Inter-American Development Bank

IDB WORKING PAPER SERIES Nº IDB-WP-1346

Assessing Macro-Fiscal Risk for Latin American and Caribbean Countries

Oscar M. Valencia
Juan Camilo Díaz
Diego A. Parra

Inter-American Development Bank
Institutions for Development Sector
Fiscal Management Division

October 2022

IDB
Inter-American
Development Bank

Assessing Macro-Fiscal Risk for Latin American and Caribbean Countries

Oscar M. Valencia
Juan Camilo Díaz
Diego A. Parra

October 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Valencia Arana, Oscar.
Assessing macro-fiscal risk for Latin American and Caribbean Countries / Oscar Valencia, Juan Camilo Díaz, Diego Parra. p. cm. — (IDB Working Paper Series ; 1346)
Includes bibliographic references.
1. Fiscal risk management-Latin America-Econometric models. 2. Fiscal risk management-Caribbean Area-Econometric models. 3. Fiscal crisis-Latin America-Econometric models. 4. Fiscal crisis-Caribbean Area-Econometric models. I. Díaz, Juan Camilo. II. Parra, Diego. III. Inter-American Development Bank. Fiscal Management Division. IV. Title. V. Series.

IDB-WP-1346

http://www.iadb.org

Assessing Macro-Fiscal Risk for Latin American and Caribbean Countries

Oscar M. Valencia* Juan C. Díaz† Diego A. Parra‡

October 27, 2022

Abstract

This paper provides a comprehensive early warning system (EWS) that balances the classical signaling approach with the best-realized machine learning (ML) model for predicting fiscal stress episodes. Using accumulated local effects (ALE), we compute a set of thresholds for the most informative variables that drive the correlation between predictors. In addition, to evaluate the main country risks, we propose a leading fiscal risk indicator, highlighting macro, fiscal and institutional attributes. Estimates from different models suggest significant heterogeneity among the most critical variables in determining fiscal risk across countries. While macro variables have higher relevance for advanced countries, fiscal variables were more significant for Latin American and Caribbean (LAC) and emerging economies. These results are consistent under different liquidity-solvency metrics and have deepened since the global financial crisis.

Keywords: forecasting, early warning system, fiscal policy
JEL Classi ication: C53, H63, E62

*Oscar M. Valencia, Inter-American Development Bank (IDB), oscarva@iadb.org
†Juan Camilo Díaz H, jdiazh@javeriana.edu.co
‡Diego Alejandro Parra, da.parra12@uniandes.edu.co

1 Introduction

The COVID-19 crisis highlighted the vulnerability of countries' macro-fiscal structure, with public and private debt levels climbing at unprecedented rates. After the severe crash of 2020, most economies faced great uncertainty about a sustained recovery. Policymakers in the region must now sharpen their tools to anticipate shocks by measuring macro-fiscal vulnerabilities to adopt preemptive policy measures. The design of effective and timely policy actions can be achieved through tools like early warning systems (EWS), which involve the identification of potential vulnerabilities through indicators that have systematically different behaviors in periods preceding a crisis. Our analysis addresses the determination of macro-fiscal risk, defined as the short-term risk of facing a sovereign liquidity and/or default crisis (Baldacci et al., 2011; Hernández de Cos et al., 2014; Gerling et al., 2017; Beers and de Leon-Manlagnit, 2021).

Since the 1990s, the literature has included several methodologies to detect fiscal risks. A widely used approach is the signaling methodology developed by Kaminsky, Lizondo, and Reinhart (1998), which relies on signal variables calculated for each indicator that best predicts an upcoming fiscal distress event. Each time an indicator exceeds a critical value, a signal is sent. The signaling methodology also enables the creation of leading indicators by grouping relevant variables to anticipate fiscal stress events. For example, Hernández de Cos et al. (2014) construct leading indicators that capture fiscal stress for each country in the Euro area. The study finds that the signaling method is more accurate when it includes a country-specific threshold for the signaling variables for both fiscal and financial variables. Similarly, Baldacci et al. (2011) use this approach through a fiscal stress index that provides early warning signals of fiscal sustainability problems for advanced and emerging economies.

An additional EWS trend employs several econometric models such as logit and probit. Lane and Milesi-Ferretti (2017) use a probit model to measure risk in low-income countries. Foreign income, risk premium, conflict dummy variable, global growth, and institutions are highly relevant variables predicting risk in low-income countries. Hilscher and Nosbusch (2007) find that terms-of-trade volatility is significant for changes in spreads and default probabilities. Maltritz and Molchanov (2014) use a Bayesian averaging model and find similar results, but with more significant competitiveness and governance variables, such as corruption-free property rights, governance practices, and countries' economic activities.

While these methodologies pioneered outcome prediction, there have been weaknesses in predicting fiscal stress events, particularly out-of-sample events (Dawood, Horsewood, and Strobel, 2017; Berg, Borensztein, and Pattillo, 2005). Recently, the use of machine learning (ML) models has increased to capture nonlinear phenomena and correct this poor out-of-sample performance. For example, several studies employing random forest models achieved a higher performance than traditional approaches to countries' risk analysis like linear and logistic regression models (Jarmulska, 2020; Moreno Badia et al., 2020).

However, one shortcoming of using ML models is that they are often considered "black box" models, which means that these models are difficult to interpret due to their high level of complexity. Estimating the marginal effects of the predictors is critically important in applications such as predicting fiscal stress episodes. The fitted model would be unreliable if the effect of a predictor violates intuition. To address this issue, this study uses the accumulated local effects (ALE) technique proposed by Apley and Zhu (2020). This methodology allows us to isolate each predictor's effect considering the impact of correlated variables.

In that sense, this paper presents a methodology that systematically integrates signaling and ML tools for EWS consolidation, contributing to the literature on three fronts. First, it derives the critical values associated with each variable within ML models using the ALE technique, which yields a deeper understanding of the factors contributing to fiscal stress. Second, it proposes a leading indicator encompassing fiscal, macroeconomic, and institutional factors by employing both methodologies. Third, it adds to the large and growing literature on EWS for emerging economies while comparing LAC with other emerging and advanced economies (Hellwig, 2021; Jarmulska, 2020; Hernández de Cos et al., 2014).

This study finds a strong heterogeneity in the variables that shape fiscal stress episodes across countries. Macroeconomic variables (different from fiscal variables) play an essential role for advanced economies, while fiscal variables have a more significant impact on estimating fiscal stress for emerging economies and LAC. Foreign currency debt is the most crucial variable for LAC and emerging countries. On the other hand, the human capital index and unemployment rates are the best predictors of fiscal stress in advanced economies. The risk maps indicate a considerable increase in fiscal risk based on the upward trend in public debt, further accentuated after the international financial crisis.

Countries differ in terms of the risk they face. While advanced countries have higher debt levels than emerging and LAC countries, the fiscal risk is lower than other countries. Fiscal revenues, currency mismatches, and interest payments dynamics largely explain these differences. Finally, the results of the leading indicator suggested that the recent COVID-19 crisis presented a higher macro-fiscal risk for emerging and LAC countries than the global financial crisis did. Also, the fiscal performance of emerging economies, including LAC, is the main factor contributing to an increase in vulnerabilities since the global financial crisis.

The structure of this paper is as follows: Section 2 describes the dataset and presents the definition of the fiscal stress–dependent variable as well as the predictors to be used in the estimations. Section 3 details the methodology, starting with the signaling model and going through the relevant ML approaches and the ALE technique. Section 4 shows the estimation process results, starting with the selection of the most relevant variables, the estimation and selection of models, the estimation of the critical values per variable and country group, and, finally, the leading early warning indicator. Finally, Section 5 concludes.

2 Data

The data for the analysis consists of panel data of 180 countries and 89 variables, with annual figures from 1990 to 2020. In addition, we classified the countries into three groups:

advanced (ADV), emerging (EME), and Latin America and the Caribbean (LAC). In turn, each predictor was categorized as: fiscal, macroeconomic, or institutional. This section presents the definition of fiscal stress, the main stylized facts, and the predictors used in the estimation.

2.1 Fiscal Stress Episode Definition

A fiscal stress episode refers to a period in which a country suffers extreme shocks on its debt dynamics, forcing it to take extraordinary measures. These difficulties can create a certain degree of distress, as they could lead to a fiscal crisis if a country fails to adjust its budget constraint to correct imbalances. The severity of the fiscal stress episode would force the country to tighten its financing by taking exceptional measures such as debt default, restructuring, debt reduction, or money printing.

The criteria adopted in this study are consistent with Gerling et al. (2017), who identifies tax stress when at least one of the following conditions is met:

1. Any credit event associated with sovereign debt. For instance, debt default or restructuring.

2. Recourse to IMF financing on a large scale. This event occurs when the country intends to use exceptional support as an alternative to default. It is captured through an agreement on 100% of the quota and an adjustment in fiscal programs.

3. Periods of implicit domestic default through (i) high inflation rates due to money printing, and (ii) domestic arrears accumulation represented by other payables.

4. Loss of market confidence in the sovereign. This event captures any extreme change in market pressures and results via two channels: (i) when the sovereign ceases issuing bonds and (ii) when the price exceeds a 1,000 basis point barrier.

Following these criteria, Gerling et al. (2017) constructed a database of fiscal crises from 1970 to 2015 for advanced, emerging, and low-income countries. To complete the data from 2016 to 2020, we use multiple sources. For sovereign debt default episodes and national arrears data, we relied on the BoC-BoE sovereign defaults database from Beers and de Leon-Manlagnit (2021). For IMF extraordinary financing, we used publicly available data on loan commitments.[1] As for inflation rates, we used inflation rates from the World Economic Outlook (WEO) database and proxied sovereign spreads from JP Morgan's Emerging Market Bond Index (EMBI).

[1]Data is available at https://www.imf.org/external/np/fin/tad/extarr1.aspx.

2.1.1 Stylized Facts

According to the latter, we identified 1,813 stress episodes for the 180 countries from 1990 to 2020. This represents a percentage of crisis of 32%. As for LAC, we identified 282 episodes averaging 11 episodes per country, which is relatively close to the average value for emerging economies (12.5). Conversely, advanced economies had an average of only 1.5 episodes per country throughout the sample, yielding a higher degree of imbalance with only 5% of crisis episodes. Table 1 summarizes the fiscal stress events identified in the dataset. Figures 1, 2, and 3 show the evolution of the number of countries under fiscal stress for LAC, emerging and advanced economies, segmented by the nature of the event.

	ADV	EME	LAC	TOTAL
Total countries	37	118	25	180
Total number of episodes	56	1475	282	1813
Average episodes per country	1.5	12.5	11.3	10.1
Percentage of crisis	5%	40%	36%	32%
Credit event	6	962	142	1110
Official financing	35	537	109	681
Implicit default	23	184	31	238
Market confidence	27	235	95	357

Table 1: Fiscal Stress Episodes Summary, from 1990–2020

Fiscal Crises in Advanced Countries

As described in Table 1, most advanced economies did not face fiscal stress episodes in the observed period. However, it is possible to identify three periods in which the number of stressed countries reached a peak. The first one occurred during the early 1980s due to the banking stock crisis in Israel, special IMF financing in Portugal and Korea, loss of market confidence in Cyprus, and high inflation rates in Iceland and Slovenia. The second peak materialized in the early 1990s due to episodes of implicit default in the Czech Republic, Estonia, Latvia, Lithuania, Slovakia, and Slovenia. The final peak transpired in the wake of the global financial crisis and the European debt crisis in Greece, Portugal, Iceland, Ireland, and Latvia.

Fiscal Crises in Emerging Countries

Sovereign debt defaults triggered significant fiscal stress episodes in emerging countries in the mid-1980s. There was a default debt restructuring in most countries, with a significant decline in the mid-1990s (Beers and de Leon-Manlagnit, 2021). Figure 2 shows that the footprint left by these stress episodes has faded over the years: despite some peaks during the global financial crisis (2008–09) and the European debt crisis (2012–13), the number of emerging

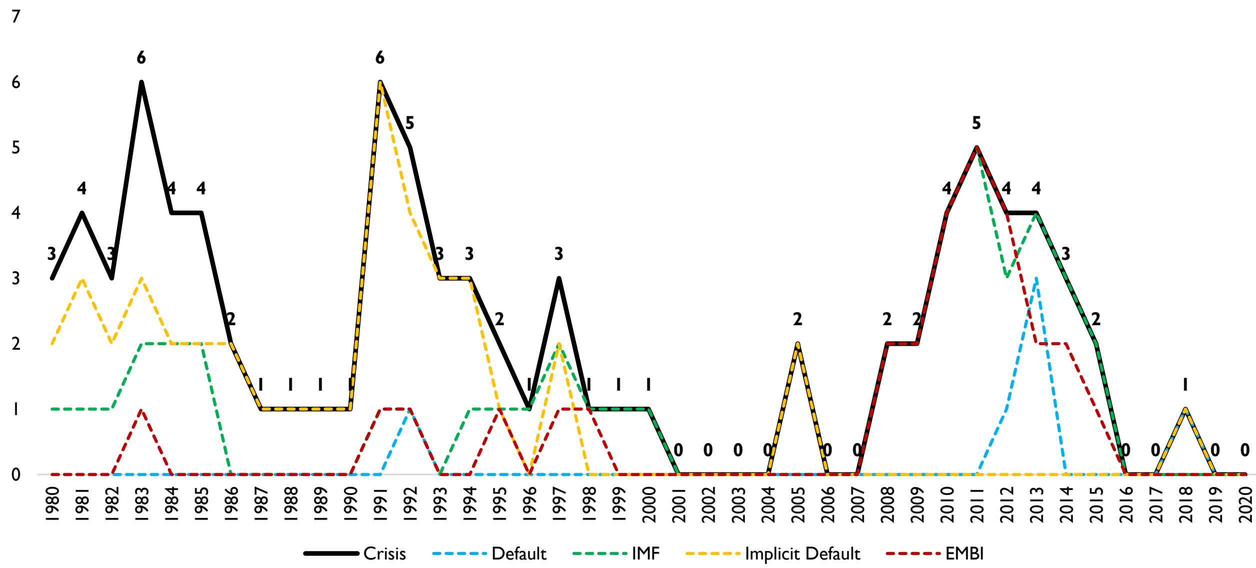

Figure 1: Advanced Countries with Fiscal Stress Episodes, by Type of Event
Source: Bank of Canada.

countries with fiscal stress episodes reached a minimum in 2015. However, it increased moderately in the following years as the commodity boom ended in the '00s. During the global economic shock caused by the COVID-19 pandemic, 54 out of 118 emerging countries faced fiscal stress in 2020, mainly driven by extraordinary IMF financing.

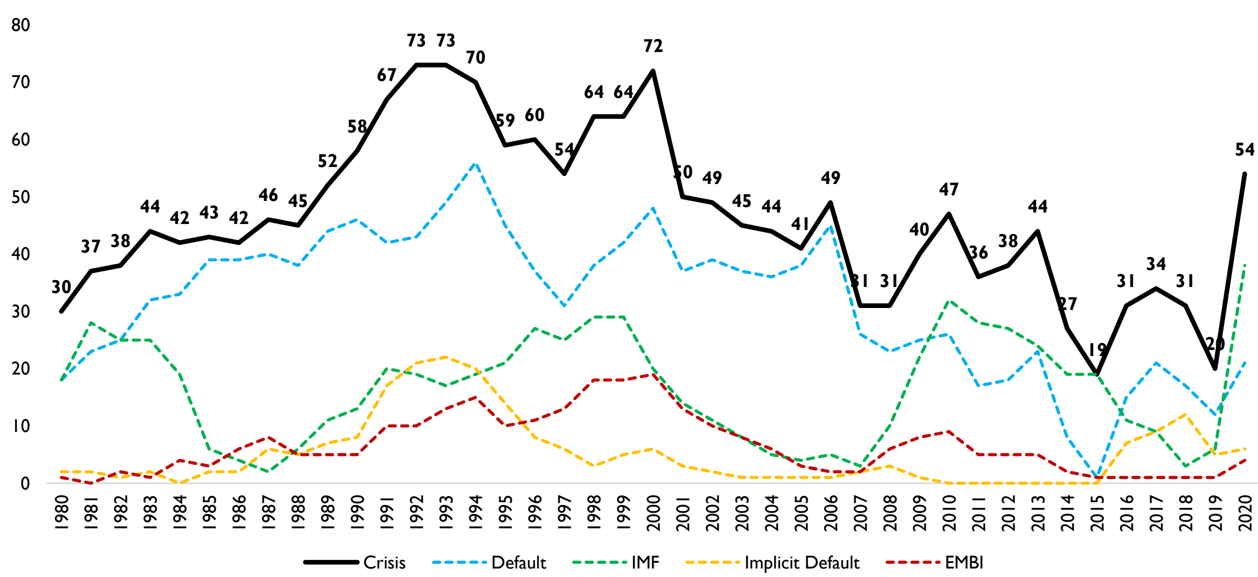

Figure 2: Number of Emerging Countries with Fiscal Stress Episodes, by Type of Event
Source: Bank of Canada.

Fiscal Crises in LAC Countries

The 1980s in LAC were shaped by a pronounced external debt crisis resulting in high fiscal deficits. The aftermath of this "lost decade" had a strong long-term impact and was characterized by a slow recovery. It was not until the 2000s that economic growth gained momentum as commodity prices rose. The years 2008 and 2009 saw the global financial crisis, in which most countries in the region faced fiscal stress due to the markets' loss of confidence, resulting in a halt in bond market issuance. More recently, the COVID-19 pandemic has now put many countries under fiscal stress. The main triggering event is an increased need for IMF financing with quotas above 100% or adjustments in their fiscal programs (Figure 3).

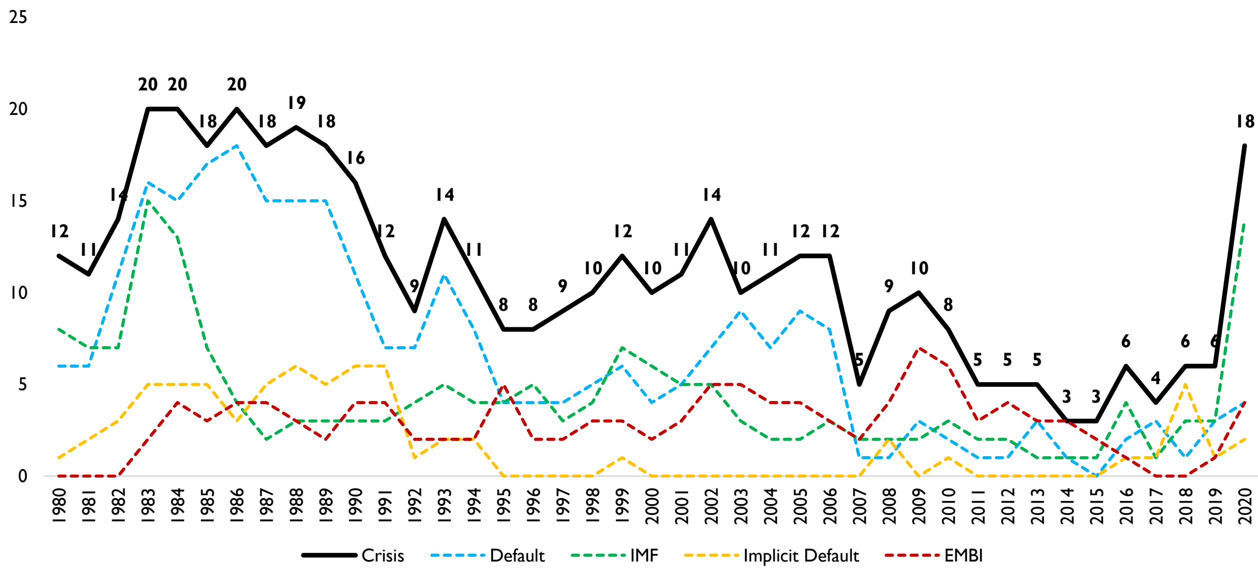

Figure 3: Number of Countries with Fiscal Stress Episodes in LAC Countries by Type of Event

Source: Bank of Canada.

2.2 Fiscal Stress Predictors

The predictors are classified in the following three groups:

1. The first group of predictors represents the *fiscal factor* and includes variables related to general government gross debt, fiscal balances, revenues, and expenditures, obtained from WEO. This group includes fiscal space and debt sustainability indicators based on data from Kose et al. (2017) and the World Bank.

2. The second group of predictors is the *macroeconomic factor*, which includes variables related to GDP, the labor market, domestic demand, macroeconomic stability, liquidity, the balance of payments, natural resource rents, and commercial and external shocks. The main source of data is the World Economic Outlook (WEO) (IMF, 2020) from

the International Monetary Fund. It is also enriched with variables from the Penn World Tables (Feenstra, Inklaar, and Timmer, 2015), the cross-country database of fiscal space from Kose et al. (2017), the external wealth of nations database from Lane and Milesi-Ferretti (2007), the International Labour Organization (ILOSTAT), and the United Nations Statistics Division.

3. Finally, the *institutional factor* includes variables related to corruption, the rule of law, regulatory quality, government effectiveness, political stability, and voice and accountability, obtained from the Worldwide Governance Indicators (Kaufmann, Kraay, and Mastruzzi, 2010). It also includes a binary democracy indicator with data from Coppedge et al. (2021).

Appendix A provides a detailed list of the variables, including sources and the percentage of missing values. We calculated the mean for each variable in both tranquil and stress periods and ran a t-test to determine whether there was a significant difference between them. The results suggest that, on average, most variables deteriorate in periods of fiscal stress. Finally, we treated missing values using linear interpolation and a random forest algorithm.

3 Methodology

This section describes the underlying methodology used for predicting fiscal risk and building the EWS. First, we present the implemented models, both the signal-based approach initially proposed by Kaminsky, Lizondo, and Reinhart (1998) and the four machine learning models to be tested: logit, random forest, gradient boosting, and support vector machine. In each model, we calculate the critical thresholds for each variable; these values determine the maximum level each variable can reach before entering the risk zone. In the case of the signals model, we implemented the classic approach described in the literature. In contrast, for the ML models, we used the ALE methodology based on Apley and Zhu (2020) to determine these risk thresholds. After obtaining these critical values, it is possible to build the early detection system for each methodology, supported by the construction of the leading indicator and the liquidity and solvency analyses based on fiscal, macroeconomic, and institutional variables.

3.1 Signaling Approach

This methodology proposed by Kaminsky, Lizondo, and Reinhart (1998) relies on an endogenously calculated signal variable for each indicator that best predicts an upcoming fiscal distress event. Aggregating these signals weighted by the predictive power of each indicator results in the composite index that works as an early warning indicator of a fiscal stress episode. This methodology makes it possible to obtain two important data points: a threshold level for each indicator and a leading indicator that provides information on fiscal distress in each country.

The main advantage of the signaling approach is the interpretability of the results in the analysis of the fiscal vulnerability since it allows aggregating multiple indices of specific variables into a single composite indicator, resulting in a more straightforward interpretation. However, its main weakness is that it ignores any existing correlation between the independent variables, as it only considers the bivariate relationship between each indicator and the fiscal stress variable. Another significant drawback is the lack of statistical significance of the indicators, posing the risk of obtaining a biased critical value (Sumner and Berti, 2017).

3.1.1 Fiscal Stress Threshold

Each time an indicator exceeds a critical value, a signal is sent. This critical value is the maximum level that a variable can withstand before falling into an unsafe zone where the system receives a warning. According to the literature, this value can be either identical for all countries, like Baldacci et al. (2011) or Berti, Salto, and Lequien (2012), or country-specific, such as Hernández de Cos et al. (2014) or Cerovic et al. (2018). However, Hernández de Cos et al. (2014) shows that the country-specific approach could eventually increase the predictive power of early warning systems.

We define the following binary variable to calculate the signal that determines an episode of fiscal distress:

$$d^i_{jt} = \begin{cases} 1, & \text{if } x^i_{jt} > \zeta^i_j \\ 0, & otherwise \end{cases} \tag{1}$$

Where d^i_{jt} stands for the binary signal emitted by variable i in period t for sovereign j, x^i_{jt} corresponds to the variable under consideration, and ζ^i_j refers to the threshold or critical value for a said variable, specific for each country. The value of this threshold matches a given percentile, which will be the same for each country in the entire study sample. The use of percentiles to define critical values takes into account the structural differences between countries and thus is a reliable determinant for a critical value (Cerovic et al., 2018).

The critical value is the percentile that maximizes —or minimizes— a specific performance metric, which is calculated using the confusion matrix for binary classification, like the one presented in Table 2.

		Real	
		Crisis	**Non Crisis**
Predicted	**Crisis**	True Positive (TP)	False Positive (FP)
	Non Crisis	False Negative (FN)	True Negative (TN)
	Total	*Total Crises (TC)*	*Total Tranquil (TNC)*

Table 2: Confusion Matrix

Two of the most common performance measures reported in the literature are signal-to-noise ratio (SNR) (Equation 2) and total misclassification error (TME) (Equation 3). SNR maximizes the percentage of true positives versus the noise given by the proportion of false positives. Ideally, the SNR should be above 1 to send more positive signals than noise. The aim of TME is to minimize the total number of errors sent to the system, given by the sum of the false-negative rate and the false-positive rate. Studies like Baldacci et al. (2011), Berti, Salto, and Lequien (2012), and Hernández de Cos et al. (2014) have shown a preference for TME over SNR, based on the size of the total errors produced and the preference for assigning a higher weight to avoid false-negative signals.

$$SNR = \frac{\text{TP}}{\text{TC}} \times \left[\frac{\text{FP}}{\text{TNC}}\right]^{-1} \tag{2}$$

$$TME = \frac{\text{FN}}{\text{TC}} + \frac{\text{FP}}{\text{TNC}} \tag{3}$$

Once the selected performance metric is optimized, the corresponding percentile is used as a threshold to calculate the signal indicator d^i_{jt}. As mentioned above, this threshold is the best predictor of a crisis event (i.e., ζ^i_j in Equation 1). The predictive power of the signal sent by each variable i is determined by the complement of the total number of misclassifications. Thus,

$$z_i = 1 - TME_i \tag{4}$$

3.1.2 Leading Index of Fiscal Stress for the Signaling Approach

Grouping different indicators into a leading index improves the predictive power of a potential debt crisis event and the consistency of the results within the signaling methodology (Kaminsky, Lizondo, and Reinhart, 1998). The calculation of the leading indicator (L) for each country j in each period t can be estimated by the following formula:

$$\mathrm{L}_{jt} = 100 * \sum_{i=1}^{n} \frac{z_i}{\sum_{k=1}^{n} h^k_{jt} z_k} d^i_{jt} \tag{5}$$

Where, z_i stands for the signal power (SP) defined by Equation 4, h^k_{jt} corresponds to an auxiliary binary variable that takes the value of 1 whenever the variable k is observed for country j at time t. Variable d^i_{jt} represents the binary signal sent by variable i, which is defined in Equation 1.

The index L_{jt} is bounded between 0 and 1, as it is a weighted measure of the binary signal. However, it scales between 0 and 100: An index of 0 means that none of the indicators sent a signal (i.e., none of the indicators exceeded the corresponding critical value), meaning a low-risk episode. On the other hand, an index of 100 implies that all indicators sent a signal to the system, suggesting a high level of fiscal stress risk.

Finally, since the leading indicator derives from varying signal strengths for different indicators, this allows observing the behavior of sovereigns under multiple fiscal, macroeconomic, and institutional factors.

3.2 Machine Learning Models

Although recent analyses have widely used the signaling approach when addressing fiscal vulnerability, this method has two critical problems that can be solved using ML methods. The first problem is the lack of interaction between variables, which biases results. The second relates to the considerable difference between the in-sample prediction performance and the out-of-sample performance, as noted above and by IMF (2021).

To solve this issue, fiscal stress episodes are also predicted using four supervised learning models (i.e., logit, random forest, gradient boosting, and support vector machine) and thresholds on predictors are calculated using the ALE method developed by Apley and Zhu (2020).

Logit Model

The first model we use to predict fiscal stress episodes $y_{i,t}$ is a logit regression. In this approach, input values $X_{i,t}$ are linearly combined using weights β_j. Then, this linear specification transfers to an activation function (i.e., sigmoid function) which restricts the outcome into values within the range 0–1:

$$y = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^{N} \beta_i X_i\right)}} \tag{6}$$

The weights β_j are trained from the training data in order to obtain the smallest error. For this purpose the loss function is defined as $L = \sum_i^n (y_i - \hat{y}_i)^2$, where $\hat{y}_i$ are predicted values. To avoid overfitting, regularization methods such as lasso (Santosa and Symes, 1986), ridge (Hoerl and Kennard, 1970), and elastic net (Zou and Hastie, 2005) are used to reduce the variance in the model. These are also known as shrinkage methods, as they reduce the estimated weights in the regression. For example, lasso adds to the loss function the sum of all weights' magnitudes: $\sum_i^n (y_i - \hat{y}_i)^2 + \lambda \sum_j^p \|\beta_j\|$, where λ is a tuning parameter that adjusts how large the penalty is. In ridge regularization, the penalty term is the sum of the squared weights: $\sum_j^p \beta_j^2$. Finally, elastic net linearly combines the penalties of the lasso and ridge methods: $\alpha \sum_j^p \|\beta_j\| + (1 - \alpha) \sum_j^p \beta_j^2$. We used a cross-validation procedure to select the penalization and its shrinkage parameter λ.

Logit models have been broadly used in the EWS literature. For instance Jarmulska (2020) predicted fiscal stress in 43 countries using this method, whereas Hellwig (2021) implemented it with an elastic net penalty in 188 advanced and emerging countries. Beutel, List, and von Schweinitz (2019) used a dataset encompassing systemic banking crises for 15 advanced countries between 1970 and 2016 to show that a simple logit model could outperform ML approaches under a variety of robustness circumstances. Although logit models can have a standard econometric interpretation based on the estimated coefficients and their significance (Jarmulska, 2020), other supervised learning models may have a better performance depending on data complexity. Below, we describe the applications for some of them.

Decision Trees

Decision trees are one of the simplest, most easily interpretable and popular models used for classification and regression problems. A decision tree, like the one presented in Figure 4, resembles an inverted tree with the trunk at the top and the leaves at the bottom. The top of the tree is called the root node and the next ones are internal nodes or branches, which extend into leaf nodes. A decision tree can be defined as a collection of decision nodes connected by branches, extending downward from the root node to the leaf nodes. When starting at the root node, the attributes undergo a series of splitting rules, and each possible outcome gives rise to a branch. Each branch leads to another decision node or a terminal leaf node.

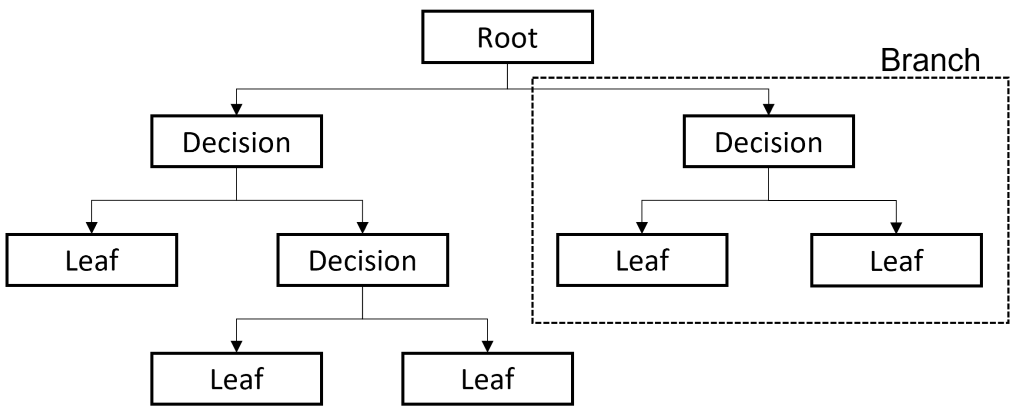

Figure 4: Decision Tree

Source: Authors' elaboration.

The recursive splitting process of building a decision tree begins by selecting the root node variable and its splitting point that best separates the classes between a fiscal crisis event and a tranquil event. Since this is measured by the node impurity that determines the homogeneity between the classes at each node, the ultimate goal of a decision tree is to reduce this impurity. This process is recursively repeated until it meets a stopping criterion. For example, if a node contains data from a single class (i.e., a pure node), there is no reason to reduce impurity, and this node is considered a leaf node. Other stopping criteria, such as the depth or the minimum number of observations required by a leaf node, could be defined ex ante.

A key disadvantage of decision trees is that the model will be prone to build large, complicated trees with output nodes containing only one observation in the absence of stopping criteria. That is, decision trees are likely to overfit the training data and fail to generalize accurately to unseen data. These complex decision trees are models with low bias and high variance. On the other hand, a smaller tree is likely to perform better generalizations at the expense of a small amount of bias.

Ensemble learning is one of the most practical approaches to address the poor performance of decision trees. It is a technique based on the idea of the "wisdom of crowds," which suggests that the decision making of a larger group of individuals is often better than that of a single expert. Thus, ensemble models aggregate individual models to achieve a better final prediction. These individual models, also known as weak learners, work together to form a strong learner and reduce bias or variance. Bagging [2] and boosting are two main types of ensemble methods. While the bagging model reduces variance, the boosting model reduces bias. An example of bagging is the random forest algorithm, whereas the most common boosting model is the gradient boosting algorithm. We describe the two algorithms below.

Random Forest

The random forest (RF) algorithm developed by Breiman (2001) represents the second established method for predicting fiscal stress episodes. The RF model is one of the ensemble methods that use the bagging technique. In other words, RF combines different individual classifiers (decision trees) to solve classification and regression problems.

Figure 5 shows the RF algorithm process for a classification problem. Initially, the dataset goes through a bootstrap process, which means that random samples are selected from the original data with the option of choosing the same sample more than once. A decision tree is then constructed using the bootstrap dataset, but instead of using all variables, a random subset of variables is selected. This process repeats a certain number of times, and then the final predictions made by each tree are aggregated. Thus, for classification purposes, the majority vote of the tree sets is used as the final prediction.

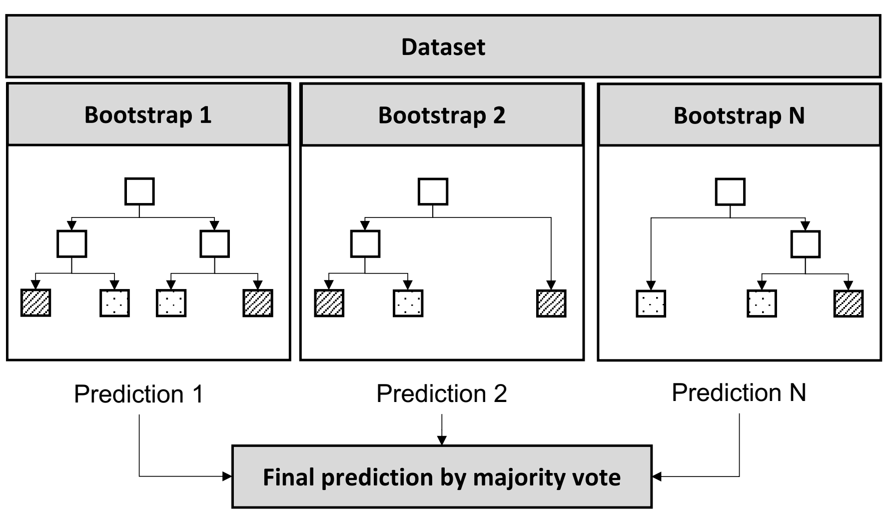

Figure 5: Random Forest Architecture

Source: Authors' elaboration.

[2]"Bagging" a shortened form of the term "bootstrapping aggregation".

Individual decision trees, also known as weak learners, form powerful learners trained in parallel and combine predictions to reduce variance. In addition, the variables evaluated at each node are randomly selected so individual trees are not correlated. RF generalizes with greater accuracy and increases its predictive power by aggregating predictions from many trees.

The RF has played a relevant role in ML lately due to key advantages such as overfitting risk reduction and convenience in assessing the importance of variables or their contribution to the model. However, the algorithm also presents challenges such as the high amount of training time and resources it requires. In addition, the output interpretability of a RF is more challenging than that of a simple decision tree.

In the literature, authors such as Moreno Badia et al. (2020), Jarmulska (2020) and Hellwig (2021) have found that the random forest performed well in predicting fiscal crises and, in most cases, proved to be one of the best performing models. For example, Jarmulska (2020), found that the efficiency of the RF model was slightly above logit models. Specifically, the RF yields an average accuracy of 80%, while the logit model is 70%-75% accuracy accurate. In this regard, Hellwig (2021) found that RF approaches consistently outperform heuristic benchmarks and other statistical models. Other studies such as IMF (2021) found that the RF model performed better compared to other ML models in emerging markets and low-income countries, based on the sum of errors and the area under the curve (AUC).

Gradient Boosting

The third model used to predict fiscal stress was the gradient boosting (GB) algorithm of Friedman (2001). Like RF, it consists of a set of multiple decision trees. However, GB uses a type of ensemble technique called boosting, in which the classifiers train sequentially, such that each new tree improves the residuals of the previous one. GB uses the prediction error gradient to identify the deficiencies of previous trees (weak learners), and a learning rate parameter controls how quickly these errors correct from each tree to the next. This method can reduce bias in predictions; however, it can also quickly overfit a training dataset. Therefore, adjustment methods play an important role in its implementation. We used cross-validation procedures to tune hyperparameters[3] such as the number of sequential trees, the maximal quantity of tree leaves, the peak depth of a tree, and the learning rate, among others.

Hellwig (2021) also used GB to predict fiscal crises and found it outperforms standard econometric approaches. He discovered that GB improves output at a marginal rate relative to the elastic net for market access countries (MACs). For low-income countries (LICs), performance improves with logit approaches but is not comparable to elastic net.

[3]The term "hyperparameters" refers to the model parameters that control the learning process in a specific ML model.

Support Vector Machine

The last model we implemented was the support vector machine (SVM) model. Developed by Boser, Guyon, and Vapnik (1996), this technique is used for both classification and regression problems. The literature considers this approach an "out of the box" model since it performs well in many scenarios, including the small dataset scenario (this paper's case). In SVM, each data corresponds to a point in an n-dimensional space. In this space, each variable corresponds to one dimension. Then, one hyperplane[4] is calculated to classify and correctly divide the fiscal crisis or tranquil event points. The coordinates of each point relative to the n-dimensional space are the support vectors. The main challenge for the algorithm is to identify this decision boundary, for which it employs the kernel trick. This technique seeks to transform a nonseparable problem into a separable one by taking a low-dimensional input space and transforming it into a higher-dimensional one. In the literature, IMF (2021) revealed that SVM performed well in the backtesting exercise for the financial sector model, ranking as the second-best model of the eight ML approaches. However, SVM performed poorly in the real sector model, with RF and signal extraction having the best performance. Beutel, List, and von Schweinitz (2019) have shown similar results where SVM displayed poor out-of-sample predictive power in terms of AUC compared to other ML models.

3.2.1 Variable Importance

Variable importance measures the contribution of each variable to the prediction of fiscal stress. This step is crucial to analyze the predictive power of the models. We used the best predictors to reduce the complexity and dimensionality of the model without sacrificing its performance, using the following techniques:

1. **Signal Power:** This method is based on the signal power score defined above in Equation 4 that has been widely used in the signaling approach theory.

2. **Random Forest:** This tree-based technique uses the RF model to sort the importance of the variables. As discussed above in Section 3.2, this model allows calculating the degree of impurity reduction. The more the variable contributes to segregate the classes, the more important the variable is.

3. **Gradient Boosting:** As in the RF model, the GB model uses impurity to rank the importance of the variables.

4. **Recursive Feature Elimination (RFE):** The last technique implemented uses an algorithm that recursively eliminates groups of features until a certain predefined number of variables is reached. In order to have a wider scope, the logit model was used as the base model in the algorithm.

[4] In a p-dimensional space, a hyperplane is a flat subspace with dimension p-1. For example, in two dimensions, a hyperplane is a straight line, and in three dimensions, a plane.

Scores for each technique were normalized from 0 (least relevant variable) to 1 (most relevant variable) to estimate the importance of the variables using the methods presented above. The final score is equal to the sum of the scores. Thus, a variable with a score of 4 means it's the variable scored the highest in all methods.

3.2.2 Model Selection and Testing Procedure

We divided the dataset into training and test samples. Since the splitting method should consider both the time and country dimensions, our training sample contains all countries from 2005 to 2017. The test sample includes the remaining three years until 2020 for all countries —data preprocessing involved addressing outliers and missing values through RF imputation. We also normalized the dataset to improve the speed of the computational process for the classifiers.

Three resampling methods were tested to correct the 21% class imbalance found in the target variable for fiscal stress events. The first method considered random oversampling, in which observations were randomly selected from the minority class with replacement and included in the training set until the classes were equally distributed. The second method involved the synthetic minority oversampling technique (SMOTE), involving the selection of neighboring observations in the same feature space and creating a hyperplane of synthetic observations. This approach produces reasonable observations since they are close in feature space (Chawla et al., 2002). The last resampling technique consists of adjusting the weights of the training samples to emphasize the minority class during estimation.

The selected performance measure was F2, calculated as the harmonic mean between precision and recall[5] giving a higher weight to the recall, since it seeks to minimize the number of false negatives. Additionally, this metric takes into account class imbalance, which allows to better approximate the real distribution of the data.

The F2 metric also considers the cost associated with incorrectly forecasting a false negative event, (i.e., predicting that there will be no fiscal stress). This cost is higher than that of a false positive (indicating there is a fiscal stress event when there is not) since, in the former, governments will not be able to anticipate a crisis episode adequately. The latter is supposed to have far worse implications than the government preparing for a stress event that eventually does not occur.

The hyperparameters for each classifier were determined using heuristic optimization for 100 iterations with the F2 metric as the target variable. During each iteration, the model was trained using a continuous cross-validation technique with an extended training window and a test window size of three periods ahead. We describe this methodology in Figure 6. At each step of the cross-validation, the model was trained with a part of the training sample and then validated out-of-sample using the following three periods. The lamination process

[5] Precision is defined as the rate between crises that were correctly identified over the total number of times the model predicted a crisis. Recall is defined as the total number of times the model correctly predicted a crisis over the total number of real crises.

is performed for each country in the dataset. This process continues until it reaches the minimum number of five observations in the training sample.

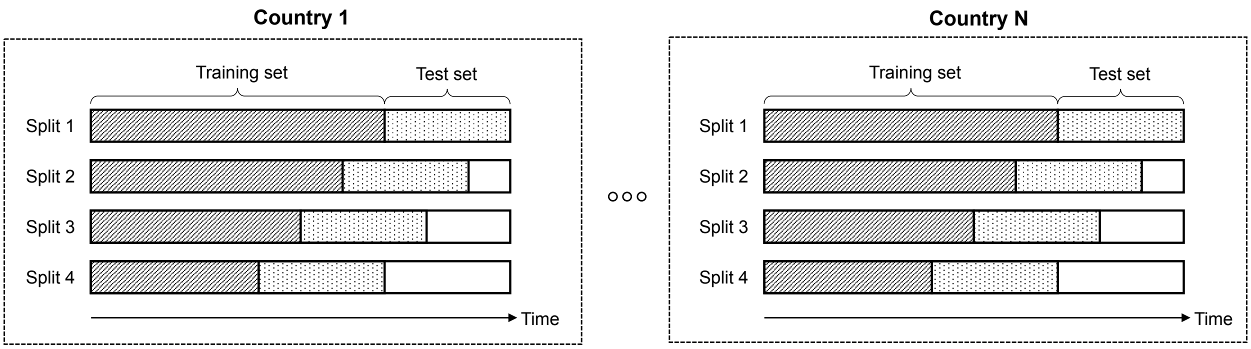

Figure 6: Example of Grouped Rolling Cross-Validation with Expanded Window.
Source: Authors' elaboration.

Following the optimization process, optimal classifiers are selected and estimated one last time to obtain the performance metrics for the training, validation, and test samples.

3.2.3 Accumulated Local Effects

We use the ALE method to estimate thresholds on each variable by using the predictions from the models described above. Let $\hat{f}(x_C, x_S)$ denote a prediction function on the feature values x_S and x_C. Then the gradient $\hat{f}^S\left(x_S, x_C\right) = \frac{\partial \hat{f}(x_S, x_C)}{\partial x_S}$ denotes the change of predictions $\hat{f}$ with respect to x_S. It isolates the effect of the feature of interest and blocks the effect of correlated features (Molnar, 2020). The ALE method averages these changes in the predictions conditional on the S features and integrates the derivative over the S features. Formally, following Molnar (2020):

$$\hat{f}_{x_S,ALE}(x_S) = \int_{z_{0,1}}^{x_S} \int_{X_C} \hat{f}^S(z_S, x_C) \mathbb{P}\left(x_C, z_S\right) dx_C dz_S - \text{constant}$$

The marginal expectation of changes in predictions is calculated over the features in set C conditional on each feature grid value of interest. The additional integral over z accumulates the local gradients over the range of features in set S. Finally, a constant is subtracted from the results so that the ALE is centered (i.e., the average effect is 0). To approximate the gradients the z's are replaced by a grid of intervals over which the changes in predictions are computed (Molnar, 2020). Then, the uncentered ALE can be computed as:

$$\hat{\tilde{f}}_{j,ALE}(x) = \underbrace{\sum_{k=1}^{k_j(x)} \underbrace{\frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in N_j(k)} \overbrace{\left[f\left(z_{k,j}, x_j^{(i)}\right) - f\left(z_{k-1,j}, x_j^{(i)}\right)\right]}^{\text{Effect for an individual instance at interval } k}}_{\text{Local (average) effect}}}_{\text{Accumulated local effect}} \tag{7}$$

In Equation 7, the feature of interest $x_j^{(i)}$ is replaced by the grid values $z_{k,j}$, so the difference in prediction corresponds to the effect that the feature has for an instance in a given interval. Next, the effects of all instances (neighborhood $N_j(k)$) are added and divided by the number of instances in that interval $(n_j(k))$. This average equals the local effect. It is accumulated across all intervals $k_j(x)$ so that the accumulated local effect of a feature that lies in the $j-th$ interval corresponds to the sum of the local effects in all preceding intervals. Based on Molnar (2020), Figure 7 illustrates the calculation of ALE for debt as a percentage of GDP in the presence of another correlated feature: debt as a percentage of revenues. The feature of interest is divided into $k = 5$ intervals. Then, for each data instance $z_{k,j}$ within an interval (blue dots), the difference in prediction is calculated when substituted for the upper and lower bound of the interval (blue horizontal lines). These differences are averaged within each interval to obtain local effects and finally accumulated across all intervals.

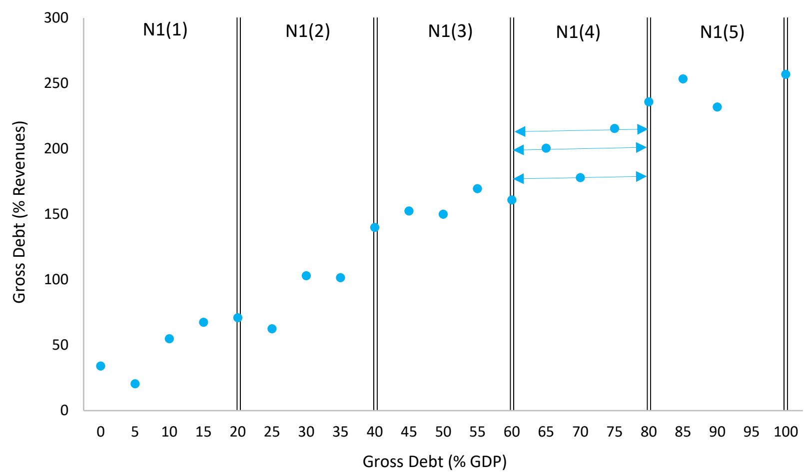

Figure 7: Accumulated Local Effects Calculation

Source: Based on Molnar (2020).

Finally, the ALE is centered so that the mean effect is 0:

$$\hat{f}_{j,ALE}(x) = \hat{\hat{f}}_{j,ALE}(x) - \frac{1}{n}\sum_{i=1}^{n}\hat{\hat{f}}_{j,ALE}(x_j^{(i)}) \tag{8}$$

The centered ALE (Equation 8) can be interpreted as the effect of the feature at a certain value compared to the average prediction of the data (Molnar, 2020). estimate the thresholds for predicting fiscal stress, we calculated the ALE centered on each characteristic and assumed that the threshold is the value at which it is 0. The probability of predicting fiscal stress is higher than the mean when a particular characteristic takes values above this threshold. Specifically, for each variable, we estimate the centered ALE by setting the grid size with $k \in \{25, 50, ..., 300\}$ and average the threshold estimates. The test data employ the ALE method, as the best ML models seek to improve the out-of-sample F2 metric. This enables estimating the critical values of the variables in the out-of-sample prediction, which allows us to properly construct the leading indicator in the following years (2021–26).

3.2.4 Leading Index of Fiscal Stress for Machine Learning Approaches

Similar to the signaling methodology, we follow Equation 5 to calculate the leading index of fiscal stress for ML models. Now, the signal d^i_{jt} sent by the variable i is equal to 1 when it exceeds its critical value estimated with the ALE method. Finally, the weights z_i are defined by variable importance estimations that we implemented to measure the individual contribution of each variable to the stress prediction. This is described in detail in Section 4.1.

Similar to the signaling methodology, we follow Equation 5 to compute the leading fiscal stress index for ML models. The d^i_{jt} signal sent by variable i is equal to 1 when it exceeds its critical value estimated with the ALE method. Weights z_i are defined by estimating the importance of the variables to measure each variable's contribution to the stress prediction. This is detailed in Section 4.1.

As in the signaling methodology, the resulting composite index L_{jt} is bounded between 0 (low-risk episode) and 100 (high-risk episode). We implemented this methodology for each ML approach described in Section 3.2 and we were also able to estimate leading indicators disaggregated by regions (advanced, emerging and Latin American economies). Results are discussed in the following section.

The leading indicator presented in this paper is based on the nonparametric approach proposed by Kaminsky, Lizondo, and Reinhart (1998) and incorporates calculations of the risk threshold level for each variable using the already proposed ALE technique. Contributions such as those made by Jarmulska (2020) or Hellwig (2021) demonstrate how ML models can be very good at forecasting fiscal stress. However, none of them takes these models to the construction of a leading indicator of macro-fiscal stress using risk thresholds.

4 Results

4.1 Variable Importance

Figure 8 presents the 20 most important variables for each of the country groups (advanced, emerging, and LAC), as well as for the aggregate of countries, which is used for the estimation of the model. The results of this exercise reveal differences between the variables that contribute most to each group. In particular, for advanced countries, the macroeconomic component plays a critical role in determining fiscal risk. For this group, there are variables of the labor market such as the Human Capital Index (measured as the return to education and the unemployment rate), macroeconomic stability variables -such as inflation volatility, and domestic demand variables such as gross fixed capital formation. Fiscal variables include domestic currency debt, tax revenues from wage tax and goods and services tax, total external debt, and fiscal balance.

On the other hand, emerging and LAC countries show a similar composition, with the payment of foreign currency liabilities as a percentage of GDP being the most relevant vari-

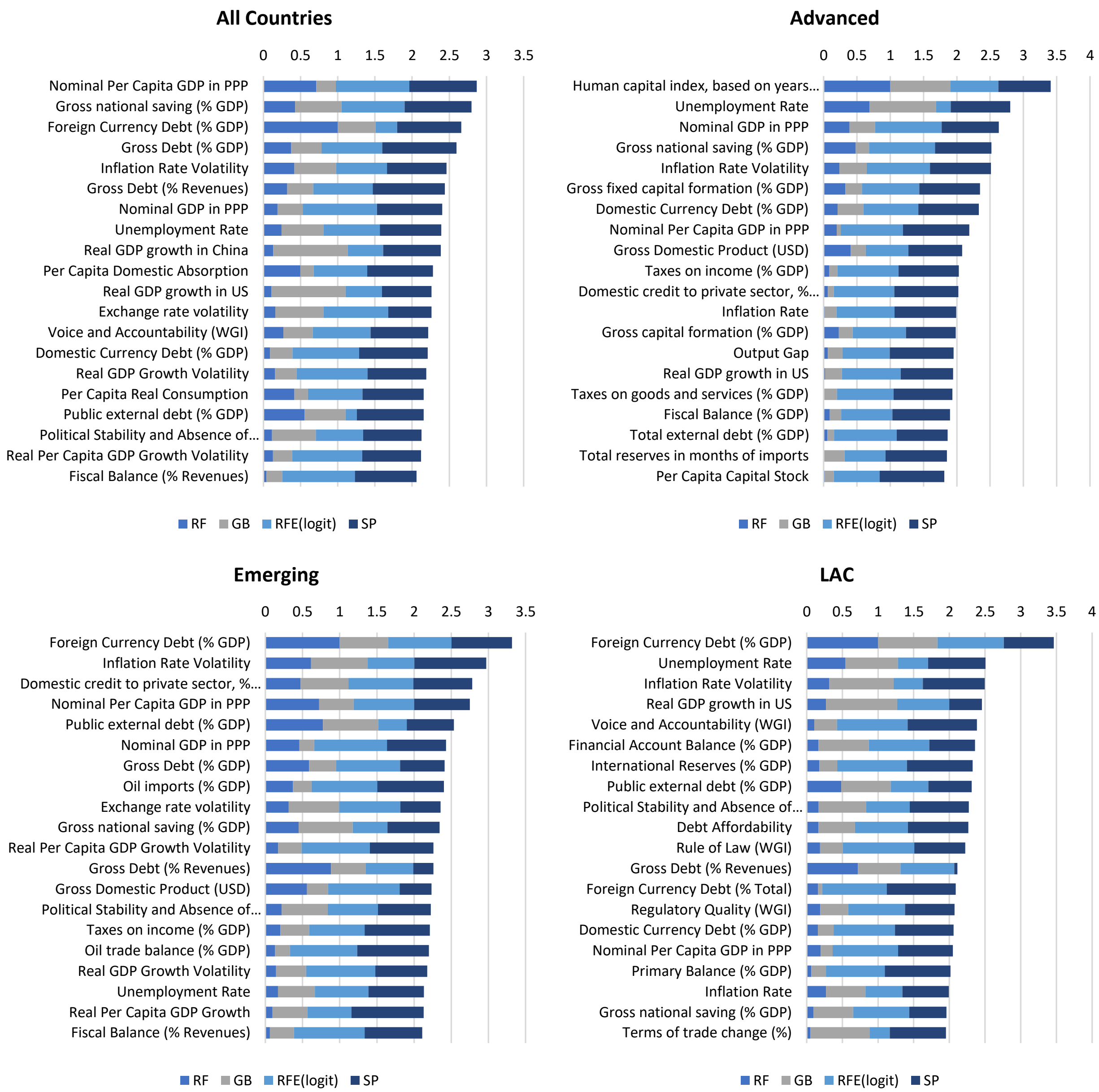

Figure 8: Variable Importance in All Country Groups

Source: Authors' elaboration.

able. This finding is consistent with different studies showing that the exposure of foreign currency liabilities can trigger fiscal stress events. Other relevant variables are indicators commonly related to a country's level of solvency and liquidity, such as gross debt or debt interest payments over tax revenues. Surprisingly, institutional variables play a crucial role in predicting fiscal stress events. This finding may evidence the relevance of political stability and good governance in the proper administration of LAC countries' fiscal accounts.

4.2 Forecast Performance

Once the variables that contribute most to the prediction of fiscal stress episodes are determined, the next step is to use these variables within the models that will be tested to find the best out-of-sample performance in predicting the crisis event. After an exhaustive hyperparameter optimization process, these models were selected to improve the out-of-sample performance metric and avoid overfitting the classifier.

Table 3 shows the results of the different performance metrics of the various models analyzed. Under the F2 metric, the best out-of-sample performance model with approximately 74.6% corresponds to the gradient boosting model, followed by the RF model with 74.3%. This is in line with the decision tree–based models over other supervised learning models such as SVM or the logistic model, which is consistent with Jarmulska (2020) or Hellwig (2021). In addition, there is a low out-of-sample predictive power of the signal model compared to its ML counterparts. Besides being consistent with results such as those presented by IMF (2021), this finding shows the importance of a metric such as F2 to measure forecasting performance, which is more appropriate under class imbalance scenarios such as the one presented here.

The predictive power of the logit model is also remarkable, as shown by metrics such as accuracy or by using receiver operating characteristic (ROC) curves[6] (see Figure 9) where the performance of this model is relatively better than other approaches. However, since it is more relevant for this study to minimize false negatives (number of undetected crises) than false positives (number of reported false crises), the gradient boosting model is considered to perform better.

4.3 Indicators Based on Critical Values

We estimate the critical values for each indicator and methodology. In the case of ML models, the ALE technique provides an advantage to study interactions and nonlinearities in the data to produce the critical values. Figures 10 and 11 show the graphical representation of the methodology results, including all country groups and taking the gross debt and interest payment to tax revenue variables as examples. We use the same variables throughout this section to show the indicators associated with liquidity and solvency and keep the results consistent. However, we could analogously perform the same exercise with any of the variables in the model. Moreover, we allow comparability between critical values under different models, enhancing the robustness of the results under various methodologies.

As expected, both variables show an increasing trend behavior. The higher the gross debt or interest payments relative to fiscal revenues, the higher the probability of incurring a fiscal stress event. Taking gross debt as an example, the risk observed for all countries in the

[6] ROC curves allow one to observe how well a model can distinguish between a fiscal stress episode and a tranquil one. They represent the rate of true positives (sensitivity) versus false positives (100 - specificity) for different cutoff points. The closer the curve is to the upper left corner, the greater the area under the curve (AUC) and thus the better the class discrimination.

Metric	Model	Train	Validation	Test
F2	Gradient Boosting	78.2	69.3	74.6
	Random Forest	78.0	69.1	74.3
	Support Vector	76.6	68.1	68.4
	Logit Regression	70.8	69.2	65.3
	Signaling Approach	61.9	-	26.4
F1	Support Vector	63.7	54.4	60.4
	Logit Regression	59.1	53.4	59.6
	Random Forest	61.9	51.5	59.1
	Gradient Boosting	61.9	51.7	59.0
	Signaling Approach	65.8	-	30.3
Accuracy	Logit Regression	71.8	69.8	76.1
	Support Vector	74.9	72.4	75.2
	Signaling Approach	84.6	-	71.9
	Random Forest	71.1	65.9	68.7
	Gradient Boosting	70.9	66.0	68.3
Precision	Logit Regression	46.3	38.9	51.9
	Support Vector	49.8	40.8	50.5
	Random Forest	46.1	36.3	44.0
	Gradient Boosting	45.9	36.5	43.8
	Signaling Approach	73.7	-	40.2
Recall	Gradient Boosting	94.9	90.3	90.4
	Random Forest	94.3	90.0	89.7
	Support Vector	88.5	82.3	75.0
	Logit Regression	81.7	86.6	69.9
	Signaling Approach	59.5	-	24.3
AUC	Logit Regression	82.9	76.2	82.4
	Support Vector	87.4	76.1	80.4
	Random Forest	89.6	75.0	79.6
	Gradient Boosting	89.4	75.1	79.4
	Signaling Approach	59.5	-	24.3

Table 3: Performance by Model

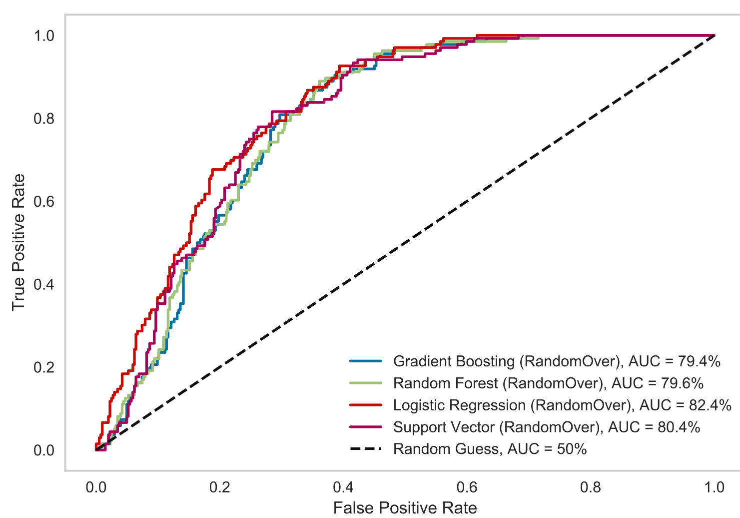

Figure 9: ROC Curves for All

Source: Authors' elaboration.

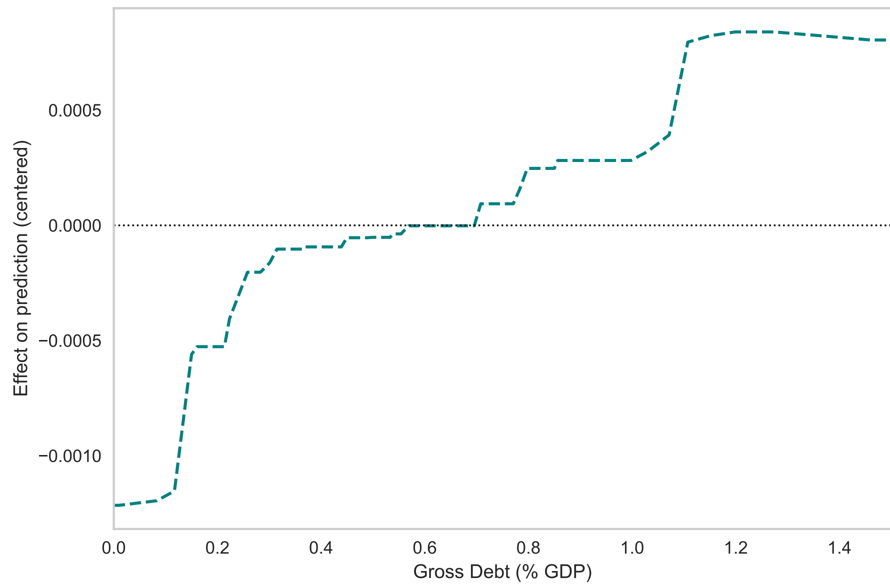

Figure 10: Marginal Plots for Gross Debt (% of GDP)

Source: Authors' elaboration.

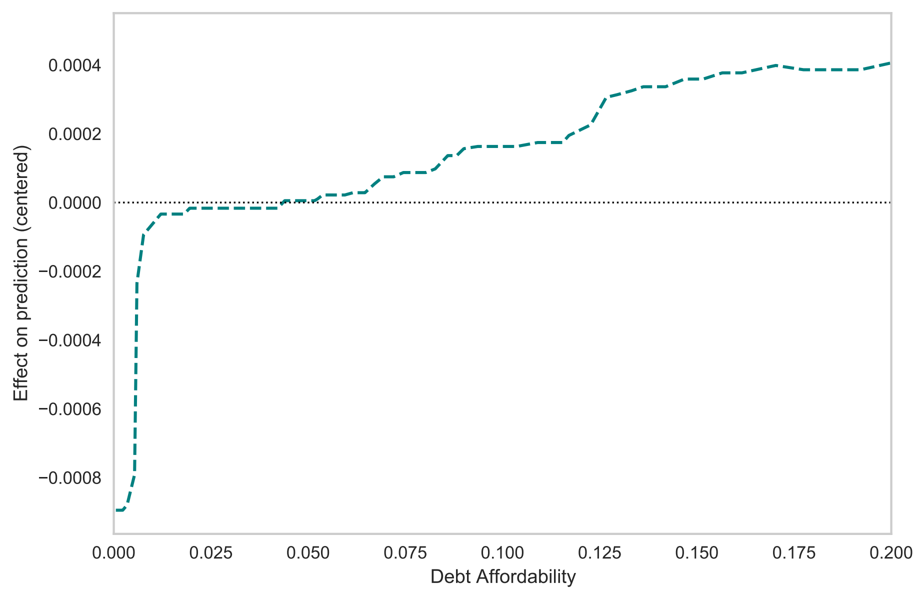

Figure 11: Marginal Plots for Debt Affordability

Source: Authors' elaboration.

aggregate has a cut-off point close to 56%. We also find that the effect on the determination of fiscal stress is less sensitive to marginal changes in gross debt than to marginal changes in interest payments to income. The shock probability increases rapidly between values of 1% and 5%.

4.3.1 Cartesian Risk Planes

The first visualization tool that uses the critical values is the Cartesian plane. This type of graph shows the evolution of each group of countries for a specific pair of variables. As seen in Figures 12 and 13, the white area corresponds to a low-risk region where neither of the two indicators exceeds the critical value. The light red area indicates a region in which one of the two indicators exceeds the threshold. Finally, the dark red zone represents the highest risk region, where both variables exceed the threshold. This same exercise could be performed with three variables using a three-dimensional hyperplane.

Figures 12 and 13 show the variation over time of the critical value found for the pair of variables gross debt and interest payment to fiscal revenues (debt affordability). The top-left graph shows at the same time the evolution of the three groups of countries using the overall critical values. The other three graphs show the evolution of each group using the critical value associated with that group as a comparison. This analysis reveals how countries or groups can be compared under two different scopes: within and between.

Cartesian Planes Using Signal Approach

Taking LAC as an example, the Cartesian risk planes based on the signaling model show the indicator's trend over time. This indicator deteriorated the most during the COVID-19 crisis, as both variables reached their historical peak and entered the high-risk zone. Regarding debt affordability, relative to its group, the signaling approach did not send risk signals in this indicator before 2019. This indicator sent signals throughout all periods if viewed comparatively with the overall threshold. Thus, the risk perception of the level of indebtedness concerning fiscal revenues for LAC is higher than the global one, as it is close to 10.5%. At the same time, it is 7.7% and 10.5% for emerging and advanced economies, respectively.

Cartesian Planes Using Gradient Boosting

Risk levels in LAC show similar trends to the critical values obtained by the signaling method, except for some crucial differences. We found a higher global and regional gross debt threshold than the signals model and a lower level for the debt affordability metric. Compared to the global threshold, the region only enters the higher risk zone after 2020. In addition, and in contrast to the signals model, it is observed that by 2015 and compared to its group, the region entered the riskier zone for the debt affordability metric, but not for gross debt, which entered the riskier zone threshold from 2019 onwards.

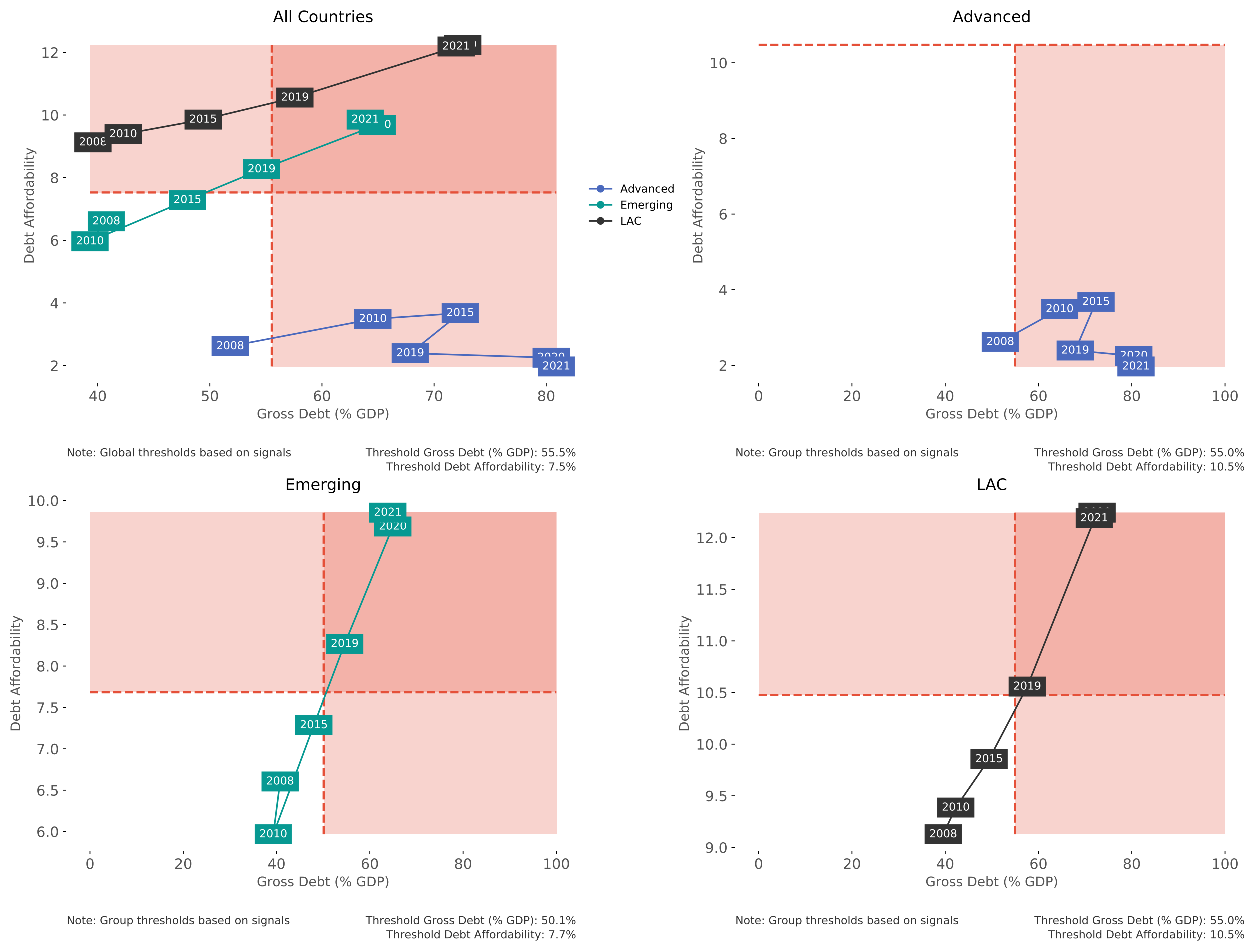

Figure 12: Cartesian Planes Using Signaling Model

Source: Authors' elaboration.

4.4 Heat Maps

A second visualization tool used for the EWS consists of heat maps relative to the difference between the indicator value and the critical threshold. Unlike Cartesian risk planes, this form of visualization shows more directly how far the indicator is from its risk zone. These graphs show a color spectrum ranging between green and red, where a dark green color depicts a state in which the group of countries is far away from the risk zone. Light yellow represents similar or equal to the risk limit values. Conversely, a dark red color characterizes states where the indicator is well above the risk level.

Figures 14 to 17 display the heat maps regarding the critical value for gross debt and interest payments on the income variable for the periods 2019, 2020, and 2021. The risk measures the distance of the present value of gross debt to the thresholds. We performed two types of analysis: the within analysis corresponds to studying the risk within each group of countries, and the between analysis compares the risk of each group with the global level.

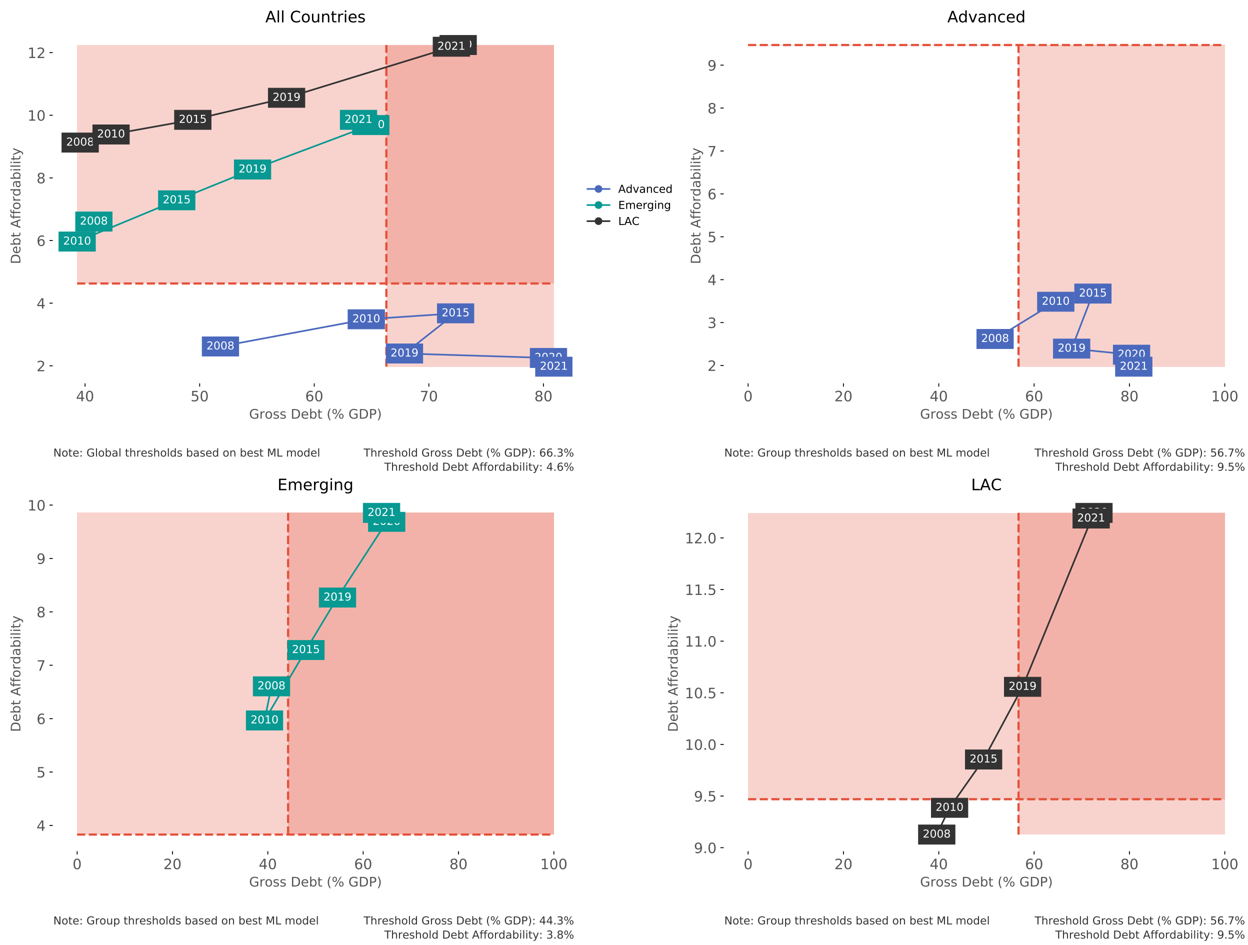

Figure 13: Cartesian Planes Using Gradient Boosting.

Source: Authors.

4.4.1 Heat Maps Using the Signal Approach

The heat map of gross debt using the signaling model (Figure 14) shows the distance of debt levels to risk threshold values for each country group. This threshold corresponds to 55.5% for all countries, 73.2% for advanced countries, 50.1% for emerging countries, and 55.0% for LAC. The LAC region goes from being relatively close to the risk limit in 2019 and increases in deviation from the risk limit in 2020. The projections show a slow recovery in terms of indebtedness for the LAC region, as the level of debt will still be in the risk zone by 2021, exceeding the limit by about 16pp. Emerging countries show similar behavior, but a less pronounced increase in debt levels, rising by 10pp. Advanced countries will move from the nonrisk zone to the risk zone, increasing by 13pp. In summary, these findings show how the impact of the pandemic's impact on debt levels affected countries in the LAC region to a greater extent than advanced or emerging countries when compared against risk thresholds that adjust for each group.

On the other hand, the heat map for interest payments to fiscal revenue of the signaling

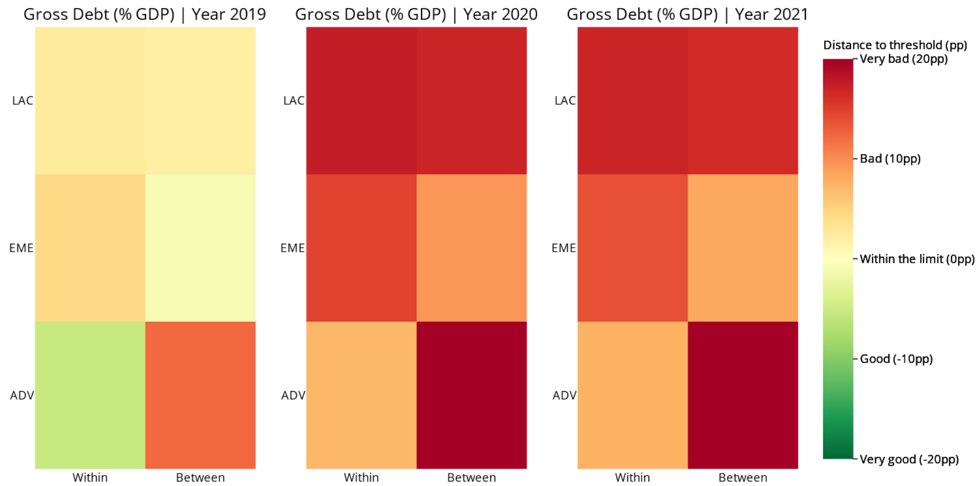

Figure 14: Map of Gross Debt (% of GDP) Using Signaling Approach

Source: Authors' elaboration.

Figure 15: Heat Map of Debt Affordability Using Signaling Approach

Source: Authors' elaboration.

model (Figure 15) is based on a critical level of 7.5% for the global level, 5% for advanced countries, 7.7% for emerging countries, and 10.5% for LAC. Generally speaking, the LAC region and emerging countries suffered a constant deterioration close to 2pp in their ability to pay their obligations, with projections to remain at this level in 2021. Advanced countries showed greater strength in interest payments since their level never exceeded the critical value and remained constant.

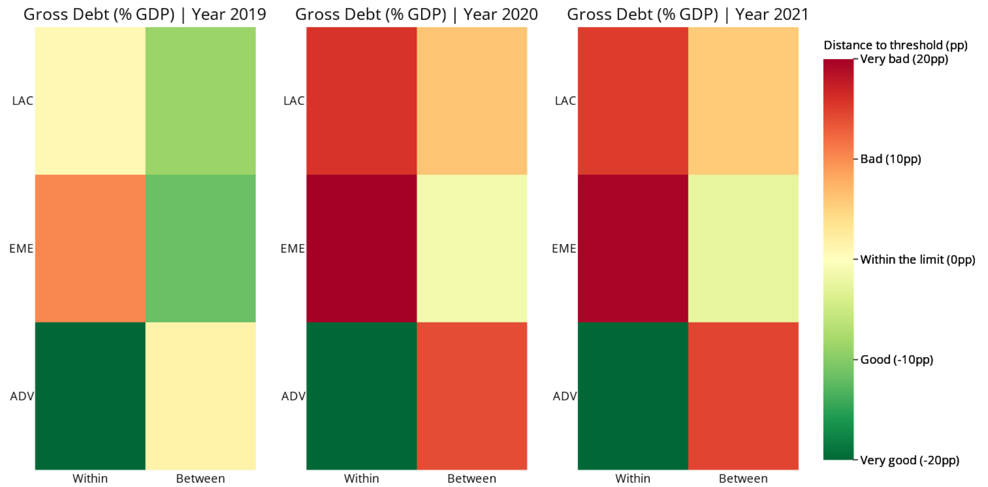

Figure 16: Heat Map of Gross Debt (% of GDP) Using Gradient Boosting Model
Source: Authors' elaboration.

Figure 17: Heat map of debt Map of Debt Affordability Using Gradient Boosting Model
Source: Authors' elaboration.

4.4.2 Heat Maps Using Gradient Boosting

The heat maps for the gross debt variable using the GB model (Figure 16) also show an increased risk by 2020 and slow recovery by 2021. Nevertheless, the thresholds are generally higher than in the signaling model: 66.3% at the global level, 109.4% for advanced countries, and LAC with 56.7%. Emerging countries exhibit a lower risk threshold of 44.3%.

The results show risk metric deterioration for gross debt. Within countries, the group of countries showing the highest risk is emerging countries, while advanced countries deteriorate the most when compared between groups of countries. This is because advanced countries historically have had higher debt levels than other countries. It turns out that compared to the global threshold, the advanced countries are in the risk zone.

The heat map analysis for interest payments to fiscal revenue using the gradient boosting model (Figure 17) shows a critical level of 4.6% at the global level, 3.8% for emerging countries, and 9.5% for LAC. These thresholds are lower for all cases than the signal model. Given that this variable is not a good predictor of fiscal stress in this group of countries (i.e., it doesn't properly distinguish tranquil and stress episodes), the ALE doesn't estimate differences in the prediction of probabilities and therefore the threshold is not calculated. The results reveal a loss in the ability of LAC and emerging countries to pay debt obligations, which is partly due to the contractions in fiscal revenue and sharp exchange rate depreciations in both regions in recent years.

4.5 Leading Indicator

The last EWS visualization tool is the macro-fiscal risk leading indicator, which reduces dimensionality, facilitates grouping relevant variables to explain fiscal stress episodes, and propels the evolution of macro-fiscal imbalances. This indicator captures the idea that a fiscal stress episode is more likely to occur if several variables show signals simultaneously.

The leading indicator presents a fundamental tool for government decision making as it allows for monitoring risk dynamics through different factors. This indicator uses early warning signals sent through different variables that are relevant to determine an episode of macro-fiscal stress, making it possible to anticipate shocks that may affect public finances and debt.

The leading indicator is on a scale from 0 to 100, where 0 means low risk of a fiscal stress episode and 100 means high risk. This indicator measures short-term macro-fiscal vulnerability at the aggregate level and for each component factor. The higher the indicator's value, the more variables send signals, indicating a possible episode of fiscal stress.

The indicator uses the signal approximation methodology and includes the critical values calculated by the ALE with weights given by the variable importance index. This form of estimation adds value to calculate the leading indicator by combining the results from the best ML model.

Figures 18, 19, and 20 show the leading indicator from 1990 to 2020 with 2025 projections for each country group and each estimation methodology. The vertical bars corresponding to the right axis show the number of countries presenting an actual fiscal stress episode. The estimation of the indicator uses the global critical values so that the results are comparative across country groups.

Results show how the two indicators associated with each of the methodologies present a somewhat similar behavior, mainly as the series follows the trend of actual episodes of fiscal stress over time. However, there are considerable differences between their dynamics, especially during the recovery times of the crises and the magnitudes reached by the indicator in response to episodes of fiscal stress.

In addition, there are different quantitative responses between the global financial crisis and the recent COVID-19 pandemic. Both crashes seem to have increased the fiscal stress indicator in the same proportion for advanced countries, to about 40 points. However, based on projections, the recovery time from the pandemic-associated crisis is expected to be faster than that presented during the 2008–2009 financial crisis.

In the case of emerging economies, the crisis associated with the pandemic has a slightly higher effect than the financial crisis. The projection presented by the indicator associated with the gradient boosting model is more conservative than the signal model in terms of the value that will be reached in the subsequent periods. Finally, in the case of LAC, the recent crisis had a more substantial impact on the region compared to previous financial crises. Unlike advanced countries, the recovery for the 2020 crisis is slower than the recovery time for the 2008 crisis.

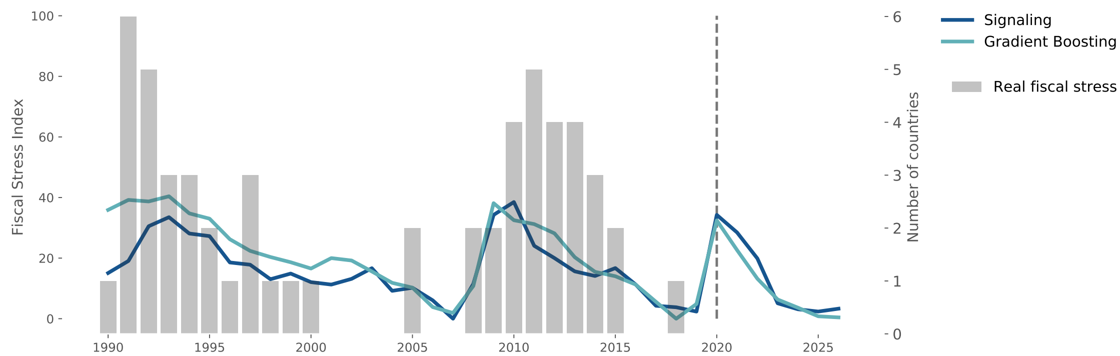

Figure 18: Leading Indicator for Advanced Countries
Source: Authors' elaboration.

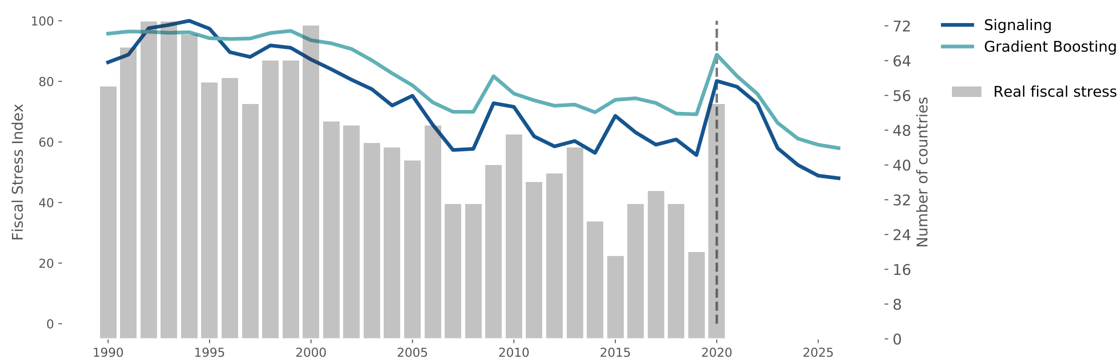

Figure 19: Leading Indicator for Emerging Countries
Source: Authors' elaboration.

4.5.1 Leading Indicator Decomposition

Figures 21, 22, and 23 display the decomposition of the leading indicator. It shows the evolution of the indicator in each group of countries for fiscal, macroeconomic, and institutional

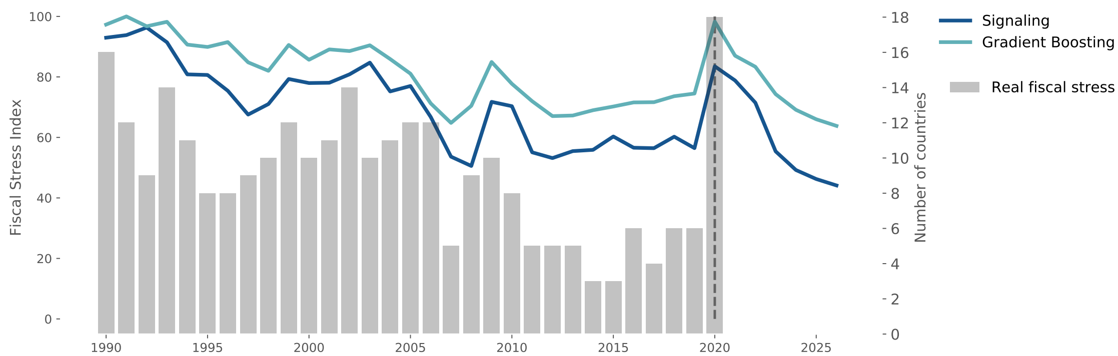

Figure 20: Overall Leading Indicator for LAC Countries

Source: Authors' elaboration.

components. The construction of the indicators uses the predictors of each group (Appendix A). It takes as a reference the global critical value, which allows comparing the gradient boosting methodology (solid lines) with the signal model (dotted lines).

The fiscal risk indicator for the LAC and emerging economies has been increasing since 2010 and reached its historical maximum during 2020. As for the advanced economies, fiscal risk levels fell from 2010 after the international financial crisis and rose again to a higher risk level in 2020.

For all economies, the macroeconomic indicator shows a constant reduction in the risk level since 2009 that is only interrupted in 2020. Figures 21, 22, and 23 also show that the risk of a crisis associated with the macroeconomic factor was higher during the international financial crisis than the risk associated with the 2020 pandemic.

The institutional indicator presents a steady behavior over time for all three economies. This indicator presents a consistently high risk for emerging and LAC economies, while it presents a constantly low risk for advanced economies. This result shows the need to strengthen institutions in emerging countries in terms of governance, control of corruption, and political stability to achieve adequate control of fiscal vulnerabilities.

The same analysis can be further disaggregated by using the subfactors of each of the components. This would provide an even more disaggregated view of fiscal, macroeconomic, and institutional vulnerabilities. Additionally, it is possible to perform a similar analysis, but looking at risk from an internal perspective (i.e., using the within-groups critical values instead of the between-groups critical values, which were used in this example).

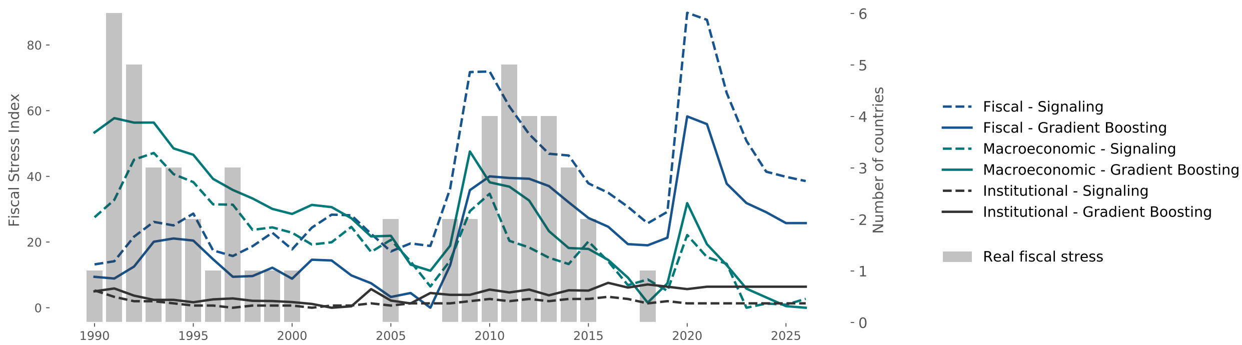

Figure 21: Decomposition of leading indicator for advances countries

Source: Authors.

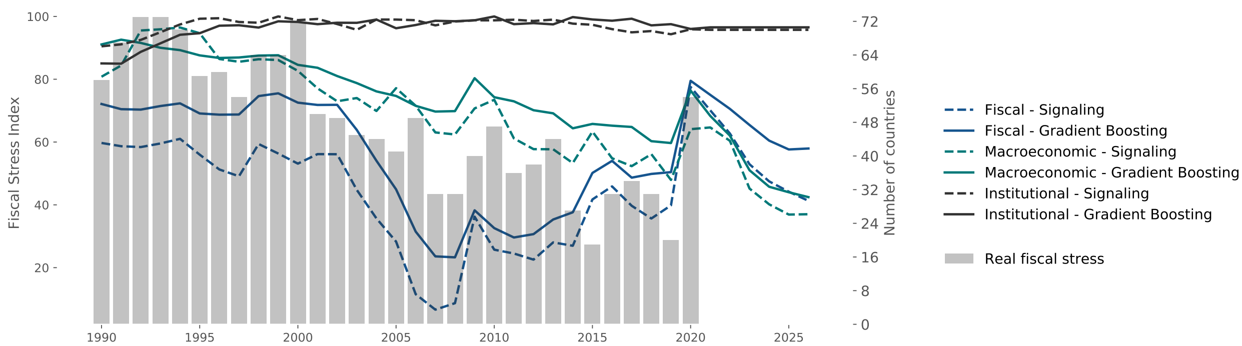

Figure 22: Decomposition of Leading Indicator for Emerging Countries

Source: Authors' elaboration.

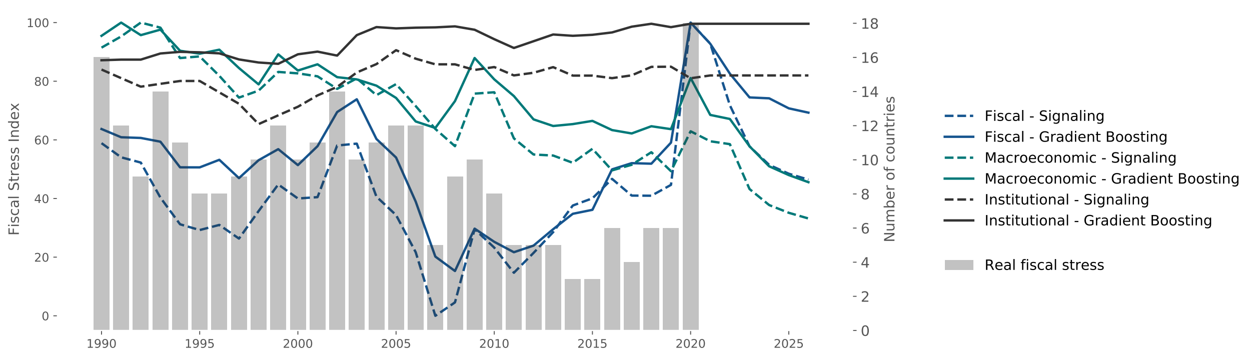

Figure 23: Decomposition of Leading Indicator for LAC Countries

Source: Authors' elaboration.

5 Conclusions

This study aims to extend the current literature on fiscal stress determination in the EWS framework. In particular, it tackles two of the major disadvantages of the traditional signals approach: overlooking the correlation between predictors and poor out-of-sample performance. To address the first issue, we used several ML models and estimated critical values in the predictors with the ALE technique proposed by Apley and Zhu (2020) to isolate each predictor's effect, considering the impact of correlated variables. The second was addressed using models that learned to predict fiscal distress events in validation and test datasets, revealing the weak predictive power of the signal model compared to its Machine Learning counterparts.

This paper finds that macro and long-term fundamentals-related variables are more relevant in explaining fiscal risk in advanced economies. In contrast, fiscal variables are more important determinants of fiscal stress for emerging economies and LAC. The risk dynamics are accentuated for all country groups with increases in debt levels and are more persistent after the global financial crisis.

Leading indicators show slower recovery times for emerging economies and LAC than for advanced economies in the coming years. In addition, the incidence of fiscal stress was higher during the pandemic than in the global financial crisis. Leading indicators show the importance of institutional factors in predicting fiscal stress events. This result suggests strengthening and monitoring institutional variables to mitigate and reduce macro-fiscal risk.

References

Apley, D. W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):1059–1086. Available at: https://arxiv.org/abs/1612.08468.

Baldacci, E., Petrova, I., Belhocine, N., Dobrescu, G., and Mazraani, S. (2011). Assessing fiscal stress. No. 11/100. Available at: https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Assessing-Fiscal-Stress-24822.

Beers, D. and de Leon-Manlagnit, P. (2021). The boc-boe sovereign default database: What's new in 2021? *Staff Analytical Notes*, 15. Available at: https://EconPapers.repec.org/RePEc:bca:bocsan:21-15.

Berg, A., Borensztein, E., and Pattillo, C. (2005). Assessing early warning systems: How have they worked in practice? *IMF Working Papers*, 04. doi:10.5089/9781451847284.001.

Berti, K., Salto, M., and Lequien, M. (2012). An early-detection index of fiscal stress for eu countries. *European Commission, DG ECFIN - European Economy Economic Paper*. ISBN: 978-92-79-22996-1.

Beutel, J., List, S., and von Schweinitz, G. (2019). An evaluation of early warning models for systemic banking crises: Does machine learning improve predictions? IWH Discussion Papers 2/2019, Halle (Saale). Available at: http://hdl.handle.net/10419/191248.

Boser, B., Guyon, I., and Vapnik, V. (1996). A training algorithm for optimal margin classifier. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 5.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32. Available at: https://link.springer.com/article/10.1023/A:1010933404324.

Cerovic, S., Gerling, K., Hodge, A., and Medas, P. (2018). Predicting fiscal crises. *IMF Working Papers*, 18:1. Available at: https://www.imf.org/en/Publications/WP/Issues/2018/08/03/Predicting-Fiscal-Crises-46098.

Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res. (JAIR)*, 16:321–357.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Alizada, N., Altman, D., Bernhard, M., Cornell, A., Fish, M. S., et al. (2021). V-dem dataset v11. 1. Available at: https://www.v-dem.net/vdemds.html.

Dawood, M., Horsewood, N., and Strobel, F. (2017). Predicting sovereign debt crises: An Early Warning System approach. *Journal of Financial Stability*, 28(C):16–28.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–82. Available at: https://www.rug.nl/ggdc/productivity/pwt/?lang=en.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29:1189–1232. Available at: https://www.bibsonomy.org/bibtex/237ca72b4c7f9383050b7c50da4356802/nosebrain.

Gerling, K., Medas, P., Poghosyan, T., Farah-Yacoub, J., and Xu, Y. (2017). Fiscal crises. *IMF Working Paper*, No. 17/86. Available at: https://www.imf.org/en/Publications/WP/Issues/2017/04/03/Fiscal-Crises-44795.

Hellwig, K.-P. (2021). Predicting fiscal crises: A machine learning approach. *IMF Working Papers No. 2021/150*, page 150. Available at SSRN: https://ssrn.com/abstract=4026328.

Hernández de Cos, P., Koester, G. B., Moral-Benito, E., and Nickel, C. (2014). Signalling fiscal stress in the euro area: A country-specific early warning system. *European Economics: Political Economy & Public Economics eJournal*. ISBN: 978-92-899-1120-7.

Hilscher, J. and Nosbusch, Y. (2007). Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt. *Review of Finance*, 14.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

IMF (2020). World economic outlook, october 2020: A long and difficult ascent. *International Monetary Fund*. Available at: https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020.

IMF (2021). How to assess country risk: The vulnerability exercise approach using machine learning. *Technical Notes and Manuals*, 2021(003).

Jarmulska, B. (2020). Random forest versus logit models: which offers better early warning of fiscal stress? *ECB Working Paper 2408*. Available at: https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2408~aa6b05aed7.en.pdf.

Kaminsky, G., Lizondo, S., and Reinhart, C. (1998). Leading indicators of currency crises. *International Monetary Fund*, 45(1). Available at: https://www.imf.org/external/Pubs/FT/staffp/1998/03-98/pdf/kaminsky.pdf.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2010). The worldwide governance indicators: Methodology and analytical issues. *World Bank Policy Research Working Paper*, (5430). Available at: https://openknowledge.worldbank.org/handle/10986/3913.

Kose, M. A., Kurlat, S., Ohnsorge, F., and Sugawara, N. (2017). A cross-country database of fiscal space. *CAMA Working Paper 2017-48*. Available at: https://www.worldbank.org/en/research/brief/fiscal-space.

Lane, P. and Milesi-Ferretti, G. M. (2017). International financial integration in the aftermath of the global financial crisis. *IMF Working Papers*, 17:1.

Lane, P. R. and Milesi-Ferretti, G. M. (2007). The external wealth of nations mark ii: Revised and extended estimates of foreign assets and liabilities, 1970–2004. *Journal of international Economics*, 73(2):223–250. Available at: https://www.brookings.edu/2021/09/16/the-external-wealth-of-nations-september-2021-update/.

Maltritz, D. and Molchanov, A. (2014). Country credit risk determinants with model uncertainty. *International Review of Economics Finance*, 29:224–234.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com, 2 edition. Available at:christophm.github.io/interpretable-ml-book/.

Moreno Badia, M., Medas, P. A., Gupta, P., and Xiang, Y. (2020). Debt is not free. *Journal of International Money and Finance*, 15. Available at: https://www.imf.org/en/Publications/WP/Issues/2020/01/03/Debt-Is-Not-Free-48894.

Santosa, F. and Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330.

Sumner, S. P. and Berti, K. (2017). A Complementary Tool to Monitor Fiscal Stress in European Economies. (049). Available at: https://ideas.repec.org/p/euf/ecopap/0475.html.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320. Available at: http://www.jstor.org/stable/3647580.

Appendices

A Predictors of Fiscal Stress

The following table lists the predictors used in the models, as well as the mean for each under tranquil and stress periods (2000–20). The p-value of a t-test is calculated (the two means are equal under the null hypothesis). Finally, the percentage of missing values is presented.

Table 4: Predictors of Fiscal Stress: Sources and Descriptive Statistics

#	Factor	Subfactor	Variable	Source	Mean (Tranquil)	Mean (Stress)	P-value	Missings
1			Gross Domestic Product (Local Currency)	WEO	163977.72	44536.46	0.01	0.00
2			Gross Domestic Product (USD)	WEO	461.41	53.71	0.00	0.00
3			Real Gross Domestic Product (Local Currency)	WEO	133823.35	50200.18	0.00	0.00
4			Real GDP Growth	WEO	0.04	0.03	0.00	0.00
5			Real GDP Growth Volatility	WEO	0.02	0.03	0.00	0.01
6		GDP	Real GDP Per Capita	WEO	1933.36	997.08	0.00	0.00
7			Real Per Capita GDP Growth	WEO	0.02	0.01	0.00	0.00
8			Real Per Capita GDP Growth Volatility	WEO	0.02	0.03	0.00	0.01
9			Nominal GDP in PPP	WEO	642.63	129.42	0.00	0.00
10			Nominal Per Capita GDP in PPP	WEO	21531.71	7212.14	0.00	0.00
11			Potential Output	WEO	133596.78	51023.89	0.00	0.00
12			Output Gap	WEO	0.00	-0.01	0.00	0.00
13			Unemployment Rate	ILOSTAT	0.07	0.09	0.00	0.03
14		Labor Market	Labor Productivity (output per worker)	Penn World Tables	50386.84	20020.62	0.00	0.05
15			Human Capital Index	Penn World Tables	2.64	2.13	0.00	0.05
16			Total Domestic Demand (% GDP)	WEO	1.03	1.09	0.00	0.00
17			Gross Capital Formation (% GDP)	WEO	0.25	0.21	0.00	0.00
18			Gross Fixed Capital Formation (% GDP)	WEO	0.24	0.20	0.00	0.00
19		Domestic Demand	Per Capita Real Consumption	Penn World Tables	15217.31	6290.01	0.00	0.05
20			Per Capita Domestic Absorption	Penn World Tables	21271.24	7893.74	0.00	0.05
21			Domestic Credit to Private Sector, % of GDP	Kose et al (2017)	0.72	0.32	0.00	0.03
22			Per Capita Capital Stock	Penn World Tables	95866.75	33081.78	0.00	0.05
23			Real Internal Rate of Return	Penn World Tables	0.11	0.11	0.85	0.05
24			Inflation Rate	WEO	0.05	0.10	0.00	0.00
25	Macro	Macro Stability	Inflation Rate Volatility	WEO	0.02	0.06	0.00	0.01
26			Exchange Rate Change (%)	WEO	0.02	0.08	0.00	0.00
27			Exchange Rate Volatility	WEO	0.06	0.11	0.00	0.01
28			Gross National Saving (% GDP)	WEO	0.23	0.16	0.00	0.03
29		Liquidity	International Reserves (% GDP)	World Bank	0.20	0.15	0.00	0.03
30			Total Reserves in Months of Imports	World Bank	4.88	4.05	0.00	0.05
31			FX Reserves Minus Gold (% GDP)	Lane et al (2021)	0.19	0.14	0.00	0.00
32			Current Account Balance (% GDP)	WEO	-0.01	-0.04	0.00	0.00
33		Balance of Payments	Financial Account Balance (% GDP)	WEO	0.00	-0.02	0.00	0.01
34			Direct Investment, Net (% GDP)	WEO	-0.03	-0.03	0.90	0.01
35			Trade balance: Exports - Imports (% GDP)	WEO	-0.04	-0.09	0.00	0.00
36			Exports of Goods and Services (% GDP)	WEO	0.45	0.34	0.00	0.00
37			Imports of Goods and Services (% GDP)	WEO	0.49	0.43	0.00	0.00
38			Oil trade Balance (% GDP)	WEO	0.02	-0.02	0.00	0.03
39			Oil Exports (% GDP)	WEO	0.07	0.04	0.00	0.03
40		Commercial	Oil Exports (% Total exports)	WEO	0.14	0.10	0.00	0.03
41			Oil Exports Volatility	WEO	0.01	0.01	0.40	0.04
42			Oil Imports (% GDP)	WEO	0.06	0.05	0.68	0.03
43			Trade Openness: Exports + Imports (% GDP)	WEO	0.93	0.77	0.00	0.00
44			Terms of Trade Change (%)	WEO	0.01	0.01	0.58	0.03
45			Terms of Trade Volatility	WEO	0.07	0.09	0.00	0.03
46			Natural resources rents (% GDP)	World Bank	0.07	0.07	0.31	0.07
47		Natural Resources Rents	Oil Economic Rents (% GDP)	World Bank	0.04	0.03	0.00	0.07
48			Commodity Dependence	WEO	0.28	0.28	0.91	0.06
49		External	Real GDP growth in US	WEO	0.02	0.02	0.18	0.00
50			Real GDP Growth in China	WEO	0.09	0.09	0.41	0.00

Table 4: *Predictors of Fiscal Stress: Sources and Descriptive Statistics (continued)*

#	Factor	Subfactor	Variable	Source	Mean (Tranquil)	Mean (Stress)	P-value	Missings
51			Fiscal Revenue (% GDP)	WEO	0.31	0.24	0.00	0.00
52			Tax Revenue (% GDP)	WEO	0.18	0.16	0.00	0.05
53			Non-tax Revenue (% GDP)	WEO	0.13	0.08	0.00	0.05
54			Tax Revenue Volatility	WEO	0.01	0.01	0.00	0.07
55		**F. Revenues**	Taxes on Income (% GDP)	WEO (October 2019)	0.07	0.05	0.00	0.09
56			Taxes on Income Ppayable by Individuals (% GDP)	WEO	0.04	0.02	0.00	0.16
57			Taxes on Income Payable by Corporations (% GDP)	WEO (October 2019)	0.03	0.03	0.00	0.16
58			Taxes on Goods and Services (% GDP)	WEO (October 2019)	0.08	0.07	0.00	0.09
59			Taxes on International Trade and Transactions (% GDP)	WEO (October 2019)	0.02	0.03	0.00	0.42
60			Total Expenditure (% GDP)	WEO	0.32	0.27	0.00	0.00
61			Interest Expense (% GDP)	WEO	0.01	0.02	0.00	0.06
62		**F. Expenditures**	Current Expenditure (% GDP)	WEO (October 2019)	0.26	0.20	0.00	0.37
63			Capital Expenditure (% GDP)	WEO (October 2019)	0.05	0.05	0.51	0.14
64			Budget Inflexibility (% Total Expenditure)	WEO (October 2019)	0.70	0.67	0.00	0.08
65			Primary Balance (% GDP)	WEO	0.00	-0.01	0.02	0.06
66	**Fiscal**		Fiscal Balance (% GDP)	WEO	-0.02	-0.03	0.00	0.00
67		**F. Balances**	Primary Balance (% Revenues)	WEO	-0.03	-0.05	0.00	0.06
68			Fiscal Balance (% Revenues)	WEO	-0.09	-0.16	0.00	0.00
69			Cyclical Adjusted Balance (% Potential Output)	WEO	-0.02	-0.03	0.00	0.01
70			Gross Debt (% GDP)	WEO	0.49	0.68	0.00	0.00
71			Gross Debt (% Revenues)	WEO	2.01	3.76	0.00	0.00
72			Foreign Currency Debt (% GDP)	WEO (October 2019)	0.18	0.44	0.00	0.14
73			Domestic Currency Debt (% GDP)	WEO (October 2019)	0.33	0.26	0.00	0.17
74			Foreign Currency Debt (% Total)	WEO (October 2019)	0.37	0.63	0.00	0.14
75			Domestic Currency Debt (% Total)	WEO (October 2019)	0.63	0.36	0.00	0.17
76		**Debt**	Debt Affordability	WEO / Worldbank	0.06	0.11	0.00	0.06
77			Effective Interest Rate on Gross Debt (%)	WEO	-0.03	0.04	0.04	0.09
78			Total External Debt (% GDP)	Kose et al (2017)	1.49	0.76	0.00	0.16
79			Private External Debt (% GDP)	Kose et al (2017)	1.25	0.32	0.00	0.16
80			Public External Debt (% GDP)	Kose et al (2017)	0.31	0.48	0.00	0.16
81			Total Debt Held by Non-Residents (% GDP)	World Bank	0.24	0.06	0.00	0.12
82			Total Debt Held by Non-Residents (Short-Term) (% GDP)	World Bank	0.02	0.00	0.00	0.10
83			Control of Corruption (WGI)	World Bank	0.17	-0.50	0.00	0.05
84			Rule of Law (WGI)	World Bank	0.17	-0.53	0.00	0.05
85			Regulatory Quality (WGI)	World Bank	0.20	-0.47	0.00	0.05
86	**Institutional**		Government Effectiveness (WGI)	World Bank	0.21	-0.53	0.00	0.05
87			Political Stability and Absence of Violence (WGI)	World Bank	0.12	-0.43	0.00	0.05
88			Voice and Accountability (WGI)	World Bank	0.10	-0.33	0.00	0.05
89			Democracy Indicator	Coppedge et al. (2021)	0.60	0.44	0.00	0.03