public debt, but not for the Euro area. Cooray et al., (2017) found that corruption has a highly statistically significant impact on public debt for 126 countries tested for the period from 1996 to 2012. In addition, Del Monte & Pennacchio, (2020) used a panel of OECD countries for the period from 1995 to 2015 and provided evidence that corruption increases public debt. Also, Liu et al., (2017), for the same time period, examined the relationship between corruption and public debt for 120 countries. They found that higher levels of corruption led to higher levels of public debt. However, a very interesting finding is that the relationship between corruption and debt was non-linear, as in our methodology. Similar results were also found for the United States (Apergis & Apergis, 2019).

The 6th most important variable is the female employment to population ratio for the ages over 15. In the 7th place is the 10 Year bond spread lag 1. The rest of the variables are if parties or candidates receive either free or publicly financed access to national broadcast media industry during national elections (election free campaign media lag 2), the industry (including construction), value added and if during, before and after elections where any violence related to the conduct of the election and the campaigns (electoral violence).

To further explore the relationships between the public debt and the 10 year bond spread, we employed the Partial Dependence Plots (PDP) methodology. A Partial Dependence Plot reveals the marginal effect of the features on the predicted outcome of a machine learning model (Friedman, 2001). A Partial Dependence Plot, shows how the predicted outcome (target variable) changes as a specific feature (independent variable) varies while keeping all other features constant. It provides insights into the relationship between a single feature and the model’s predictions. It is a useful tool as it enhances the interpretability of the trained machine learning model in several ways providing empirical information. The reason we chose spread to further analyze is because it can be directly controlled from the governments to alter the public debt.

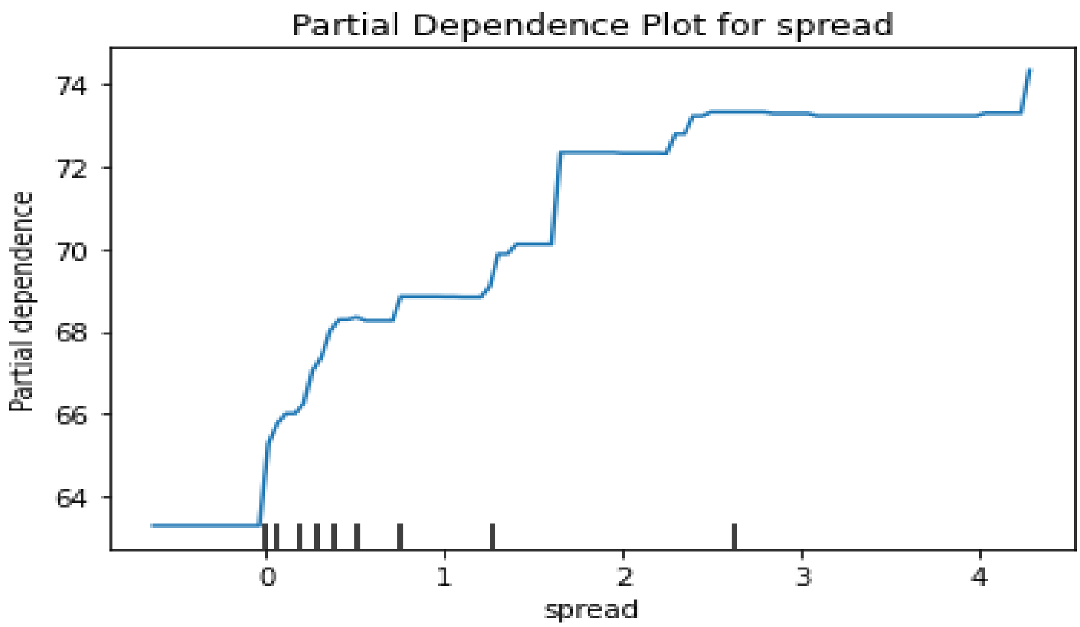


Figure 11: Partial Dependence Plot for spread.

In Figure 11, we present the PDP of the spread and the target variable. We can observe that a) when the spread increases the public debt increases as well b) when the spread is below zero the public debt is very low and increases substantially until the spread reaches a value of 0.5 and c) when the spread is more than 1.8 the public debt increases but in a very low rate.

5. Conclusion

Unpredicted high levels of public debt can have serious economic, financial and social consequences, including higher borrowing costs, reduced investor confidence and fiscal constraints that limit government spending on essential projects. It can also trigger inflationary pressures, financial instability and higher taxes, while reducing public services and employment opportunities, all of which contribute to increased uncertainty and anxiety among citizens.

In this study we attempt to forecast the public debt of Euro area countries with a universal model. We use 4 machine learning methodologies, SVM, decision trees, random forest, XGBoost and an elastic net regression from the area of econometrics. Moreover, we use a broad dataset with 566 economic, financial, political, institutional and social variables, ranging from 2000 to 2022. To the best of our knowledge, this is the first time the public debt of Euro area countries is forecasted with machine learning methodologies and such broad dataset.

The optimal XGBoost model outperformed the competition reaching an out-of-sample MAPE of 8.41% and an in-sample MAPE of 6.83%. In addition, we tested the performance of the XGBoost model by comparing its predictions with those of the European Commission and the IMF, for the years 2021 and 2022. The XGBoost model achieved a MAPE of 6.49%, outperforming the Commission's 10.66% and the IMF's 12.06%. This result highlights the accuracy of the model and the potential of machine learning models such as XGBoost to improve the accuracy of macroeconomic forecasting.

According to the XGBoost VIM, the past values of public debt to GDP ratio are the most influential. The third most important variable is the male population in the ages 50-54 lag 3 followed by the regulatory quality and the control of corruption which capture the perception by economic agents of two major characteristics for the management of public finance: the quality of regulation and the control of corruption, i.e. "*perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development"* and *"perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests".* Among the most influential variables, the 10 year bond spread also plays an important role.

In other words, while standard macroeconomic indicators such as past values of the public debt ratio, the working age population or even 10 year bond spread are among the most influential variables, the public finance management by government also seems to play a crucial role. Countries with the highest public debt have also a very low rank in the regulatory quality. In the same way, countries with the highest public debt have also a very low rank in the control of corruption. It is for this reason that rating agencies must also monitor the way in which economic agents perceive the management of public finance by governments.

Thus, these precise forecasts provide substantial contributions to the financial sector. By identifying macroeconomic variables (the historical values of the public debt to GDP ratio, employment rate, government spread), but also the control of corruption and the regulatory quality as key predictors, financial institutions can more effectively assess the creditworthiness and risk profiles of Euro area countries. This enables more informed lending and investment decisions, potentially lowering the cost of capital. Furthermore, understanding the impact of these indicators on public debt levels allows for better predictions of fiscal sustainability, aiding investors in portfolio diversification and risk management. For policymakers, these insights support the formulation of much more attention to public finance management (captured here by regulatory quality and control of corruption), thereby enhancing economic and financial stability of the Eurozone and thus increase investor confidence.

For future research, extending the forecasting horizon beyond one year could offer valuable insights, enabling a more comprehensive understanding of long-term trends and potential outcomes. Such an approach may also improve the robustness and applicability of the forecasting model across diverse time frames.

Funding: Emmanouil Sofianos would like to acknowledge that this work is part of the Interdisciplinary Thematic Institute MAKErS of the ITI 2021-2028 program of the University of Strasbourg, CNRS and INSERM. It has received financial support from the IdEx Unistra (ANR-10-IDEX-0002), and from the Programme Investissement d'Avenir as part of the SFRI-STRAT'US project(s) (ANR-20-SFRI-0012).

References

- Apergis, E. and Apergis, N. (2019) ‘New evidence on corruption and government debt from a Global Country Panel’, *Journal of Economic Studies*, 46(5), pp. 1009–1027.
- Bachleitner, A. and Prammer, D. (2024) ‘Don’t blame the government!? an assessment of debt forecast errors with a view to the EU Economic Governance Review’, *European Journal of Political Economy*, 82, p. 102524.
- Belly, G., Boeckelmann, L., Caicedo Graciar, C., Di lorio, A., Istrefi, K., Siakoulis, V., Stalla-Bourdillor, A. (2023) ‘Forecasting sovereign risk in the Euro Area via machine learning’, *Journal of Forecasting*, 42(3), pp. 657–684.
- Breiman, L. (1996) ‘Bagging predictors’, *Machine Learning*, 24(2), pp. 123–140.
- Breiman, L. (2001)a. Random forests. *Machine Learning*, 45 pp.5-32
- Chen T., Guestrin C. (2016) XGBoost: ‘A Scalable Tree Boosting System’, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16).* Association for Computing Machinery, New York, NY, USA, 785–794.
- Chen, C., Yao, S., Hu, P. and Lin, Y. (2017) ‘Optimal Government Investment and public debt in an economic growth model’, *China Economic Review*, 45, pp. 257–278.
- Cimadomo, J., Claeys, P. and Poplawski-Ribeiro, M. (2016) ‘How do experts forecast sovereign spreads?’, *European Economic Review*, 87, pp. 216–235.
- Cooray, A., Dzhumashev, R. and Schneider, F. (2017) ‘How does corruption affect public debt? an empirical analysis’, *World Development*, 90, pp. 115–127.
- Cortes, C.; Vapnik, V. (1995) Support-vector networks. *Machine Learn*ing, 20, pp. 273–297.
- D’Erasmo, P. and Mendoza, E. (2018) *History remembered: Optimal Sovereign Default on domestic and external debt* [Preprint].

- Dawood, M., Horsewood, N. and Strobel, F. (2017) ‘Predicting sovereign debt crises: An early warning system approach’, *Journal of Financial Stability*, 28, pp. 16–28.
- Del Monte, A. and Pennacchio, L. (2020) ‘Corruption, government expenditure and public debt in OECD countries’, *Comparative Economic Studies*, 62, pp. 739–771.
- Di Bartolomeo, G., Di Petro, M., Saltari, E. and Semmler, W. (2018) ‘Public debt stabilization: The relevance of policymakers’ time horizons’, *Public Choice*, 177(3–4), pp. 287–299.
- Estefania-Flores, J., Furceri, D., Kothari, S. and D. Ostry, J. (2023) ‘Worse than you think: Public debt forecast errors in advanced and developing economies’, *Journal of Forecasting*, 42(3), pp. 685–714.
- Eusepi, G. and Wagner, R.E. (2017) *Public debt: An illusion of democratic political economy*. Cheltenham, UK: Edward Elgar Publishing. Fioramanti, M. (2008) ‘Predicting sovereign debt crises using Artificial Neural Networks: A comparative approach’, *Journal of Financial Stability*, 4(2), pp. 149–164.
- Favero, C.A. (2013) ‘Modelling and forecasting government bond spreads in the Euro Area: A GVAR model’, *Journal of Econometrics*, 177(2), pp. 343–356.
- Freund, Y. and Schapire, R.E. (1997) ‘A decision-theoretic generalization of on-line learning and an application to boosting’, *Journal of Computer and System Sciences*, 55(1), pp. 119–139.
- Friedman, Jerome H. (2001) “Greedy function approximation: A gradient boosting machine.” *Annals of statistics*, 29(5), 1189-1232.
- Gogas, P. and Papadimitriou, T. (2021) ‘Machine Learning in Economics and Finance’, *Computational Economics*, 57(1), pp. 1–4.
- Gogas, P., Papadimitriou, T. and Sofianos, E. (2019) ‘Money neutrality, monetary aggregates and machine learning’, *Algorithms*, 12(7), p. 137.
- Gogas, P., Papadimitriou, T. and Sofianos, E. (2022) ‘Forecasting unemployment in the Euro area with machine learning’, *Journal of Forecasting*, 41(3), pp. 551–566.

- Kalemli-Özcan, Şebnem, Reinhart, C. and Rogoff, K. (2016) ‘Sovereign debt and financial crises: Theory and historical evidence’, *Journal of the European Economic Association*, 14(1), pp. 1–6.

- Kaufmann, D., Kraay, A. and Mastruzzi, M. (2011) ‘The worldwide governance indicators: Methodology and Analytical Issues’, *Hague Journal on the Rule of Law*, 3(02), pp. 220–246.

- Knapková, M., Kiaba, M. and Hudec, S. (2019) ‘Impact of macroeconomic indicators on public debt of Slovak Republic’, *Journal of Business Economics and Management*, 20(4), pp. 734–753.

- Liu, C., Moldogaziev, T.T. and Mikesell, J.L. (2017) ‘Corruption and state and local government debt expansion’, *Public Administration Review*, 77(5), pp. 681–690.

- Ono, T. and Uchida, Y. (2018) ‘Human capital, public debt, and economic growth: A political economy analysis’, *Journal of Macroeconomics*, 57, pp. 1–14.

- Page, A. (2018) ‘War, public debt and Richard Price’s rational dissenting radicalism’, *Historical Research*, 91(251), pp. 98–115.

- Poghosyan, T. (2017) ‘How do financial cycles affect public debt cycles?’, *Empirical Economics*, 54(2), pp. 425–460.

- Saravakos, C., Karkatsoulis, P., Stefopoulou, E., Archontas, G., (2022), ‘Regulatory Quality Index: a process-oriented approach to measure the quality of primary legislation in a country’, *SSRN Electronic Journal [Preprint]*.

- Sermpinis, G., Stasinakis, C., Theofilatos, K., & Karathanasopoulos, A. (2014). Inflation and unemployment forecasting with genetic support vector regression. *Journal of Forecasting*, 33(6), 471–487.

- Sofianos, E., Zaganidis, E., Papadimitriou, T. and Gogas, P. (2024) ‘Forecasting East and West Coast gasoline prices with tree-based machine learning algorithms’, *Energies*, 17(6), p. 1296.

- Sofianos, E., Gogas, P. and Papadimitriou, T. (2021) ‘Mind the gap: Forecasting Euro-area output gaps with machine learning’, *Applied Economics Letters*, 29(19), pp. 1824–1828.
- Solanes, A. and Radua, J. (2022). Advances in using MRI to estimate the risk of future outcomes in mental health - are we getting there? *Frontiers in Psychiatry*, 13.
- Stasavage, D. (2003) *Public debt and the birth of the Democratic state France and Great Britain, 1688-1789*. Cambridge: Cambridge University Press.
- von Hagen, J., Schuknecht, L. and Wolswijk, G. (2011) ‘Government bond risk premiums in the EU revisited: The impact of the financial crisis’, *European Journal of Political Economy*, 27(1), pp. 36–43.
- Vuckovic, V. and Basarac Sertic, M. (2013) ‘The effect of political institutions on the size of government spending in European Union member states and Croatia’, *Financial Theory and Practice*, 37(2), pp. 161–179.

Appendix

Table 3: List of all variables used with their source and short description.

#	Code	Source	Description
1	country_encoded	-	Categorical variable for countries
2	year	-	The year for each observation
3	sdg_value	Eurostat	Ratio of government debt outstanding at the end of the year to gross domestic product at current market prices
4	gov_value	Eurostat	Government deficit/surplus
5	NY.GDP.MKTP.KD.ZG	WBOD	GDP growth (annual %)
6	FP.CPI.TOTL.ZG	WBOD	Inflation, consumer prices (annual %)
7	SL.UEM.TOTL.ZS	WBOD	Unemployment, total (% of total labor force) (modeled ILO estimate)
8	NE.EXP.GNFS.CD	WBOD	Exports of goods and services (current US$)
9	NE.IMP.GNFS.CD	WBOD	Imports of goods and services (current US$)
10	NE.GDI.FTOT.CD	WBOD	Gross fixed capital formation (current US$)
11	EG.ELC.ACCS.ZS	WBOD	Access to electricity (% of population)
12	SP.DYN.IMRT.IN	WBOD	Mortality rate, infant (per 1,000 live births)
13	SP.DYN.LE00.IN	WBOD	Life expectancy at birth, total (years)
14	SM.POP.NETM	WBOD	Net migration
15	NE.RSB.GNFS.ZS	WBOD	External balance on goods and services (% of GDP)
16	NE.CON.TOTL.ZS	WBOD	Final consumption expenditure (% of GDP)
17	BX.KLT.DINV.WD.GD.ZS	WBOD	Foreign direct investment, net inflows (% of GDP)
18	BM.KLT.DINV.WD.GD.ZS	WBOD	Foreign direct investment, net outflows (% of GDP)
19	NE.CON.GOVT.KD.ZG	WBOD	General government final consumption expenditure (annual % growth)
20	NY.GNP.MKTP.CD	WBOD	Gross National Income, GNI (current US$)
21	NE.GDI.TOTL.ZS	WBOD	Gross capital formation (% of GDP)
22	NE.GDI.FTOT.ZS	WBOD	Gross fixed capital formation (% of GDP)
23	NE.DAB.TOTL.ZS	WBOD	Gross national expenditure (% of GDP)
24	NY.GNS.ICTR.ZS	WBOD	Gross savings (% of GDP)
25	NE.CON.PRVT.ZS	WBOD	Households and Non-Profit Institutions Serving Households (NPISH) final consumption expenditure (% of GDP)
26	NE.IMP.GNFS.ZS	WBOD	Imports of goods and services (% of GDP)
27	NV.IND.TOTL.ZS	WBOD	Industry (including construction), value added (% of GDP)
28	NV.IND.MANF.ZS	WBOD	Manufacturing, value added (% of GDP)
29	BG.GSR.NFSV.GD.ZS	WBOD	Trade in services (% of GDP)
30	NE.TRD.GNFS.ZS	WBOD	Trade (% of GDP)
31	NV.SRV.TOTL.ZS	WBOD	Services, value added (% of GDP)
32	BX.TRF.PWKR.DT.GD.ZS	WBOD	Personal remittances, received (% of GDP)
33	NV.IND.MANF.KD.ZG	WBOD	Manufacturing, value added (annual % growth)

34	NE.IMP.GNFS.KD.ZG	WBOD	Imports of goods and services (annual % growth)
35	NY.GDS.TOTL.ZS	WBOD	Gross domestic savings (% of GDP)
36	NE.GDI.FTOT.KD.ZG	WBOD	Gross fixed capital formation (annual % growth)
37	NE.GDI.TOTL.KD.ZG	WBOD	Gross capital formation (annual % growth)
38	NE.CON.GOVT.ZS	WBOD	General government final consumption expenditure (% of GDP)
39	NE.CON.TOTL.KD.ZG	WBOD	Final consumption expenditure (annual % growth)
40	NE.EXP.GNFS.KD.ZG	WBOD	Exports of goods and services (annual % growth)
41	FP.CPI.TOTL	WBOD	Consumer price index (2010 = 100)
42	NY.GDP.DEFL.KD.ZG	WBOD	Inflation, GDP deflator (annual %)
43	NY.GDP.PCAP.KD.ZG	WBOD	GDP per capita growth (annual %)
44	TM.VAL.MANF.ZS.UN	WBOD	Manufactures imports (% of merchandise imports)
45	TX.VAL.MANF.ZS.UN	WBOD	Manufactures exports (% of merchandise exports)
46	TG.VAL.TOTL.GD.ZS	WBOD	Merchandise trade (% of GDP)
47	GC.XPN.COMP.ZS	WBOD	Compensation of employees (% of expense)
48	CC.PER.RNK	WBOD	Control of Corruption, Percentile Rank
49	GC.XPN.TOTL.GD.ZS	WBOD	Expense (% of GDP)
50	GC.XPN.GSRV.ZS	WBOD	Goods and services expense (% of expense)
51	GC.NLD.TOTL.GD.ZS	WBOD	Net lending (+) / net borrowing (-) (% of GDP)
52	GC.XPN.OTHR.ZS	WBOD	Other expense (% of expense)
53	RQ.PER.RNK	WBOD	Regulatory Quality
54	RL.PER.RNK	WBOD	Rule of Law, Percentile Rank
55	GC.REV.SOCL.ZS	WBOD	Social contributions (% of revenue)
56	GC.TAX.TOTL.GD.ZS	WBOD	Tax revenue (% of GDP)
57	GC.TAX.GSRV.RV.ZS	WBOD	Taxes on goods and services (% of revenue)
58	GC.TAX.YPKG.ZS	WBOD	Taxes on income, profits and capital gains (% of total taxes)
59	SL.UEM.TOTL.NE.ZS	WBOD	Unemployment, total (% of total labor force) (national estimate)
60	BN.CAB.XOKA.GD.ZS	WBOD	Current account balance (% of GDP)
61	NE.EXP.GNFS.ZS	WBOD	Exports of goods and services (% of GDP)
62	NV.IND.MANF.CD	WBOD	Manufacturing, value added (current US$)
63	EN.POP.DNST	WBOD	Population density (people per sq. km of land area)
64	SP.URB.TOTL.IN.ZS	WBOD	Urban population (% of total population)
65	SE.XPD.TOTL.GB.ZS	WBOD	Government expenditure on education, total (% of government expenditure)
66	SE.XPD.TOTL.GD.ZS	WBOD	Government expenditure on education, total (% of GDP)
67	EG.FEC.RNEW.ZS	WBOD	Renewable energy consumption (% of total final energy consumption)
68	IP.JRN.ARTC.SC	WBOD	Scientific and technical journal articles
69	NY.ADJ.AEDU.GN.ZS	WBOD	Adjusted savings
70	EN.ATM.CO2E.PC	WBOD	CO2 emissions (metric tons per capita)

71	GB.XPD.RSDV.GD.ZS	WBOD	Research and development expenditure (% of GDP)
72	SE.PRM.ENRR	WBOD	School enrollment, primary (% gross)
73	SE.SEC.ENRR	WBOD	School enrollment, secondary (% gross)
74	NY.GDP.MKTP.CD	WBOD	GDP (current US$)
75	NY.GDP.PCAP.CD	WBOD	GDP per capita (current US$)
76	NY.GNP.ATLS.CD	WBOD	Gross national Income (GNI), Atlas method (current US$)
77	NY.GNP.PCAP.CD	WBOD	Gross national Income (GNI) per capita, Atlas method (current US$)
78	SG.AGE.FUPN.EQ	WBGD	The age at which men and women can retire with full pension benefits is the same (1=yes; 0=no)
79	SG.AGE.MRET.EQ	WBGD	The mandatory retirement age for men and women is the same (1=yes; 0=no)
80	SG.AGE.PAPN.EQ	WBGD	The age at which men and women can retire with partial pension benefits is the same (1=yes; 0=no)
81	SG.APL.PSPT.EQ	WBGD	A woman can apply for a passport in the same way as a man (1=yes; 0=no)
82	SG.BUS.REGT.EQ	WBGD	A woman can register a business in the same way as a man (1=yes; 0=no)
83	SG.CNT.SIGN.EQ	WBGD	A woman can sign a contract in the same way as a man (1=yes; 0=no)
84	SG.CTR.TRVL.EQ	WBGD	A woman can travel outside the country in the same way as a man (1=yes; 0=no)
85	SG.DML.PRGW	WBGD	Dismissal of pregnant workers is prohibited (1=yes; 0=no)
86	SG.DNG.WORK.DN.EQ	WBGD	A woman can work in a job deemed dangerous in the same way as a man (1=yes; 0=no)
87	SG.GET.JOBS.EQ	WBGD	A woman can get a job in the same way as a man (1=yes; 0=no)
88	SG.HLD.HEAD.EQ	WBGD	A woman can be head of household in the same way as a man (1=yes; 0=no)
89	SG.HME.TRVL.EQ	WBGD	A woman can travel outside her home in the same way as a man (1=yes; 0=no)
90	SG.IHT.ASST.EQ	WBGD	Male and female surviving spouses have equal rights to inherit assets (1=yes; 0=no)
91	SG.IHT.ASST.PT.EQ	WBGD	Sons and daughters have equal rights to inherit assets from their parents (1=yes; 0=no)
92	SG.IND.WORK.EQ	WBGD	A woman can work in an industrial job in the same way as a man (1=yes; 0=no)
93	SG.LAW.ASST.AR	WBGD	The law grants spouses equal administrative authority over assets during marriage (1=yes; 0=no)
94	SG.LAW.CRDD.GR	WBGD	The law prohibits discrimination in access to credit based on gender (1=yes; 0=no)

95	SG.LAW.EQRM.WK	WBGD	Law mandates equal remuneration for females and males for work of equal value (1=yes; 0=no)
96	SG.LAW.INDX	WBGD	Women Business and the Law Index Score (scale 1-100)
97	SG.LAW.INDX.AS	WBGD	Women, Business and the Law: Assets Indicator Score (scale 1-100)
98	SG.LAW.INDX.EN	WBGD	Women, Business and the Law: Entrepreneurship Indicator Score (scale 1-100)
99	SG.LAW.INDX.MO	WBGD	Women, Business and the Law: Mobility Indicator Score (scale 1-100)
100	SG.LAW.INDX.MR	WBGD	Women, Business and the Law: Marriage Indicator Score (scale 1-100)
101	SG.LAW.INDX.PE	WBGD	Women, Business and the Law: Pension Indicator Score (scale 1-100)
102	SG.LAW.INDX.PR	WBGD	Women, Business and the Law: Parenthood Indicator Score (scale 1-100)
103	SG.LAW.INDX.PY	WBGD	Women, Business and the Law: Pay Indicator Score (scale 1-100)
104	SG.LAW.INDX.WP	WBGD	Women, Business and the Law: Workplace Indicator Score (scale 1-100)
105	SG.LAW.NMCN	WBGD	The law provides for the valuation of nonmonetary contributions (1=yes; 0=no)
106	SG.LAW.NODC.HR	WBGD	The law prohibits discrimination in employment based on gender (1=yes; 0=no)
107	SG.LAW.OBHB.MR.NO	WBGD	The law is free of legal provisions that require a married woman to obey her husband (1=yes; 0=no)
108	SG.LEG.SXHR.EM	WBGD	There is legislation on sexual harassment in employment (1=yes; 0=no)
109	SG.LOC.LIVE.EQ	WBGD	A woman can choose where to live in the same way as a man (1=yes; 0=no)
110	SG.NGT.WORK.EQ	WBGD	A woman can work at night in the same way as a man (1=yes; 0=no)
111	SG.OBT.DVRC.EQ	WBGD	A woman can obtain a judgment of divorce in the same way as a man (1=yes; 0=no)
112	SG.OPN.BANK.EQ	WBGD	A woman can open a bank account in the same way as a man (1=yes; 0=no)
113	SG.OWN.PRRT.IM	WBGD	Women and men have equal ownership rights to immovable property (1=yes; 0=no)
114	SG.PEN.SXHR.EM	WBGD	Criminal penalties or civil remedies exist for sexual harassment in employment (1=yes; 0=no)
115	SG.REM.RIGT.EQ	WBGD	A woman has the same rights to remarry as a man (1=yes; 0=no)
116	SH.MMR.LEVE	WBGD	Length of paid maternity leave (calendar days)
117	SH.MMR.LEVE.AL	WBGD	Paid leave of at least 14 weeks available to mothers (1=yes; 0=no)
118	SH.PAR.LEVE.AL	WBGD	There is paid parental leave (1=yes; 0=no)

119	SH.PAR.LEVE.FE	WBGD	Length of paid parental leave for mother (calendar days)
120	SH.PAR.LEVE.MA	WBGD	Length of paid parental leave for father (calendar days)
121	SH.PTR.LEVE.AL	WBGD	Paid leave is available to fathers (1=yes; 0=no)
122	SL.EMP.VULN.FE.ZS	WBGD	Vulnerable employment, female (% of female employment) (modeled ILO estimate)
123	SL.EMP.VULN.MA.ZS	WBGD	Vulnerable employment, male (% of male employment) (modeled ILO estimate)
124	SL.EMP.VULN.ZS	WBGD	Vulnerable employment, total (% of total employment) (modeled ILO estimate)
125	SL.TLF.CACT.FM.NE.ZS	WBGD	Ratio of female to male labor force participation rate (%) (national estimate)
126	SL.TLF.CACT.FM.ZS	WBGD	Ratio of female to male labor force participation rate (%) (modeled ILO estimate)
127	SL.TLF.TOTL.FE.ZS	WBGD	Labor force, female (% of total labor force)
128	SL.TLF.TOTL.IN	WBGD	Labor force, total
129	SL.UEM.1524.FM.NE.ZS	WBGD	Ratio of female to male youth unemployment rate (% ages 15-24) (national estimate)
130	SL.UEM.1524.FM.ZS	WBGD	Ratio of female to male youth unemployment rate (% ages 15-24) (modeled ILO estimate)
131	SP.POP.0014.FE.IN	WBGD	Population ages 0-14, female
132	SP.POP.0014.MA.IN	WBGD	Population ages 0-14, male
133	SP.POP.0014.TO	WBGD	Population ages 0-14, total
134	SP.POP.0509.FE	WBGD	Population ages 05-09, female
135	SP.POP.0509.MA	WBGD	Population ages 05-09, male
136	SP.POP.1014.FE	WBGD	Population ages 10-14, female
137	SP.POP.1014.MA	WBGD	Population ages 10-14, male
138	SP.POP.1519.FE	WBGD	Population ages 15-19, female
139	SP.POP.1519.MA	WBGD	Population ages 15-19, male
140	SP.POP.1564.FE.IN	WBGD	Population ages 15-64, female
141	SP.POP.1564.MA.IN	WBGD	Population ages 15-64, male
142	SP.POP.1564.TO	WBGD	Population ages 15-64, total
143	SP.POP.2024.FE	WBGD	Population ages 20-24, female
144	SP.POP.2024.MA	WBGD	Population ages 20-24, male
145	SP.POP.2529.FE	WBGD	Population ages 25-29, female
146	SP.POP.2529.MA	WBGD	Population ages 25-29, male
147	SP.POP.3034.FE	WBGD	Population ages 30-34, female
148	SP.POP.3034.MA	WBGD	Population ages 30-34, male
149	SP.POP.3539.FE	WBGD	Population ages 35-39, female
150	SP.POP.3539.MA	WBGD	Population ages 35-39, male
151	SP.POP.4044.FE	WBGD	Population ages 40-44, female
152	SP.POP.4044.MA	WBGD	Population ages 40-44, male
153	SP.POP.4549.FE	WBGD	Population ages 45-49, female
154	SP.POP.4549.MA	WBGD	Population ages 45-49, male
155	SP.POP.5054.FE	WBGD	Population ages 50-54, female
156	SP.POP.5054.MA	WBGD	Population ages 50-54, male
157	SP.POP.5559.FE	WBGD	Population ages 55-59, female

158	SP.POP.5559.MA	WBGD	Population ages 55-59, male
159	SP.POP.6064.FE	WBGD	Population ages 60-64, female
160	SP.POP.6064.MA	WBGD	Population ages 60-64, male
161	SP.POP.6569.FE	WBGD	Population ages 65-69, female
162	SP.POP.6569.MA	WBGD	Population ages 65-69, male
163	SP.POP.65UP.FE.IN	WBGD	Population ages 65 and above, female
164	SP.POP.65UP.MA.IN	WBGD	Population ages 65 and above, male
165	SP.POP.65UP.TO.ZS	WBGD	Population ages 65 and above (% of total population)
166	SP.POP.7074.FE	WBGD	Population ages 70-74, female
167	SP.POP.7074.MA	WBGD	Population ages 70-74, male
168	SP.POP.7579.FE	WBGD	Population ages 75-79, female
169	SP.POP.7579.MA	WBGD	Population ages 75-79, male
170	SP.POP.80UP.FE	WBGD	Population ages 80 and above, female
171	SP.POP.80UP.MA	WBGD	Population ages 80 and above, male
172	SP.POP.DPND	WBGD	Age dependency ratio (% of working-age population)
173	SP.POP.TOTL.FE.IN	WBGD	Population, female
174	SP.POP.TOTL.FE.ZS	WBGD	Population, female (% of total population)
175	SP.POP.TOTL.MA.IN	WBGD	Population, male
176	SP.RUR.TOTL.ZS	WBGD	Rural population (% of total population)
177	SE.PRM.ENRR.FE	WBGD	School enrollment, primary, female (% gross)
178	SE.PRM.ENRR.MA	WBGD	School enrollment, primary, male (% gross)
179	SE.SEC.ENRR.FE	WBGD	School enrollment, secondary, female (% gross)
180	SE.SEC.ENRR.MA	WBGD	School enrollment, secondary, male (% gross)
181	NY.GNP.PCAP.PP.CD	WBGD	Gross National Income (GNI) per capita, Purchasing Power Parity (PPP) (current international $)
182	SG.GEN.PARL.ZS	WBGD	Proportion of seats held by women in national parliaments (%)
183	SH.DTH.IMRT	WBGD	Number of infant deaths
184	SH.DTH.IMRT.FE	WBGD	Number of infant deaths, female
185	SH.DTH.IMRT.MA	WBGD	Number of infant deaths, male
186	SH.DTH.MORT.FE	WBGD	Number of under-five deaths, female
187	SH.DTH.MORT.MA	WBGD	Number of under-five deaths, male
188	SH.DTH.STLB	WBGD	Number of stillbirths
189	SH.DYN.AIDS.FE.ZS	WBGD	Womens share of population ages 15+ living with HIV (%)
190	SH.DYN.MORT	WBGD	Mortality rate, under-5 (per 1,000 live births)
191	SH.DYN.MORT.FE	WBGD	Mortality rate, under-5, female (per 1,000 live births)
192	SH.DYN.MORT.MA	WBGD	Mortality rate, under-5, male (per 1,000 live births)
193	SH.DYN.STLB	WBGD	Stillbirth rate (per 1,000 total births)
194	SH.IMM.IDPT	WBGD	Immunization, Diphtheria, Tetanus, Pertussis (DPT) (% of children ages 12-23 months)
195	SH.IMM.MEAS	WBGD	Immunization, measles (% of children ages 12-23 months)

196	SL.AGR.EMPL.FE.ZS	WBGD	Employment in agriculture, female (% of female employment) (modeled ILO estimate)
197	SL.AGR.EMPL.MA.ZS	WBGD	Employment in agriculture, female (% of male employment) (modeled ILO estimate)
198	SL.AGR.EMPL.ZS	WBGD	Employment in agriculture (% of total employment) (modeled ILO estimate)
199	SL.EMP.1524.SP.FE.NE.ZS	WBGD	Employment to population ratio, ages 15-24, female (%) (national estimate)
200	SL.EMP.1524.SP.FE.ZS	WBGD	Employment to population ratio, ages 15-24, female (%) (modeled ILO estimate)
201	SL.EMP.1524.SP.MA.NE.ZS	WBGD	Employment to population ratio, ages 15-24, male (%) (national estimate)
202	SL.EMP.1524.SP.MA.ZS	WBGD	Employment to population ratio, ages 15-24, male (%) (modeled ILO estimate)
203	SL.EMP.1524.SP.NE.ZS	WBGD	Employment to population ratio, ages 15-24, total (%) (national estimate)
204	SL.EMP.1524.SP.ZS	WBGD	Employment to population ratio, ages 15-24, total (%) (modeled ILO estimate)
205	SL.EMP.MPYR.FE.ZS	WBGD	Employers, female (% of female employment) (modeled ILO estimate)
206	SL.EMP.MPYR.MA.ZS	WBGD	Employers, male (% of male employment) (modeled ILO estimate)
207	SL.EMP.MPYR.ZS	WBGD	Employers, total (% of total employment) (modeled ILO estimate)
208	SL.EMP.SELF.FE.ZS	WBGD	Self-employed, female (% of female employment) (modeled ILO estimate)
209	SL.EMP.SELF.MA.ZS	WBGD	Self-employed, male (% of male employment) (modeled ILO estimate)
210	SL.EMP.SELF.ZS	WBGD	Self-employed, total (% of total employment) (modeled ILO estimate)
211	SL.EMP.TOTL.SP.FE.NE.ZS	WBGD	Employment to population ratio, 15+, female (%) (national estimate)
212	SL.EMP.TOTL.SP.FE.ZS	WBGD	Employment to population ratio, 15+, female (%) (modeled ILO estimate)
213	SL.EMP.TOTL.SP.MA.NE.ZS	WBGD	Employment to population ratio, 15+, male (%) (national estimate)
214	SL.EMP.TOTL.SP.ZS	WBGD	Employment to population ratio, 15+, total (%) (modeled ILO estimate)
215	SL.EMP.WORK.FE.ZS	WBGD	Wage and salaried workers, female (% of female employment) (modeled ILO estimate)
216	SL.EMP.WORK.MA.ZS	WBGD	Wage and salaried workers, male (% of female employment) (modeled ILO estimate)
217	SL.EMP.WORK.ZS	WBGD	Wage and salaried workers, total (% of total employment) (modeled ILO estimate)
218	SL.IND.EMPL.FE.ZS	WBGD	Employment in industry, female (% of female employment) (modeled ILO estimate)
219	SL.IND.EMPL.MA.ZS	WBGD	Employment in industry, male (% of male employment) (modeled ILO estimate)
220	SL.IND.EMPL.ZS	WBGD	Employment in industry (% of total employment) (modeled ILO estimate)

221	SL.SRV.EMPL.FE.ZS	WBGD	Employment in services, female (% of female employment) (modeled ILO estimate)
222	SL.SRV.EMPL.MA.ZS	WBGD	Employment in services, male (% of male employment) (modeled ILO estimate)
223	SL.SRV.EMPL.ZS	WBGD	Employment in services (% of total employment) (modeled ILO estimate)
224	SL.TLF.ACTI.1524.FE.NE.ZS	WBGD	Labor force participation rate for ages 15-24, female (%) (national estimate)
225	SL.TLF.ACTI.1524.FE.ZS	WBGD	Labor force participation rate for ages 15-24, female (%) (modeled ILO estimate)
226	SL.TLF.ACTI.1524.MA.NE.ZS	WBGD	Labor force participation rate for ages 15-24, male (%) (national estimate)
227	SL.TLF.ACTI.1524.MA.ZS	WBGD	Labor force participation rate for ages 15-24, male (%) (modeled ILO estimate)
228	SL.TLF.ACTI.1524.NE.ZS	WBGD	Labor force participation rate for ages 15-24, total (%) (national estimate)
229	SL.TLF.ACTI.1524.ZS	WBGD	Labor force participation rate for ages 15-24, total (%) (modeled ILO estimate)
230	SL.TLF.ACTI.FE.ZS	WBGD	Labor force participation rate, female (% of female population ages 15-64) (modeled ILO estimate)
231	SL.TLF.ACTI.MA.ZS	WBGD	Labor force participation rate, male (% of male population ages 15-64) (modeled ILO estimate)
232	SL.TLF.ACTI.ZS	WBGD	Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)
233	SL.TLF.ADVN.FE.ZS	WBGD	Labor force with advanced education, female (% of female working-age population with advanced education)
234	SL.TLF.ADVN.MA.ZS	WBGD	Labor force with advanced education, male (% of male working-age population with advanced education)
235	SL.TLF.ADVN.ZS	WBGD	Labor force with advanced education (% of total working-age population with advanced education)
236	SL.TLF.BASC.FE.ZS	WBGD	Labor force with basic education, female (% of female working-age population with basic education)
237	SL.TLF.BASC.MA.ZS	WBGD	Labor force with basic education, male (% of male working-age population with basic education)
238	SL.TLF.BASC.ZS	WBGD	Labor force with basic education (% of total working-age population with basic education)
239	SL.TLF.CACT.FE.NE.ZS	WBGD	Labor force participation rate, female (% of female population ages 15+) (national estimate)
240	SL.TLF.CACT.FE.ZS	WBGD	Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)

241	SL.TLF.CACT.MA.NE.ZS	WBGD	Labor force participation rate, male (% of male population ages 15+) (national estimate)
242	SL.TLF.CACT.MA.ZS	WBGD	Labor force participation rate, male (% of male population ages 15+) (modeled ILO estimate)
243	SL.TLF.CACT.NE.ZS	WBGD	Labor force participation rate, total (% of total population ages 15+) (national estimate)
244	SL.TLF.CACT.ZS	WBGD	Labor force participation rate, total (% of total population ages 15+) (modeled ILO estimate)
245	SL.TLF.INTM.FE.ZS	WBGD	Labor force with intermediate education, female (% of female working-age population with intermediate education)
246	SL.TLF.INTM.MA.ZS	WBGD	Labor force with intermediate education, male (% of male working-age population with intermediate education)
247	SL.TLF.INTM.ZS	WBGD	Labor force with intermediate education (% of total working-age population with intermediate education)
248	SL.TLF.PART.ZS	WBGD	Part time employment, total (% of total employment)
249	SL.UEM.1524.FE.NE.ZS	WBGD	Unemployment, youth female (% of female labor force ages 15-24) (national estimate)
250	SL.UEM.1524.MA.NE.ZS	WBGD	Unemployment, youth male (% of male labor force ages 15-24) (national estimate)
251	SL.UEM.1524.MA.ZS	WBGD	Unemployment, youth male (% of male labor force ages 15-24) (modeled ILO estimate)
252	SL.UEM.1524.NE.ZS	WBGD	Unemployment, youth total (% of total labor force ages 15-24) (national estimate)
253	SL.UEM.1524.ZS	WBGD	Unemployment, youth total (% of total labor force ages 15-24) (modeled ILO estimate)
254	SL.UEM.ADVN.ZS	WBGD	Unemployment with advanced education (% of total labor force with advanced education)
255	SL.UEM.BASC.FE.ZS	WBGD	Unemployment with basic education, female (% of female labor force with basic education)
256	SL.UEM.BASC.MA.ZS	WBGD	Unemployment with basic education, male (% of male labor force with basic education)
257	SL.UEM.BASC.ZS	WBGD	Unemployment with basic education (% of total labor force with basic education)
258	SL.UEM.INTM.FE.ZS	WBGD	Unemployment with intermediate education, female (% of female labor force with intermediate education)
259	SL.UEM.INTM.MA.ZS	WBGD	Unemployment with intermediate education, male (% of male labor force with intermediate education)
260	SL.UEM.INTM.ZS	WBGD	Unemployment with intermediate education (% of total labor force with intermediate education)

261	SL.UEM.TOTL.FE.NE.ZS	WBGD	Unemployment, female (% of female labor force) (national estimate)
262	SL.UEM.TOTL.FE.ZS	WBGD	Unemployment, female (% of female labor force) (modeled ILO estimate)
263	SL.UEM.TOTL.MA.NE.ZS	WBGD	Unemployment, male (% of male labor force) (national estimate)
264	SL.UEM.TOTL.MA.ZS	WBGD	Unemployment, male (% of male labor force) (modeled ILO estimate)
265	SP.ADO.TFRT	WBGD	Adolescent fertility rate (births per 1,000 women ages 15-19)
266	SP.DYN.CBRT.IN	WBGD	Birth rate, crude (per 1,000 people)
267	SP.DYN.CDRT.IN	WBGD	Death rate, crude (per 1,000 people)
268	SP.DYN.IMRT.FE.IN	WBGD	Mortality rate, infant, female (per 1,000 live births)
269	SP.DYN.IMRT.MA.IN	WBGD	Mortality rate, infant, male (per 1,000 live births)
270	SP.DYN.LE00.FE.IN	WBGD	Life expectancy at birth, female (years)
271	SP.DYN.LE00.MA.IN	WBGD	Life expectancy at birth, male (years)
272	SP.DYN.LE60.FE.IN	WBGD	Life expectancy at age 60, female
273	SP.DYN.LE60.MA.IN	WBGD	Life expectancy at age 60, male
274	SP.DYN.TFRT.IN	WBGD	Fertility rate, total (births per woman)
275	SP.DYN.TO65.FE.ZS	WBGD	Survival to age 65, female (% of cohort)
276	SP.DYN.TO65.MA.ZS	WBGD	Survival to age 65, male (% of cohort)
277	SP.POP.0014.TO.ZS	WBGD	Population ages 0-14 (% of total population)
278	SP.POP.1564.TO.ZS	WBGD	Population ages 15-64 (% of total population)
279	SP.POP.AG00.FE.IN	WBGD	Age population, age 00, female, interpolated
280	SP.POP.AG00.MA.IN	WBGD	Age population, age 00, male, interpolated
281	SP.POP.AG01.FE.IN	WBGD	Age population, age 01, female, interpolated
282	SP.POP.AG01.MA.IN	WBGD	Age population, age 01, male, interpolated
283	SP.POP.AG02.FE.IN	WBGD	Age population, age 02, female, interpolated
284	SP.POP.AG02.MA.IN	WBGD	Age population, age 02, male, interpolated
285	SP.POP.AG03.FE.IN	WBGD	Age population, age 03, female, interpolated
286	SP.POP.AG03.MA.IN	WBGD	Age population, age 03, male, interpolated
287	SP.POP.AG04.FE.IN	WBGD	Age population, age 04, female, interpolated
288	SP.POP.AG04.MA.IN	WBGD	Age population, age 04, male, interpolated
289	SP.POP.AG05.FE.IN	WBGD	Age population, age 05, female, interpolated
290	SP.POP.AG05.MA.IN	WBGD	Age population, age 05, male, interpolated
291	SP.POP.BRTH.MF	WBGD	Sex ratio at birth (male births per female births)
292	SP.POP.TOTL	WBGD	Population, total
293	Election_Type_Binary	I-IDEA	0 if no elections and 1, 2, 3 according to the type of elections
294	Elections	I-IDEA	0 if no elections 1 if elections
295	BureaucracyQuality	ICRG	high points are given to countries where the bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services
296	Corruption	ICRG	an assessment of corruption within the political system

297	DemocraticAccountability	ICRG	a measure of how responsive government is to its people
298	EthnicTensions	ICRG	an assessment of the degree of tension within a country attributable to racial, nationality, or language divisions
299	ExternalConflict	ICRG	an assessment both of the risk to the incumbent government from foreign action to violent external pressure
300	GovernmentStability	ICRG	an assessment both of the government's ability to carry out its declared program(s), and its ability to stay in office
301	InternalConflict	ICRG	an assessment of political violence in the country and its actual or potential impact on governance
302	InvestmentProfile	ICRG	an assessment of factors affecting the risk to investment that are not covered by other political, economic and financial risk components
303	LawAndOrder	ICRG	the "Law" element is an assesment of the strength and impartiality of the legal system. The "Order" element is an assessment of popular observance of the law
304	MilitaryInPolitics	ICRG	lower risk ratings indicate a greater degree of military participation in politics and a higher level of political risk
305	ReligiousTensions	ICRG	an assesment of the domination of society and/or governance by a single religious group that seeks to replace civil law by religious law and to exclude other religions from the political and/or social process
306	SocioeconomicConditions	ICRG	an assessment of the socioeconomic pressures at work in society that could constrain government action or fuel social dissatisfaction
307	gini_disp	SWIID	Disposable Gini
308	gini_disp_se	SWIID	Disposable Gini st. error
309	gini_mkt	SWIID	Market Gini
310	gini_mkt_se	SWIID	Market Gini st. error
311	v2x_polyarchy	VDEM	Electoral democracy index
312	v2x_libdem	VDEM	Liberal democracy index
313	v2x_partipdem	VDEM	Participatory democracy index
314	v2x_freexp_altinf	VDEM	Freedomof Expression and Alternative Sources of Information index
315	v2xel_frefair	VDEM	Clean elections index
316	v2x_elecoff	VDEM	Elected officials index
317	v2xcl_rol	VDEM	Equality before the law and individual liberty index
318	v2x_jucon	VDEM	Judicial constraints on the executive index
319	v2x_partip	VDEM	Participatory component index
320	v2x_cspart	VDEM	Civil society participation index
321	v2xdd_dd	VDEM	Direct popular vote index

322	v2xel_locelec	VDEM	Local government index
323	v2xel_regelec	VDEM	Regional government index
324	v2xdl_delib	VDEM	Deliberative component index
325	v2x_egal	VDEM	Egalitarian component index
326	v2xeg_eqprotec	VDEM	Equal protection index
327	v2xeg_eqaccess	VDEM	Equal access index
328	v2elembaut	VDEM	Electoral Management Body (EMB) autonomy
329	v2elembcap	VDEM	Electoral Management Body (EMB) capacity
330	v2elmulpar	VDEM	Elections multiparty
331	v2elrgstry	VDEM	Election voter registry
332	v2elvotbuy	VDEM	Election vote buying
333	v2elirreg	VDEM	Election other voting irregularities
334	v2elintim	VDEM	Election government intimidation
335	v2elpeace	VDEM	Election other electoral violence
336	v2elboycot	VDEM	Election boycotts
337	v2elfrcamp	VDEM	Election free campaign media
338	v2elpdcamp	VDEM	Election paid campaign advertisements
339	v2elpaidig	VDEM	Election paid interest group media
340	v2elfrfair	VDEM	Election free and fair
341	v2elaccept	VDEM	Election losers accept results
342	v2elasmoff	VDEM	Election assume office
343	v2elrsthos	VDEM	Head Of State (HOS) restriction by ethnicity, race, religion, or language
344	v2ellocons	VDEM	Lower chamber election consecutive
345	v2ellocpwr	VDEM	Local offices relative power
346	v2elffelr	VDEM	Subnational elections free and fair
347	v2elsnlsff	VDEM	Subnational election unevenness
348	v2psbars	VDEM	Barriers to parties
349	v2psparban	VDEM	Party ban
350	v2psoppaut	VDEM	Opposition parties autonomy
351	v2psorgs	VDEM	Party organizations
352	v2psprbrch	VDEM	Party branches
353	v2psprlnks	VDEM	Party linkages
354	v2psplats	VDEM	Distinct party platforms
355	v2pscnslnl	VDEM	Candidate selection-national/local
356	v2pscohesv	VDEM	Legislative party cohesion
357	v2pscomprg	VDEM	Party competition across regions
358	v2psnatpar	VDEM	National party control
359	v2pssunpar	VDEM	Subnational party control
360	v2ddlexci	VDEM	Initiatives permitted
361	v2ddthreci	VDEM	Popular initiative credible threat
362	v2ddlexrf	VDEM	Referendums permitted
363	v2ddlexor	VDEM	Enforcement of Constitutional changes through popular vote
364	v2ddyrci	VDEM	Occurrence of citizen-initiatives this year
365	v2exremhsp	VDEM	Head Of State (HOS) removal by legislature in practice
366	v2exdfdshs	VDEM	Head Of State (HOS) dissolution in practice

367	v2exdfcbhs	VDEM	Head Of State (HOS) appoints cabinet in practice
368	v2exdfvths	VDEM	Head Of State (HOS) veto power in practice
369	v2exdfdmhs	VDEM	Head Of State (HOS) dismisses ministers in practice
370	v2exdfpphs	VDEM	Head Of State (HOS) proposes legislation in practice
371	v2exagehos	VDEM	Head Of State (HOS) age
372	v2exfemhos	VDEM	Head Of State (HOS) female
373	v2exfxtmhg	VDEM	Head Of Government (HOG) term length by law
374	v2ex_hogw	VDEM	Relative power of the Head Of Government (HOG)
375	v2exrescon	VDEM	Executive respects constitution
376	v2exbribe	VDEM	Executive bribery and corrupt exchanges
377	v2exembez	VDEM	Executive embezzlement and theft
378	v2excrptps	VDEM	Public sector corrupt exchanges
379	v2exthftps	VDEM	Public sector theft
380	v2regdur	VDEM	Regime duration
381	v2lgqstexp	VDEM	Legislature questions officials in practice
382	v2lginvstp	VDEM	Legislature investigates in practice
383	v2lgotovst	VDEM	Executive oversight
384	v2lgcrrpt	VDEM	Legislature corrupt activities
385	v2lgoppart	VDEM	Legislature opposition parties
386	v2lgfunds	VDEM	Legislature controls resources
387	v2lgdsadlo	VDEM	Representation of disadvantaged social groups
388	v2lgdsadlobin	VDEM	Representation of disadvantaged social groups binary
389	v2ex_hosw	VDEM	Relative power of the Head of State (HOS)
390	v2lgtreaty	VDEM	Legislature approval of treaties by law
391	v2lglegplo	VDEM	Lower chamber legislates in practice
392	v2lgcomslo	VDEM	Lower chamber committees
393	v2lgsrvlo	VDEM	Lower chamber members serve in government
394	v2lgstafflo	VDEM	Lower chamber staff
395	v2lgello	VDEM	Lower chamber elected
396	v2dlreason	VDEM	Reasoned justification
397	v2dlcommon	VDEM	Common good
398	v2dlcountr	VDEM	Respect counterarguments
399	v2dlconslt	VDEM	Range of consultation
400	v2dlengage	VDEM	Engaged society
401	v2dlencmps	VDEM	Particularistic or public goods
402	v2dlunivl	VDEM	Means-tested v. universalistic policy
403	v2jureform	VDEM	Judicial reform
404	v2jupurge	VDEM	Judicial purges
405	v2jupoatck	VDEM	Government attacks on judiciary
406	v2jupack	VDEM	Court packing
407	v2juaccnt	VDEM	Judicial accountability

408	v2jucorrdc	VDEM	Judicial corruption decision
409	v2juhcind	VDEM	High court independence
410	v2juncind	VDEM	Lower court independence
411	v2juhccomp	VDEM	Compliance with high court
412	v2jucomp	VDEM	Compliance with judiciary
413	v2jureview	VDEM	Judicial review
414	v2cltort	VDEM	Freedom from torture
415	v2clkill	VDEM	Freedom from political killings
416	v2clslavem	VDEM	Freedom from forced labor for men
417	v2cltrnslw	VDEM	Transparent laws with predictable enforcement
418	v2clrspct	VDEM	Rigorous and impartial public administration
419	v2clacjstm	VDEM	Access to justice for men
420	v2clacjstw	VDEM	Access to justice for women
421	v2clacjust	VDEM	Social class equality in respect for civil liberty
422	v2clsocgrp	VDEM	Social group equality in respect for civil liberties
423	v2clrgunev	VDEM	Subnational civil liberties unevenness
424	v2cldiscm	VDEM	Freedom of discussion for men
425	v2clacfree	VDEM	Freedom of academic and cultural expression
426	v2clrelig	VDEM	Freedom of religion
427	v2clfmove	VDEM	Freedom of foreign movement
428	v2cldmovem	VDEM	Freedom of domestic movement for men
429	v2clstown	VDEM	State ownership of economy
430	v2clprptym	VDEM	Property rights for men
431	v2clprptyw	VDEM	Property rights for women
432	v2svdomaut	VDEM	Domestic autonomy
433	v2svinlaut	VDEM	International autonomy
434	v2svstterr	VDEM	State authority over territory
435	v2stfisccap	VDEM	State fiscal source of revenue
436	v2strenadm	VDEM	Bureaucratic remuneration
437	v2stcritrecadm	VDEM	Criteria for appointment decisions in the state administration
438	v2strenarm	VDEM	Remuneration in the Armed Forces
439	v2cseeorgs	VDEM	Civil Society Organization (CSO) entry and exit
440	v2csreprss	VDEM	Civil Society Organization (CSO) repression
441	v2cscnsult	VDEM	Civil Society Organization (CSO) consultation
442	v2csprtcpt	VDEM	Civil Society Organization (CSO) participatory environment
443	v2csgender	VDEM	Civil Society Organization (CSO) women’s participation
444	v2csantimv	VDEM	Civil Society Organization (CSO) anti-system movements
445	v2csrlgrep	VDEM	Religious organization repression
446	v2mecenefm	VDEM	Government censorship effort
447	v2mecenefi	VDEM	Internet censorship effort
448	v2mecenefibin	VDEM	Internet binary
449	v2mecrit	VDEM	Print/broadcast media critical
450	v2merange	VDEM	Print/broadcast media perspectives

451	v2mefemjrn	VDEM	Female journalists
452	v2meharjrn	VDEM	Harassment of journalists
453	v2meslfcen	VDEM	Media self-censorship
454	v2mebias	VDEM	Media bias
455	v2mecorrpt	VDEM	Media corrupt
456	v2pepwrses	VDEM	Power distributed by socioeconomic position
457	v2pepwrgen	VDEM	Power distributed by gender
458	v2pepwrort	VDEM	Power distributed by sexual orientation
459	v2peedueq	VDEM	Educational equality
460	v2pehealth	VDEM	Health equality
461	v2clgencl	VDEM	Gender equality in respect for civil liberties
462	v2peapsgen	VDEM	Access to public services distributed by gender
463	v2peasjgen	VDEM	Access to state jobs by gender
464	v2peasbgen	VDEM	Access to state business opportunities by gender
465	v2pepwrgeo	VDEM	Power distributed by urban-rural location
466	v2clgeocl	VDEM	Urban-rural location equality in respect for civil liberties
467	v2peapsgeo	VDEM	Access to public services distributed by urban-rural location
468	v2peasjgeo	VDEM	Access to state jobs by urban-rural location
469	v2peasbegeo	VDEM	Access to state business opportunities by urban-rural location
470	v2clpolcl	VDEM	Political group equality in respect for civil liberties
471	v2peapspol	VDEM	Access to public services distributed by political group
472	v2peasjpol	VDEM	Access to state jobs by political group
473	v2peasbepol	VDEM	Access to state business opportunities by political group
474	v2peapssoc	VDEM	Access to public services distributed by social group
475	v2peasjsoc	VDEM	Access to state jobs by social group
476	v2peasbsoc	VDEM	Access to state business opportunities by social group
477	v2exl_legitideol	VDEM	Ideology
478	v2exl_legitlead	VDEM	Person of the Leader
479	v2exl_legitperf	VDEM	Performance legitimation
480	v2exl_legitratio	VDEM	Rational-legal legitimation
481	v2cacamps	VDEM	Political polarization
482	v2caviol	VDEM	Political violence
483	v2caassemb	VDEM	Freedom of peaceful assembly
484	v2cagenmob	VDEM	Mass mobilization
485	v2caconmob	VDEM	Mass mobilization concentration
486	v2cademmob	VDEM	Mobilization for democracy
487	v2caautmob	VDEM	Mobilization for autocracy
488	v2castate	VDEM	Engagement in state-administered mass organizations

489	v2catrauni	VDEM	Engagement in independent trade unions
490	v2cafres	VDEM	Freedom to research and teach
491	v2cafexch	VDEM	Freedom of academic exchange and dissemination
492	v2cainsaut	VDEM	Institutional autonomy
493	v2casurv	VDEM	Campus integrity
494	v2cacritic	VDEM	Academics as critics
495	v2x_accountability	VDEM	Accountability index
496	v2x_veracc	VDEM	Vertical accountability index
497	v2x_diagacc	VDEM	Diagonal accountability index
498	v2x_horacc	VDEM	Horizontal accountability index
499	v2x_ex_confidence	VDEM	Confidence dimension index
500	v2x_ex_direlect	VDEM	Direct election dimension index
501	v2x_ex_party	VDEM	Ruling party dimension index
502	v2x_neopat	VDEM	Neopatrimonial Rule Index
503	v2xnp_client	VDEM	Clientelism Index
504	v2xnp_pres	VDEM	Presidentialism Index
505	v2xnp_regcorr	VDEM	Regime corruption
506	v2x_civlib	VDEM	Civil liberties index
507	v2x_clphy	VDEM	Physical violence index
508	v2x_clpol	VDEM	Political civil liberties index
509	v2x_clpriv	VDEM	Private civil liberties index
510	v2xpe_exlecon	VDEM	Exclusion by Socio-Economic Group
511	v2xpe_exlgender	VDEM	Exclusion by Gender index
512	v2xpe_exlgeo	VDEM	Exclusion by Urban-Rural Location index
513	v2xpe_exlpol	VDEM	Exclusion by Political Group index
514	v2xpe_exlsocgr	VDEM	Exclusion by Social Group index
515	v2x_corr	VDEM	Political corruption index
516	v2x_execorr	VDEM	Executive corruption index
517	v2x_pubcorr	VDEM	Public sector corruption index
518	v2x_gender	VDEM	Women political empowerment index
519	v2x_gencl	VDEM	Women civil liberties index
520	v2x_gencs	VDEM	Women civil society participation index
521	v2x_genpp	VDEM	Women political participation index
522	v2xcl_acjst	VDEM	Access to justice
523	v2xcl_prpty	VDEM	Property rights
524	v2xdd_i_pl	VDEM	Plebiscite index
525	v2xcs_ccsi	VDEM	Core civil society index
526	v2xex_elecreg	VDEM	Executive electoral regime index
527	v2x_EDcomp_thick	VDEM	Electoral component index
528	v2x_freexp	VDEM	Freedom of expression index
529	v2xcl_disc	VDEM	Freedom of discussion
530	v2xcl_dmove	VDEM	Freedom of domestic movement
531	v2xel_elecparl	VDEM	Legislative or constituent assembly election
532	v2xel_elecpres	VDEM	Presidential election
533	v2xme_altinf	VDEM	Alternative sources of information index
534	v2xps_party	VDEM	Party institutionalization index
535	v2x_divparctrl	VDEM	Divided party control index

536	v2x_feduni	VDEM	Division of power index
537	v2xca_academ	VDEM	Academic Freedom Index
538	v2smgovdom	VDEM	Government dissemination of false information domestic
539	v2smgovab	VDEM	Government dissemination of false information abroad
540	v2smpardom	VDEM	Party dissemination of false information domestic
541	v2smfordom	VDEM	Foreign governments dissemination of false information
542	v2smforads	VDEM	Foreign governments ads
543	v2smgovfilcap	VDEM	Government Internet filtering capacity
544	v2smgovshutcap	VDEM	Government Internet shut down capacity
545	v2smgovsm	VDEM	Government social media shut down in practice
546	v2smgovcapsec	VDEM	Government cyber security capacity
547	v2smpolcap	VDEM	Political parties cyber security capacity
548	v2smregcon	VDEM	Internet legal regulation content
549	v2smprivex	VDEM	Privacy protection by law exists
550	v2smregcap	VDEM	Government capacity to regulate online content
551	v2smregapp	VDEM	Government online content regulation approach
552	v2smlawpr	VDEM	Defamation protection
553	v2smdefabu	VDEM	Abuse of defamation and copyright law by elites
554	v2smonex	VDEM	Online media existence
555	v2smonper	VDEM	Online media perspectives
556	v2smmefra	VDEM	Online media fractionalization
557	v2smorgviol	VDEM	Use of social media to organize offline violence
558	v2smorgavgact	VDEM	Average people’s use of social media to organize offline action
559	v2smorgelitact	VDEM	Elites’use of social media to organize offline action
560	v2smcamp	VDEM	Party/candidate use of social media in campaigns
561	v2smarrest	VDEM	Arrests for political content
562	v2smpolsoc	VDEM	Polarization of society
563	v2smpolhate	VDEM	Political parties hate speech
564	IR.LT	OECD	Long-term interest rates refer to government bonds maturing in ten years
565	IR.ST	OECD	Short-term interest rates, based on three-month money market rates.
566	spread	authors	Expected spread compared to the German IR.LT for the same year as a benchmark (Cimadomo et al., 2016, C. A. Favero, 2013, Von Hagen et al., 2011).
567	shifted_sdg_value	Eurostat	Target variable, government debt the next year for each country

Table 4: Variable Importance Measure results, using gain, for all 25 variables for the optimal XGBoost model.

Gain	Feature	Description
6375.472	sdg_value_lag_1	Ratio of government debt outstanding at the end of the year to gross domestic product at current market prices
2982.187	sdg_value_lag_2	Ratio of government debt outstanding at the end of the year to gross domestic product at current market prices
1872.879	SP.POP.5054.MA_lag_3	Population ages 50-54, male
598.7888	RQ.PER.RNK_lag_1	Regulatory Quality, Percentile Rank
368.8009	CC.PER.RNK_lag_1	Control of Corruption, Percentile Rank
356.1886	SL.EMP.TOTL.SP.FE.NE.ZS_lag_1	Employment to population ratio, 15+, female (%) (national estimate)
174.3274	spread_lag_1	10 Year bond spread (Cimadomo et al., 2016, C. A. Favero, 2013, Von Hagen et al., 2011)
130.6781	v2elfrcamp_lag_2	Election free campaign media
118.8208	NV.IND.TOTL.ZS_lag_1	Industry (including construction), value added (% of GDP)
115.1123	v2elpeace_lag_2	Election other electoral violence
104.4191	SL.EMP.VULN.MA.ZS_lag_1	Vulnerable employment, male (% of male employment) (modeled ILO estimate)
96.91984	SL.EMP.TOTL.SP.ZS_lag_1	Employment to population ratio, 15+, total (%) (modeled ILO estimate)
94.93802	NE.CON.PRVT.ZS_lag_2	Households and Non-Profit Institutions Serving Households (NPISH) final consumption expenditure (% of GDP)
89.91057	IP.JRN.ARTC.SC_lag_1	Scientific and technical journal articles
52.44279	v2xpe_exlgeo_lag_1	Exclusion by Urban-Rural Location index
23.27445	NE.GDI.FTOT.ZS_lag_3	Gross fixed capital formation (% of GDP)
19.20233	SL.UEM.INTM.FE.ZS_lag_1	Unemployment with intermediate education, female (% of female labor force with intermediate education)
18.2017	NE.CON.TOTL.ZS_lag_2	Final consumption expenditure (% of GDP)
14.89868	TX.VAL.MANF.ZS.UN_lag_2	Manufactures exports (% of merchandise exports)
14.48244	NV.SRV.TOTL.ZS_lag_3	Services, value added (% of GDP)
13.98627	NV.IND.MANF.ZS_lag_2	Manufacturing, value added (% of GDP)
13.96676	SL.AGR.EMPL.FE.ZS_lag_1	Employment in agriculture, female (% of female employment) (modeled ILO estimate)
12.82722	NY.GDS.TOTL.ZS_lag_2	Gross domestic savings (% of GDP)
10.71115	TX.VAL.MANF.ZS.UN_lag_3	Manufactures exports (% of merchandise exports)
7.277377	NV.IND.TOTL.ZS_lag_2	Industry (including construction), value added (% of GDP)